

# Incidenti da arma da fuoco in Texas

Statistic for Spatial Data  
Corso di laurea in Statistica e Data Science  
Andrea Marchetti



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

# *Introduzione*

Gli Stati Uniti hanno il più alto numero di armi da fuoco al mondo: nel 2018 il Congressional Research Service, ha stimato che ci sono circa 393 milioni armi da fuoco in possesso ai civili.

In termini di rapporto tra il numero di armi da fuoco e la popolazione, negli Stati Uniti, è di circa 120 armi da fuoco ogni 100 persone.

Secondo il rapporto "2019 Gun Law Scorecard", il Texas ha il secondo più alto numero di armi da fuoco registrate negli Stati Uniti, con un totale di circa 588.696 armi registrate.

Tuttavia, è importante notare che questo dato non rappresenta il numero totale di armi da fuoco possedute dai residenti del Texas, poiché molte armi da fuoco non sono registrate.

Purtroppo, il numero di vittime delle armi negli Stati Uniti e in Texas è in aumento negli ultimi anni. Nel 2019, sono state registrate 39.707 morti per armi negli Stati Uniti, di queste 3353 avvenute in Texas con un tasso di mortalità, rispettivamente, del 12.0 e 11.7 per 100.000 abitanti.

## *Obiettivo dell'analisi*

Nel mio progetto ho analizzato gli incidenti da arma da fuoco nello stato del Texas, suddividendo gli eventi mortali da quelli con solo feriti o persone incolumi.

Dopo una veloce analisi esplorativa per lo stato del Texas mi concentrerò sulla contea di Harris, verificando attraverso un'analisi univariata, per entrambi gli eventi, la presenza di qualche relazione spaziale: cluster, regolarità o csr.

Analizzerò quanti eventi avvengono in prossimità dei dipartimenti della polizia locale.

Infine, tramite un'analisi multivariata, verificherò se è presente un'interazione spaziale tra i due tipi di evento, attrazione, repulsione o regolarità, e la presenza di qualche interazione spazio-temporale.

# *Dati*

- Le informazioni sugli incidenti sono dati di processo di punto e sono reperibili al link:  
<https://data.virginia.gov/Public-Safety/Open-View-Gun-Violence-Incidents/88rd-6gkj>
- I dati relativi alla popolazione del Texas:  
<https://demographics.texas.gov/data/decennial/2020/>
- Per le stazioni di polizia di Houston:  
<https://cohgis-my.city.opendata.arcgis.com/datasets/MyCity::coh-police-stations/about>
- Gli shapefile:  
<https://catalog.data.gov/dataset/tiger-line-shapefile-2016-state-texas-current-county-subdivision-state-based>
- Sistema di riferimento delle coordinate utilizzato: WGS 84

# *Texas*

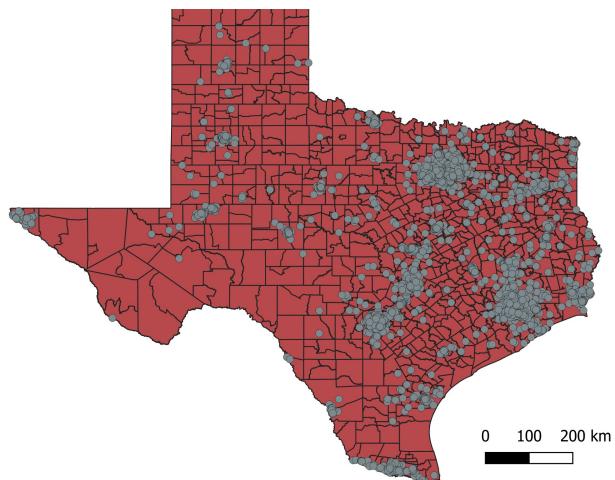
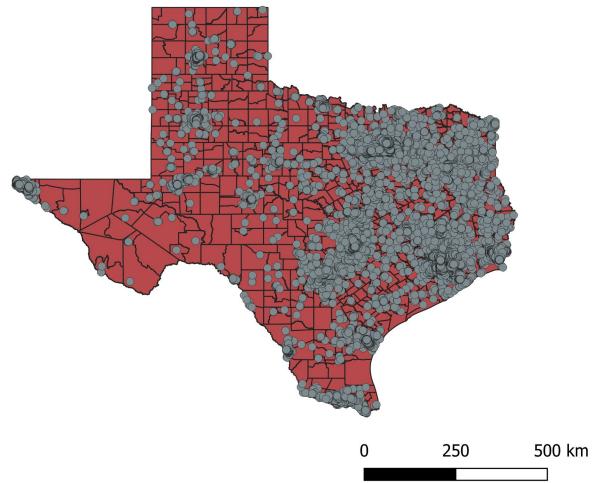
Dopo aver caricato il layer per lo stato del Texas ho modificato lo shape file degli eventi:

tabella attributi → seleziona elementi con espressione → salva layer solo elementi selezionati.

Ottenendo tutti gli incidenti da arma da fuoco dal 2014 a fine gennaio 2023 accaduti in Texas (30 557).

Successivamente ho ripetuto le modifiche selezionando solo gli incidenti relativi al 2020 (4904) ed eliminando gli errori:

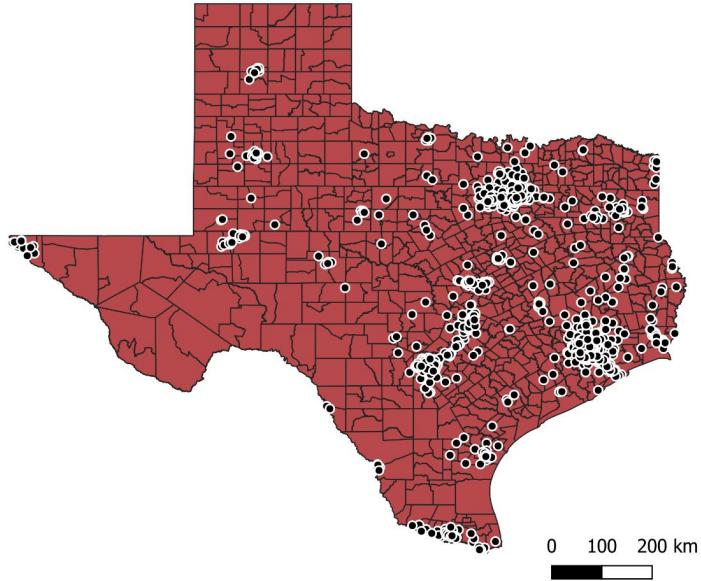
seleziona elementi con area o tocco → tabella attributi → attiva modifiche → elimina elementi selezionati.



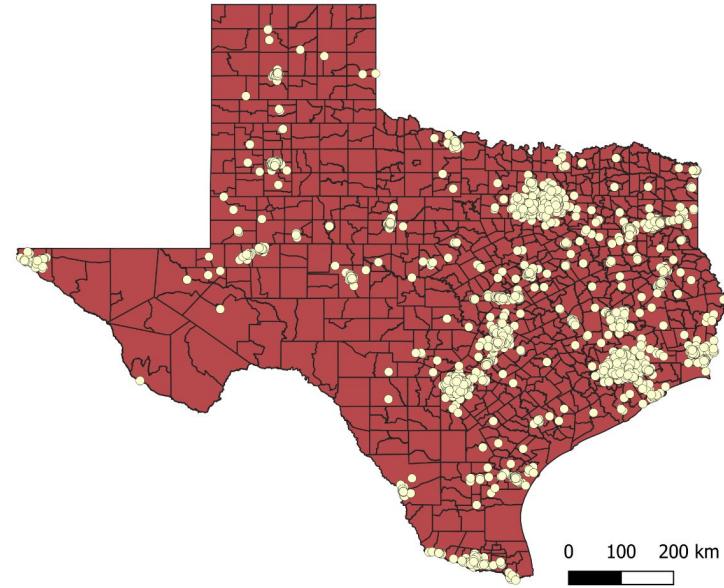
# *Texas*

Ho creato i due nuovi file shape dividendo gli eventi mortali (1615) da quelli con solo feriti o incolumi (3289).

*Morti*



*Feriti o incolumi*

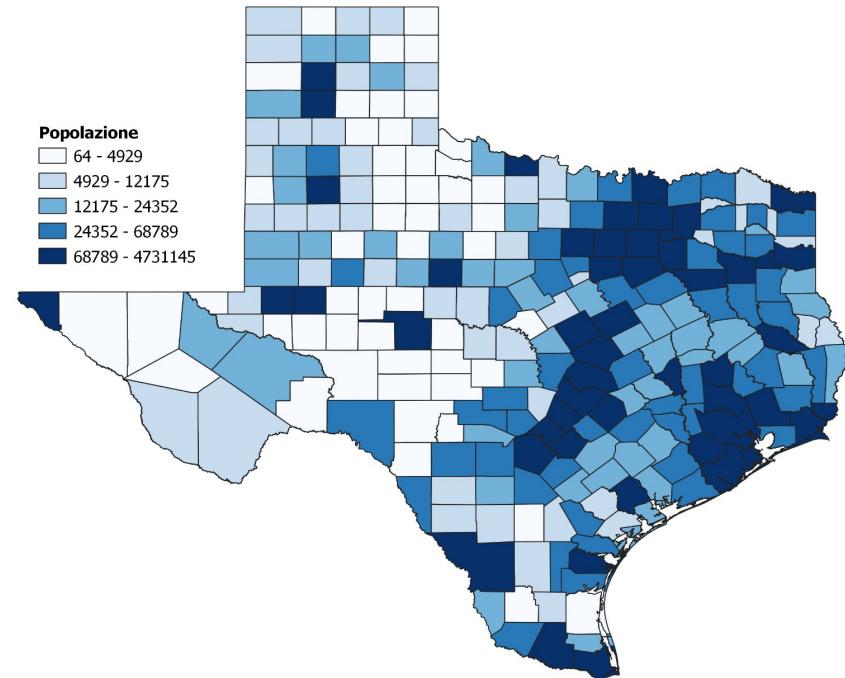


# *Texas: popolazione residente*

Importata la tabella della popolazione residente nel 2020 per ciascuna contea, tramite join l'ho unita al layer del Texas e ho colorato ogni contea:

proprietà → simbologia → graduato → modalità: quantile.

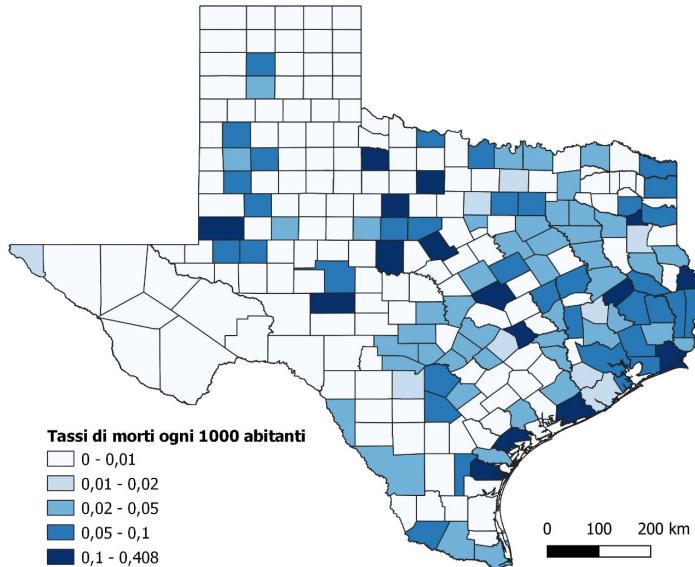
Ad Est risiede la maggior parte della popolazione e le contee di Harris, Dallas e Tarrant hanno il maggior numero di abitanti.



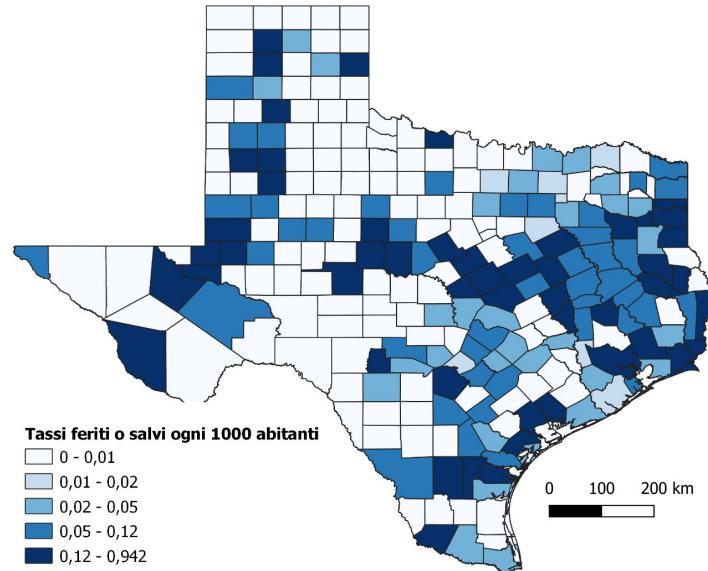
# *Texas: tassi ogni 1000 abitanti*

Calcolando il numero di eventi per ogni contea, vettore → strumenti di analisi → conta punti, posso ottenere i tassi per ogni contea attraverso il calcolatore dei campi.

*Morti*



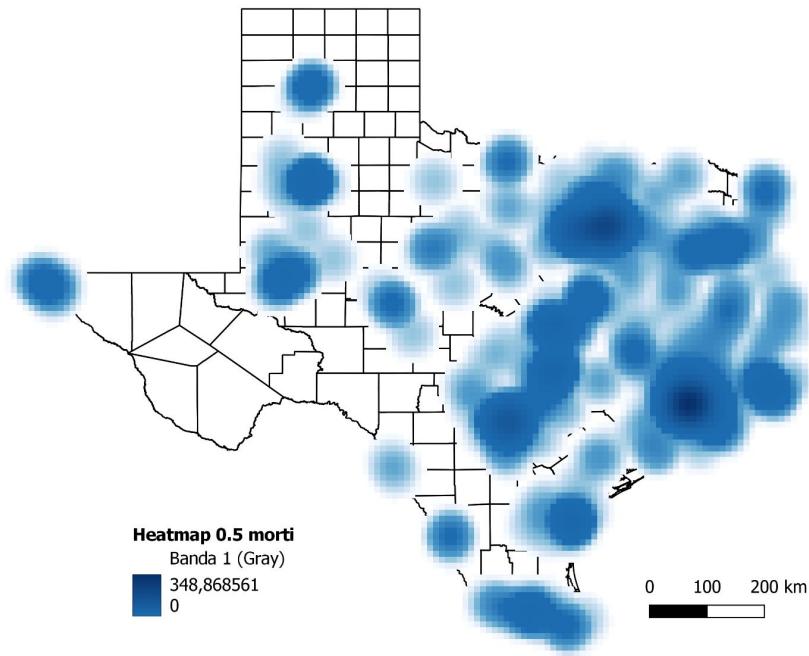
*Feriti o incolumi*



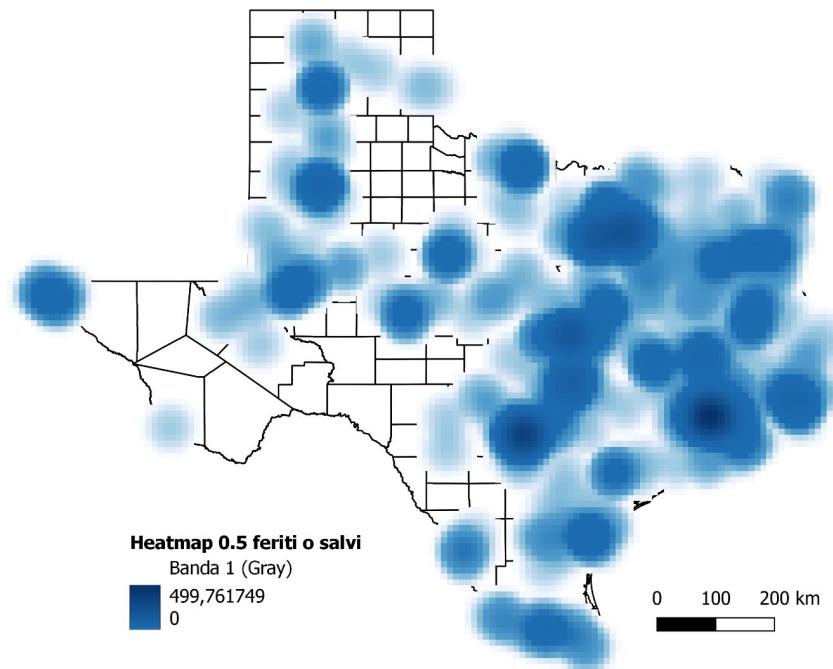
# *Texas: heatmap*

Utilizzando gli strumenti di geoprocessing ho stimato l'intensità di primo ordine:

*Morti*



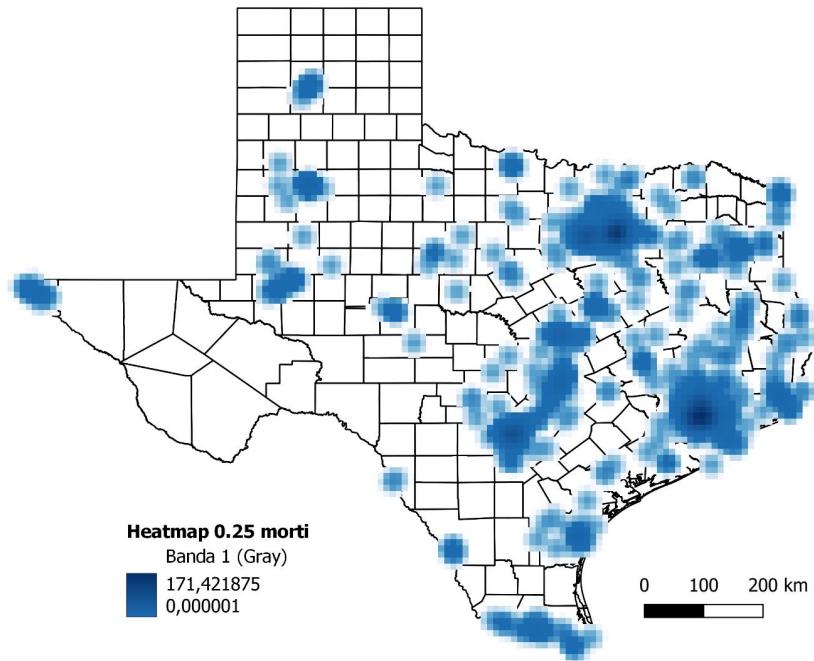
*Feriti o incolumi*



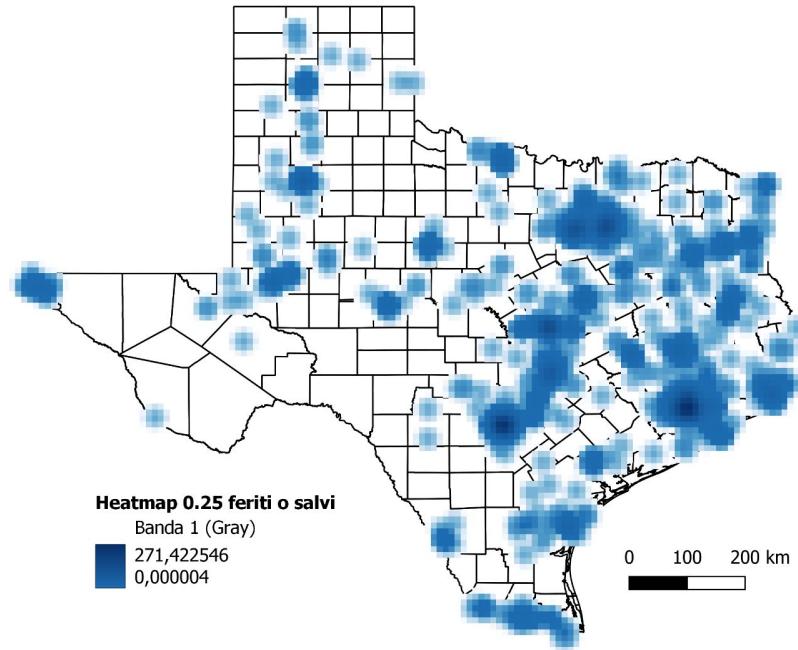
# *Texas: heatmap*

Utilizzando gli strumenti di geoprocessing ho stimato l'intensità di primo ordine:

*Morti*



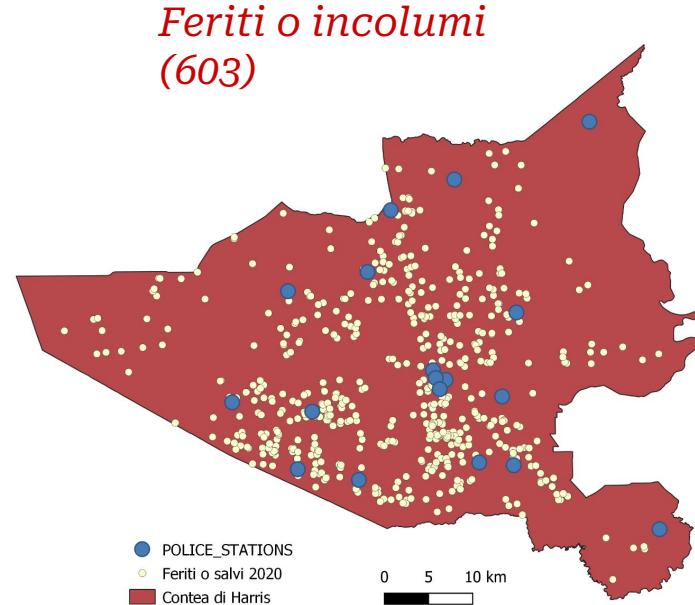
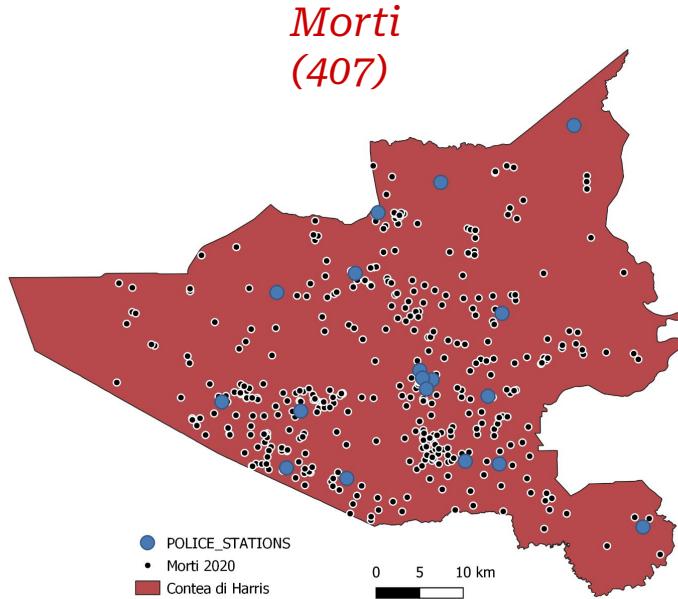
*Feriti o incolumi*



# Contea di Harris

Dopo, ho scelto di concentrarmi sulla contea di Harris, zona nella quale è presente un elevato tasso di eventi, per entrambi i tipi.

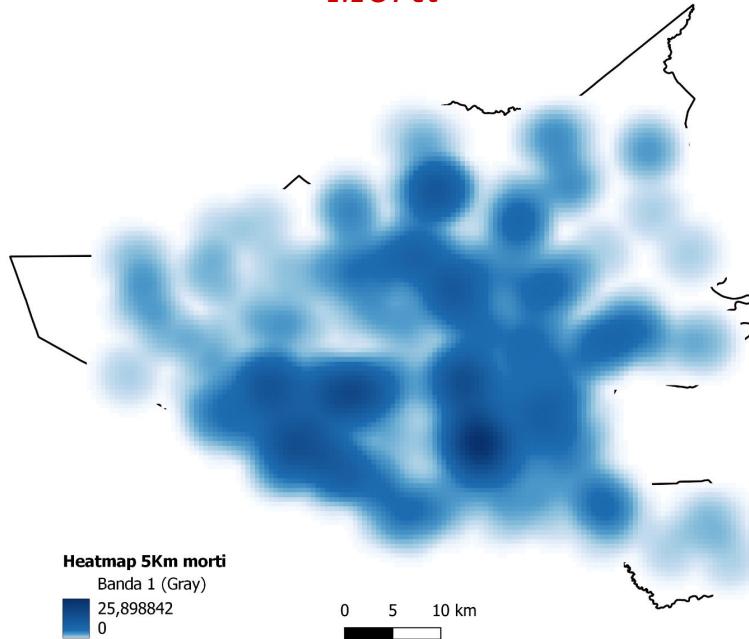
Creando due nuove file shape: seleziona per posizione → 2 shape → contiene.



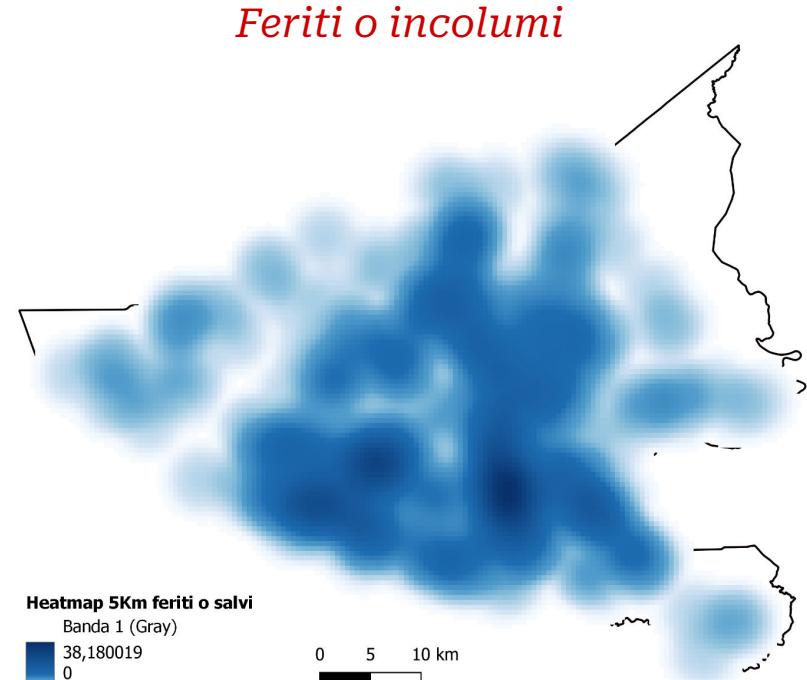
# *Contea di Harris: heatmap*

Utilizzando gli strumenti di geoprocessing ho stimato l'intensità di primo ordine:

*Morti*

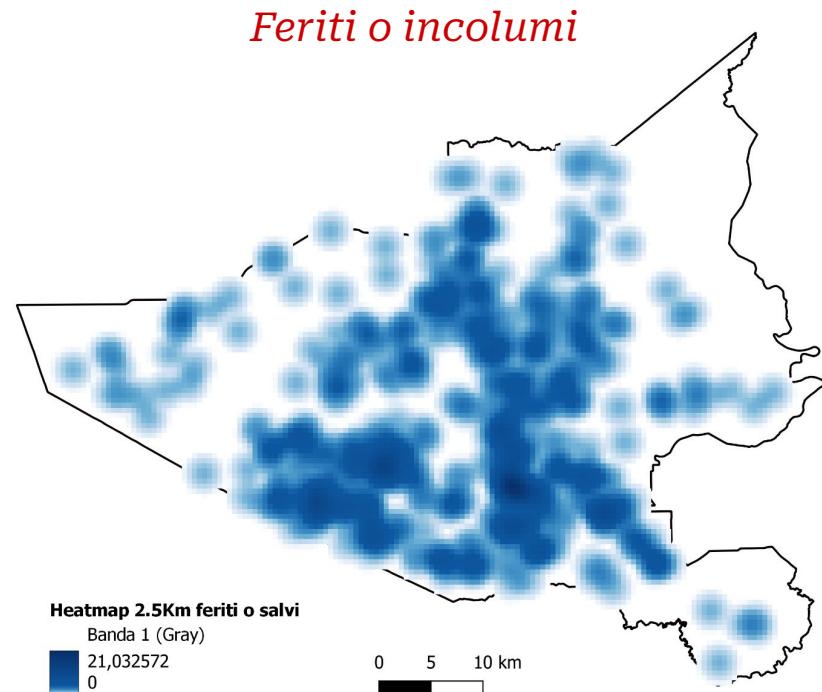
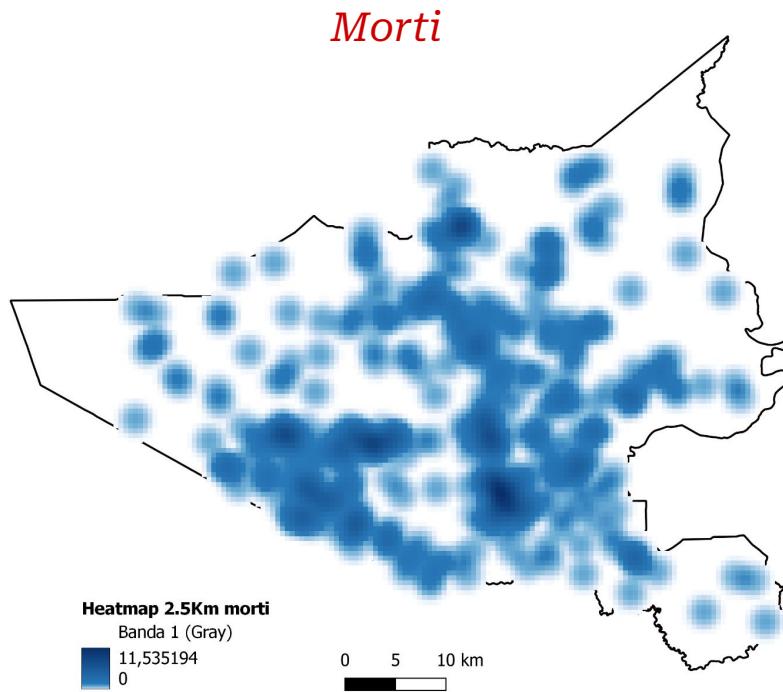


*Feriti o incolumi*



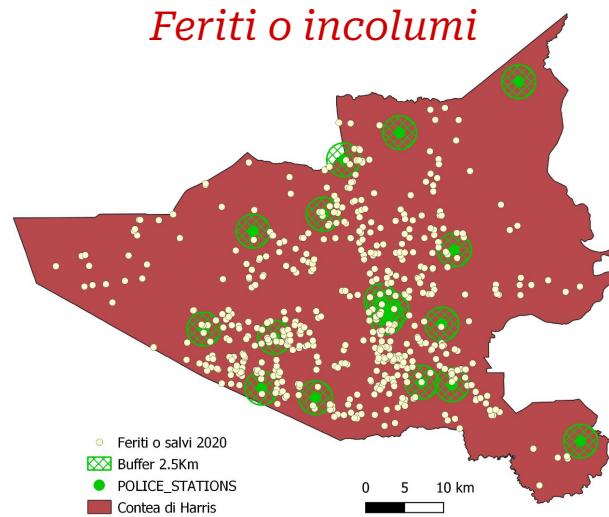
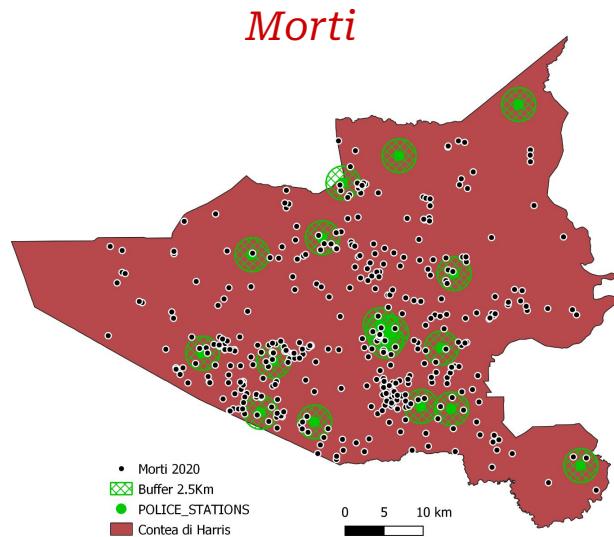
# *Contea di Harris: heatmap*

Utilizzando gli strumenti di geoprocessing ho stimato l'intensità di primo ordine:



# *Contea di Harris: buffering*

Per analizzare quanti eventi avvengono in prossimità dei dipartimenti della polizia ho utilizzato un buffer con un raggio di 2.5 chilometri.



Il numero medio di feriti o salvi entro un raggio di 2.5Km da ogni dipartimento è pari a 10.3 e il dipartimento del Midwest ne ha il maggior numero, 29. Mentre è il dipartimento del South Central ad avere il maggior numero di incidenti mortali entro 2.5Km, 18, e il numero medio di morti in ogni dipartimento è 6.5.

## *R: analisi dell'intensità del secondo ordine*

Per analizzare l'intensità del secondo ordine ho utilizzato le funzioni F, G e K.

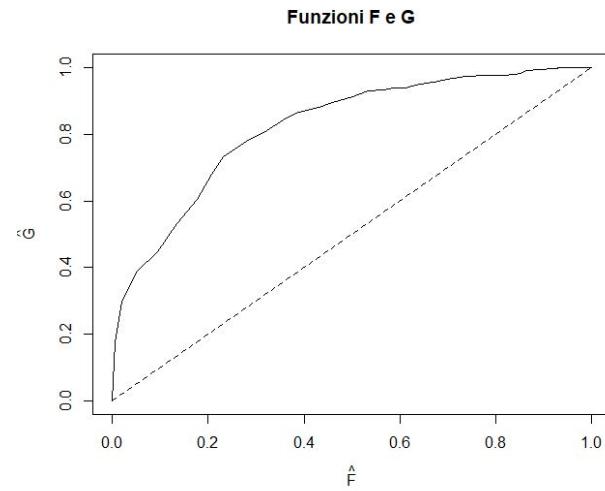
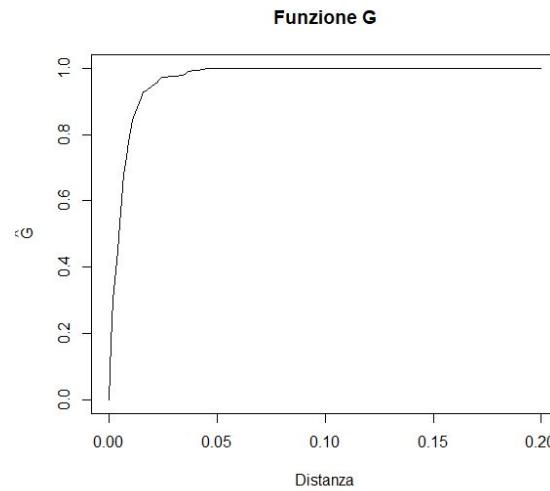
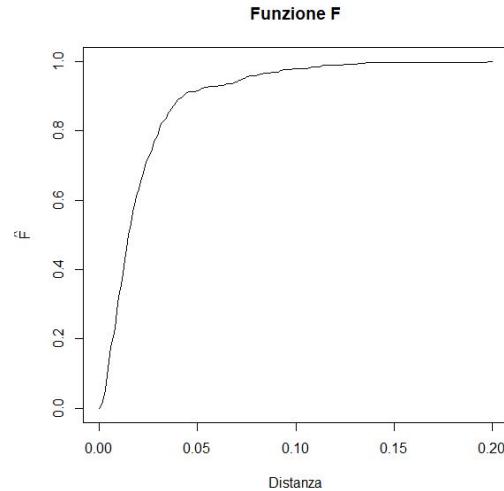
Le funzioni G(w) e F(x) si basano sulle distanze evento vicino più prossimo e la stima delle distribuzioni di ripartizione empiriche sono date da:

$$\hat{G}(w) = \frac{\#(w_i \leq w)}{n} \quad \hat{F}(x) = \frac{\#(x_i \leq x)}{m}$$

Dove:

- G(w) è la funzione di ripartizione empirica delle distanze “evento-evento” più prossimo
- F(x) è la funzione di ripartizione empirica delle distanze “evento-punto”
- n è il numero di eventi
- m è il numero di punti casuali

# *Funzioni F e G: feriti o incolumi*



Dai plots delle funzioni empiriche si osserva che la funzione G cresce più velocemente rispetto alla funzione F. Tale risultato viene confermato dal grafico della funzione F contro la funzione G.

Dunque le distanze “punto-evento” sono maggiori delle distanze “evento-evento”, indicando la presenza di cluster.

# *Funzione K*

Le funzioni F(x) ed G(w) hanno il difetto di considerare solo le distanze da eventi vicini, dando informazioni solo sull'andamento in piccola scala del fenomeno.

La funzione K(.) fornisce una informazione sulla struttura spaziale su ampia scala, utilizza tutte le distanze evento-evento, ed ha una corrispondenza biunivoca con l'intensità del 2° ordine. È definita come:

$$K(h) = \frac{1}{\lambda} E[\# \text{di eventi entro una distanza } h \text{ da un evento arbitrario}]$$

E lo stimatore è:

$$\hat{K}(h) = \frac{1}{\lambda^2 |R|} \sum_i \sum_{j \neq i}^n \frac{I_h(d_{ij})}{w_{ij}}$$

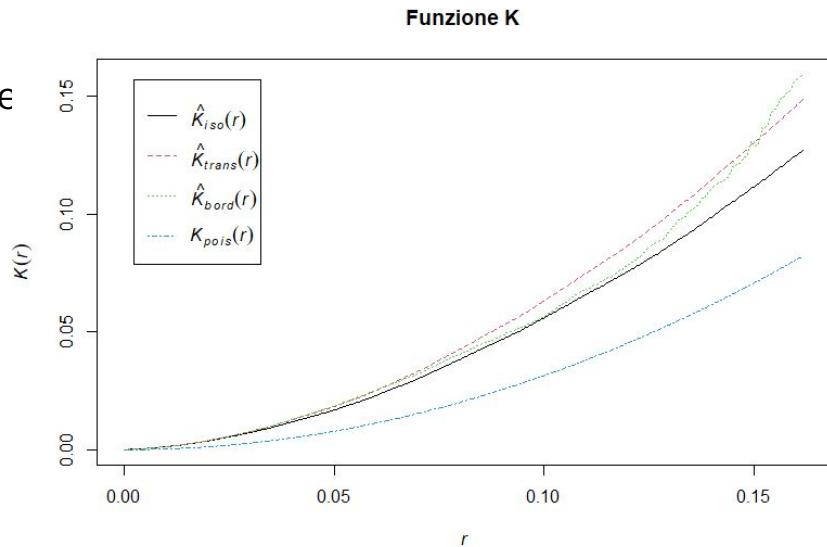
# *Funzione K e test Clark-Evans: feriti o incolumi*

La funzione K conferma la presenza di cluster:  
le funzioni stimate sono maggiori della funzione  
K sotto CSR, ovvero  $\pi h^2$ .

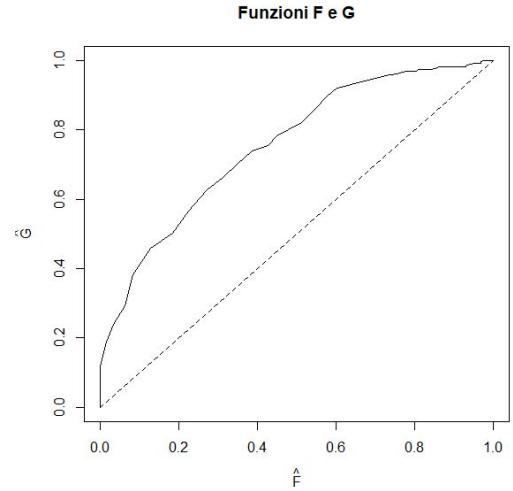
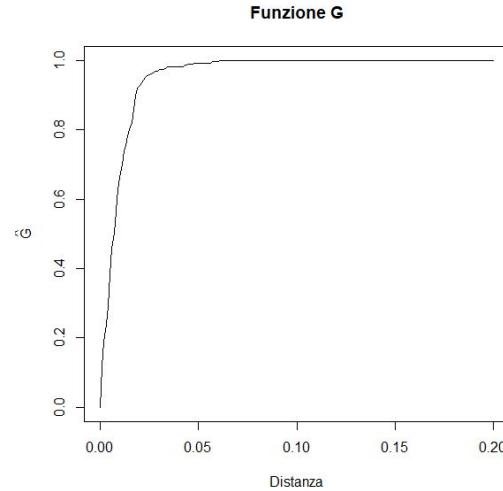
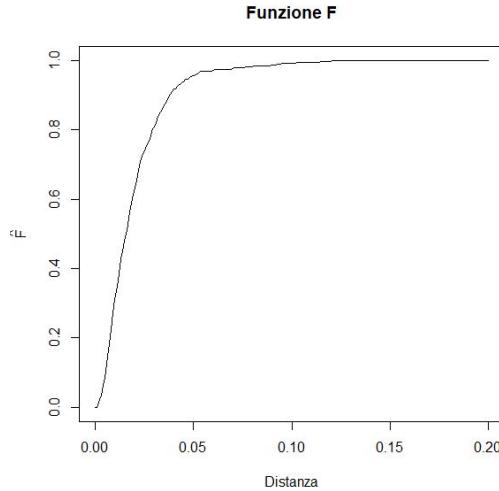
## **Test di Clark-Evans:**

R = 0.64427, p-value < 2.2e-16

alternative hypothesis: clustered (R < 1)



# *Funzioni F e G: morti*



Dai plots delle funzioni empiriche si osserva che la funzione G cresce più velocemente rispetto alla funzione F. Tale risultato viene confermato dal grafico della funzione F contro la funzione G. Indicando dunque cluster anche per gli incidenti mortali.

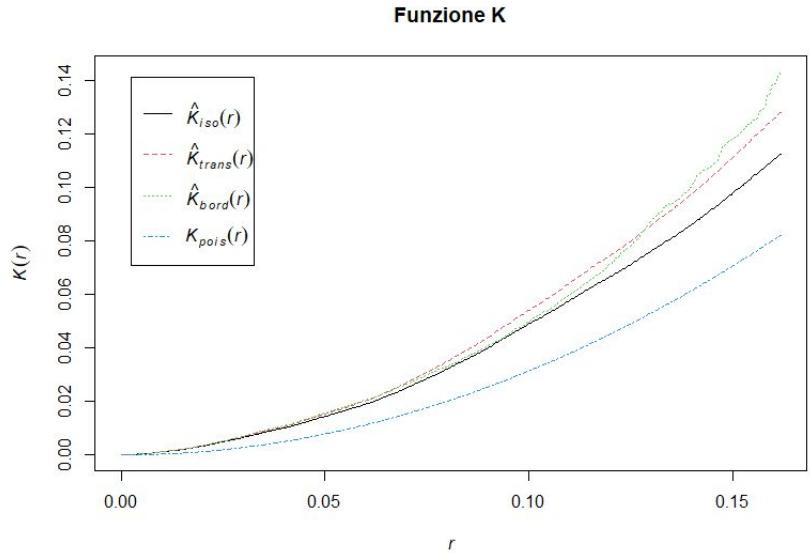
# *Funzione K e test Clark-Evans: morti*

La funzione K conferma la presenza di cluster anche per gli eventi mortali: le funzioni stimate sono maggiori della funzione K sotto CSR, ovvero  $\pi h^2$ .

## **Test di Clark-Evans:**

R = 0.7421, p-value < 2.2e-16

alternative hypothesis: clustered (R < 1)



## *Cross K function*

Per verificare la presenza di qualche legame tra i due tipi di evento, mortali e non, ad esempio se sono indipendenti, si attraggono o sono repulsivi ho utilizzato la Cross K function, definita come:

$$K_{ij}(h) = \frac{1}{\lambda_j} E[\# \text{ di eventi del tipo } j \text{ entro una distanza } \leq h \text{ da un evento del tipo } i \text{ arbitrario}]$$

Sotto l'ipotesi di indipendenza tra gli eventi abbiamo:

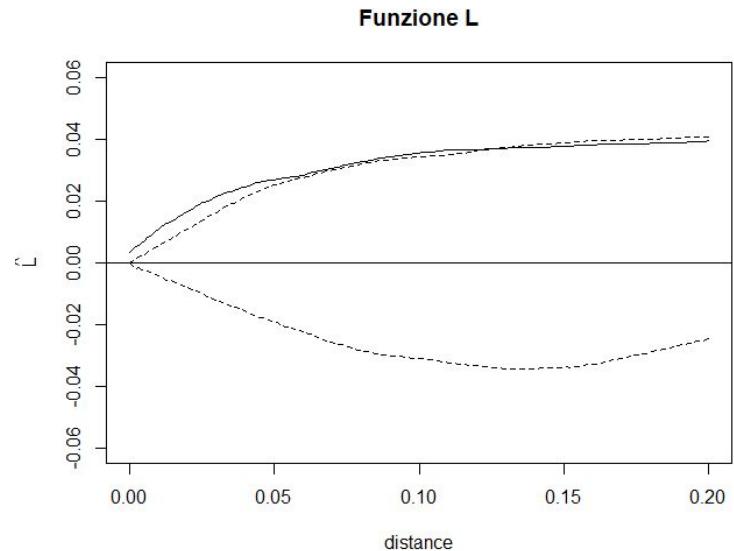
$$K_{ij}(h) = \pi h^2$$

possiamo anche utilizzare la trasformata:

$$\hat{L}_{ij}(h) = \sqrt{\frac{\hat{K}_{ij}(h)}{\pi}} - h$$

# *Cross K function*

Tra i due eventi sembra esserci una leggera attrazione fino ai 5km circa, ma a distanze maggiori la funzione è molto simile alla banda superiore ottenuta tramite shift toroidale e dunque osserviamo indipendenza tra essi.



## *Analisi spazio-tempo*

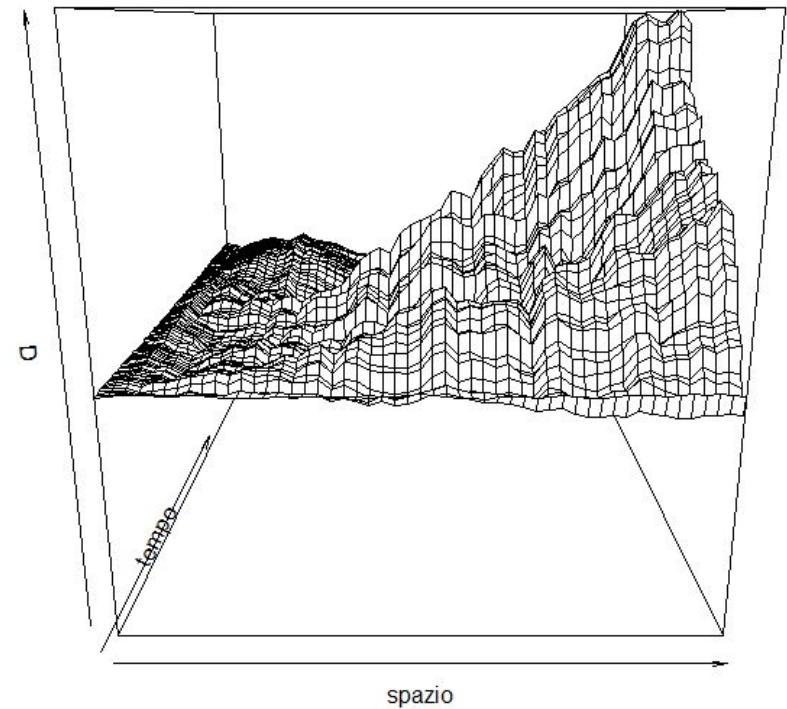
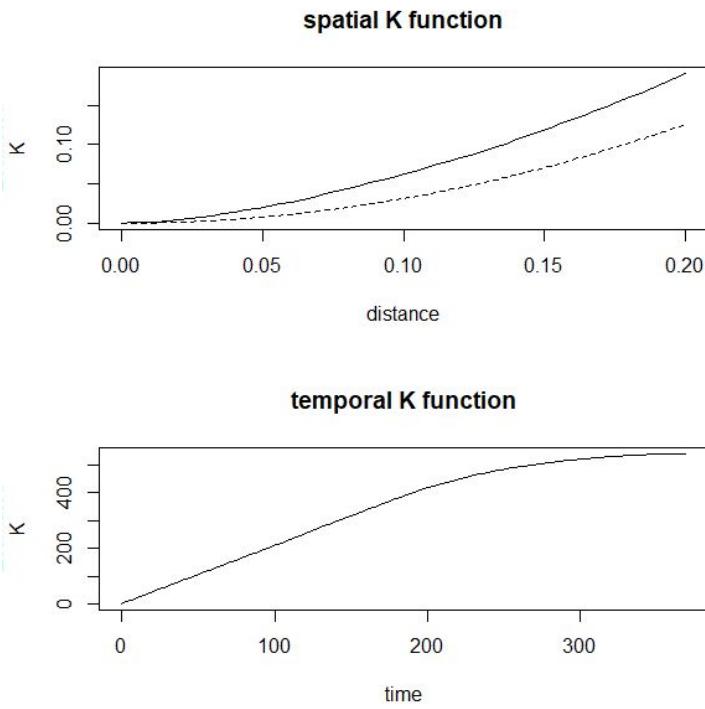
Infine con la funzione  $K(h,t)$  ho verificato la presenza di interazione spazio-tempo.

Il test può essere basato sulle differenze osservate tra le funzioni:

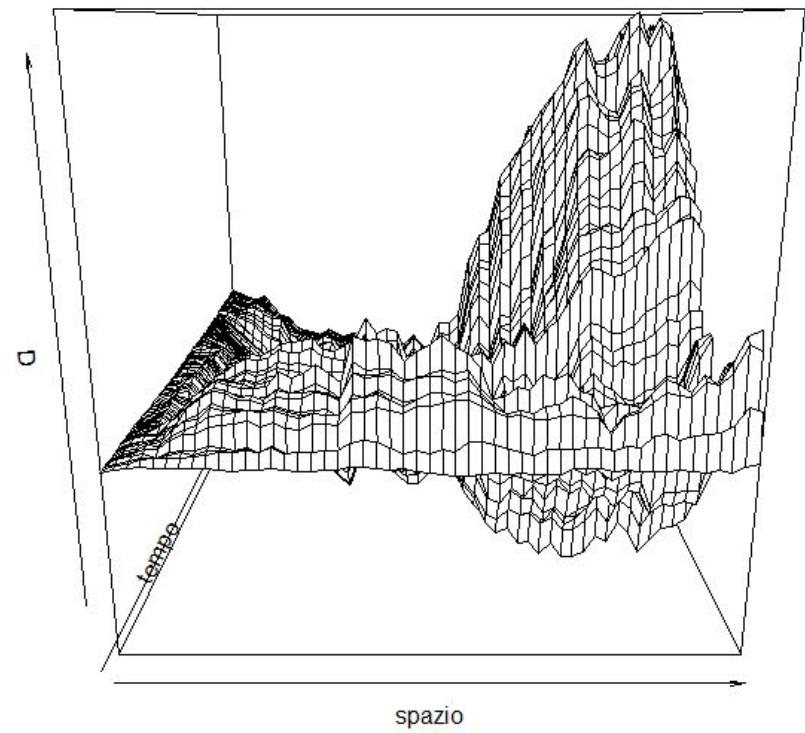
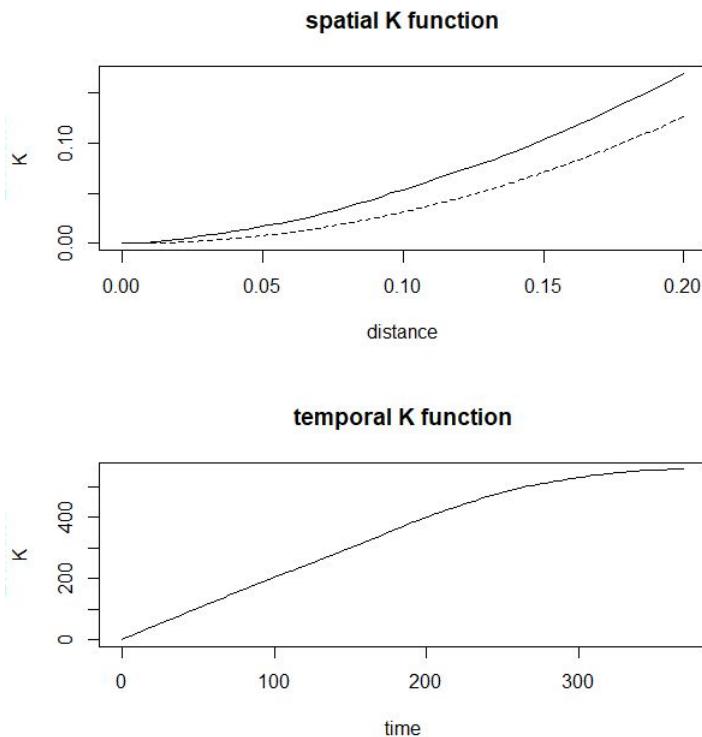
$$\hat{D}(h, t) = \hat{K}(h, t) - \hat{K}_S(h)\hat{K}_T(t)$$

Se i processi che operano nel tempo e nello spazio sono indipendenti (nessuna interazione spazio-temporale), allora  $K(h, t)$  dovrebbe essere solo il prodotto di due funzioni  $K$ , una nello spazio,  $K_S(h)$ , e l'altra nel tempo,  $K_T(t)$ .

# *Analisi spazio-tempo: feriti o incolumi*



# *Analisi spazio-tempo: morti*



## *Conclusioni*

Attraverso questa analisi abbiamo osservato la presenza di cluster per gli incidenti d'arma da fuoco in Texas e in particolare nella contea di Harris: il maggior numero di eventi si verificano in prossimità della capitale, Houston.

C'è una leggera attrazione tra gli incidenti mortali e gli incidenti con solo feriti o salvi, entro una distanza di 5km.

E infine abbiamo osservato la presenza di interazione spazio-tempo in entrambi i tipi di eventi.

*Grazie per l'attenzione*