



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Analisi Spaziale degli Omicidi nel Mississippi negli anni '90

Progetto di Statistics for Spatial data

a cura di Jacopo Manetti

Introduzione

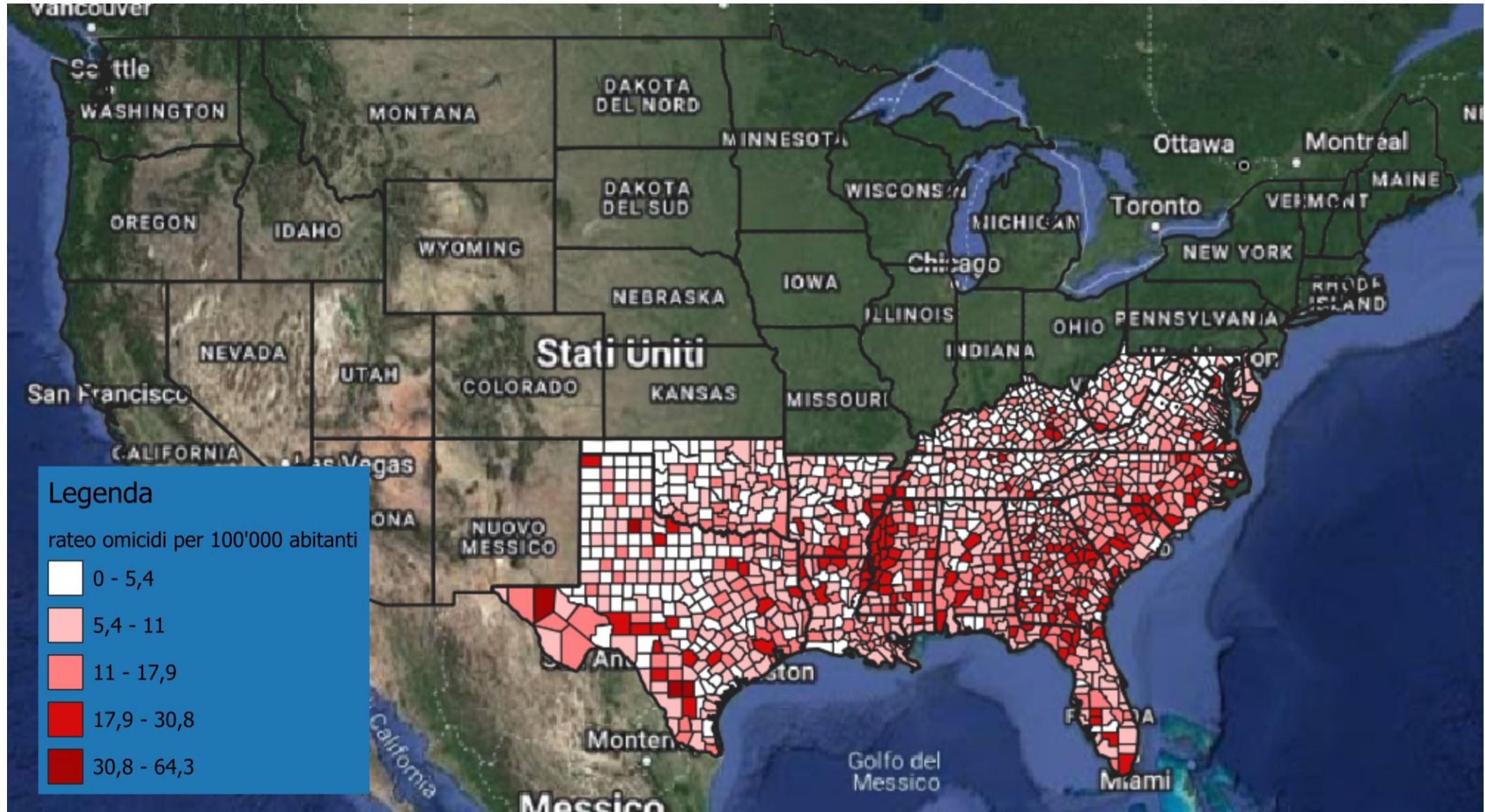
Il presente studio si inserisce nel contesto dell'analisi spaziale degli omicidi negli stati del sud degli Stati Uniti negli anni '90. Inizialmente, si è posta l'attenzione sulla scelta di un contesto geografico specifico per l'analisi. La decisione di concentrarsi su un singolo stato è stata guidata dalla necessità di preservare la ricchezza dei dettagli e delle informazioni che potevano andare perdute in un'analisi più estesa che coinvolgesse più stati. Inoltre, un'analisi che includesse tutte le contee di tutti gli stati del sud si è rivelata troppo ampia e dispersiva per gli scopi di questo studio. Pertanto, si è optato per la focalizzazione su un singolo stato, selezionato in base ai risultati di una fase esplorativa preliminare che ha evidenziato modelli e tendenze particolarmente significative e interessanti.

Step seguiti:

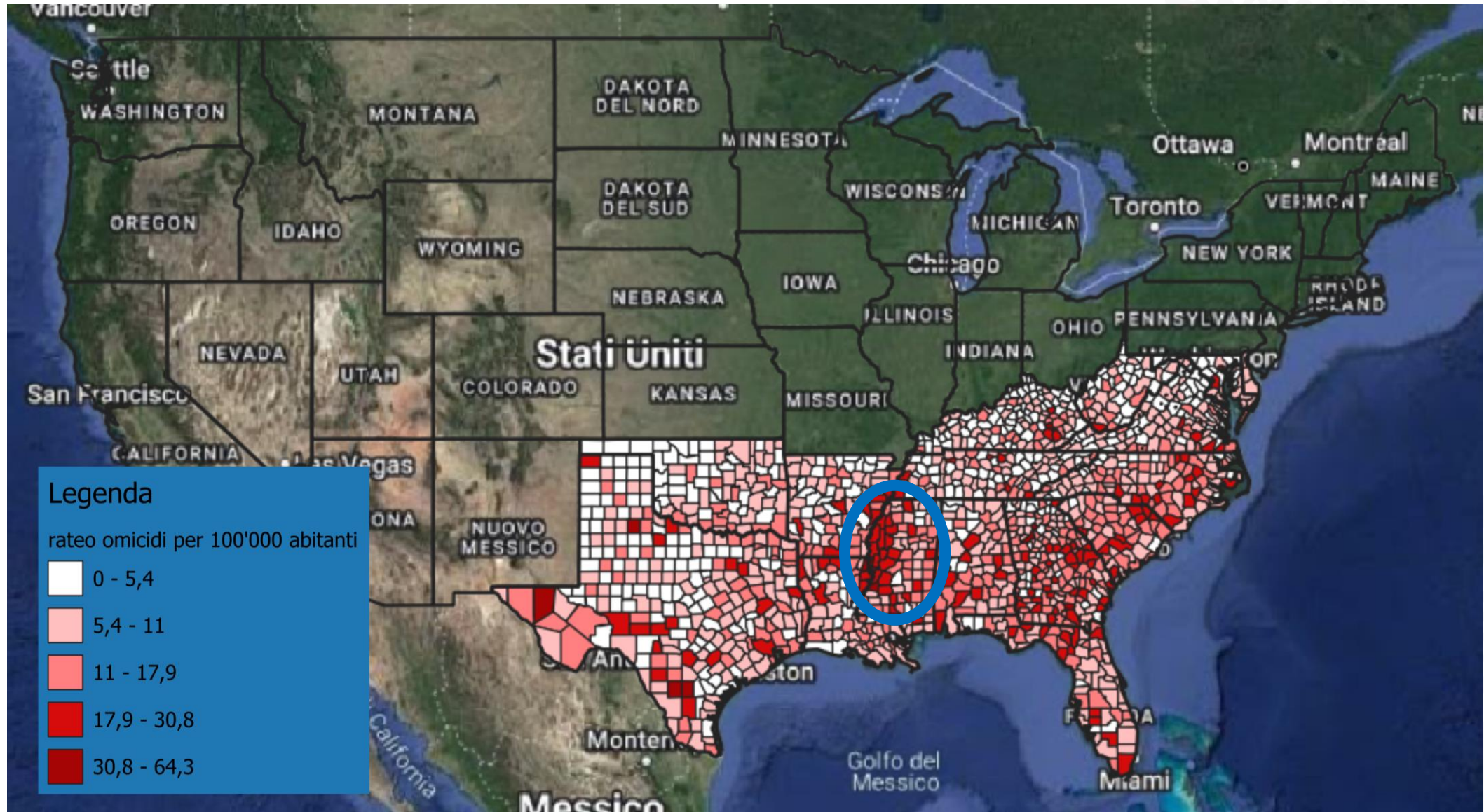
Gli step principali del progetto sono:

- **Scelta di uno Stato per l'Analisi Approfondita:** Determinare uno stato del sud degli USA come punto focale per l'analisi degli omicidi, basandosi sui dati esplorativi e sui modelli criminali osservati.
- **Valutazione dell'Influenza dei Fattori Socio-Economici:** Analizzare come le variabili socio-economiche a disposizione possano essere correlate con i tassi di omicidio.
- **Sviluppo e Implementazione di Modelli Statistici:** Utilizzare modelli lineari per i dati e valutare l'autocorrelazione spaziale dei residui tramite il test di Moran, per poi implementare un modello statistico spaziale appropriato.

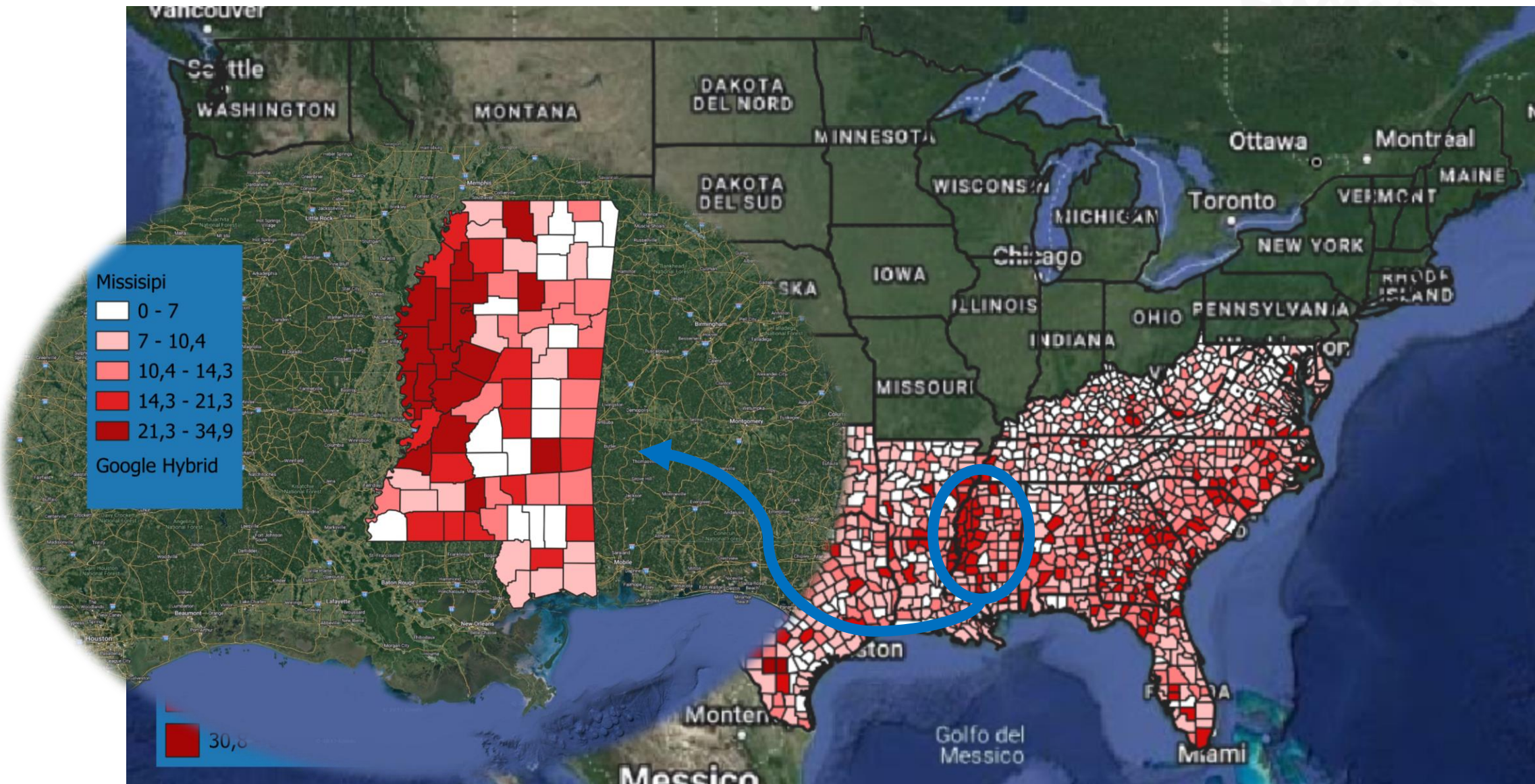
Variabile di Studio



Variabile di Studio



Variabile di Studio



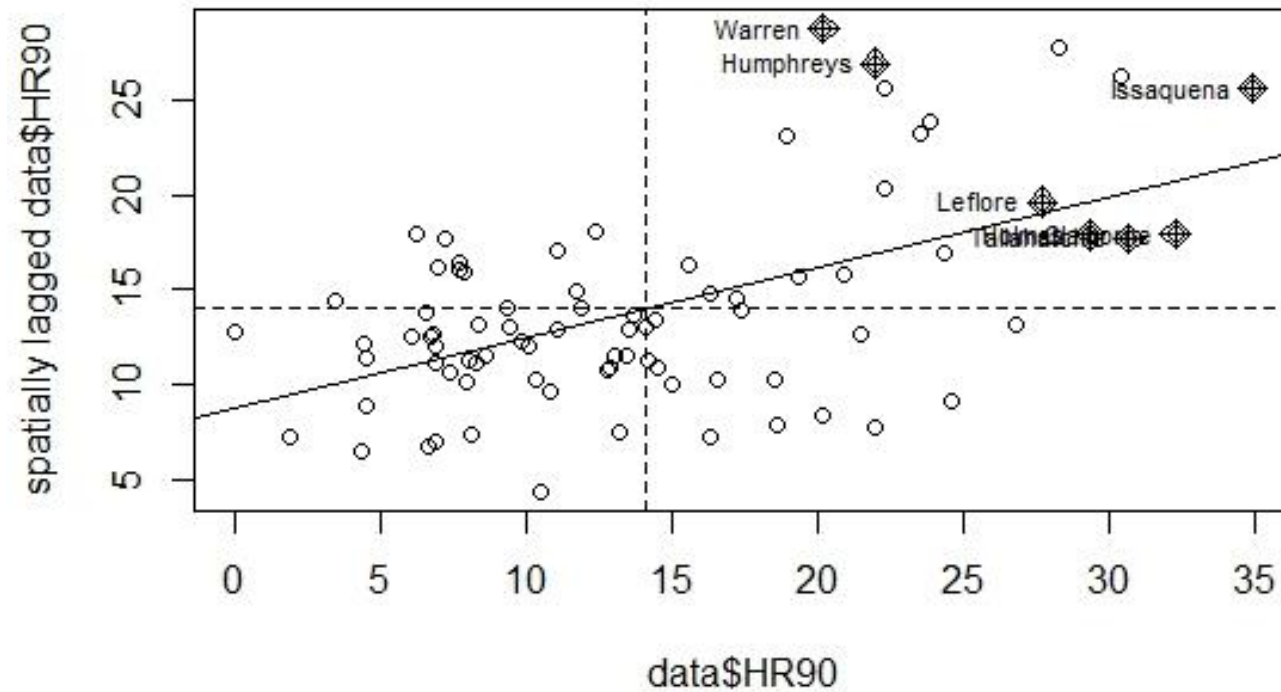
I dati

Per il nostro studio avevamo a disposizione diverse variabili socio-economiche legate agli omicidi avvenuti nelle contee dello stato americano del Mississippi negli anni '90.

Escludendo le variabili poco rilevanti come quelle utili al calcolo del rateo di omicidi e le PCA otteniamo le seguenti variabili.

variabile	descrizione	media	mediana	Dev. standard
HR90	Rateo degli omicidi per 100'000 abitanti	14,05	12,81	7,87
DV90	Percentuale di uomini divorziati sopra i 15 anni	6,63	6,70	0,94
BLK90	Percentuale di neri	38,39	36,96	19,07
FP89	Percentuale di famiglie sotto la soglia di povertà	23,75	22,53	8,45

Global Moran



Moran	P-value
0,369	7.904e-09

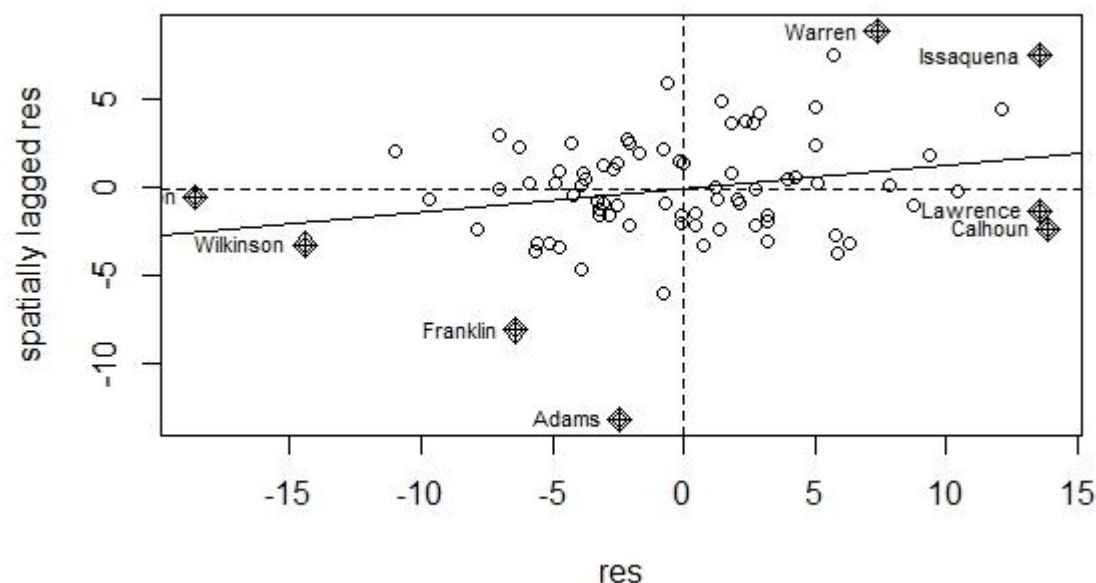
Come ci si aspettava il calcolo del moran individua degli hotspot nella zona ovest del mississippi

Map of Mississippi Counties

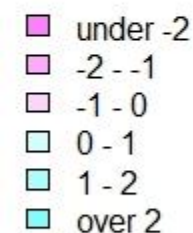


Il modello

In questa fase dell'analisi, viene costruito un modello di regressione lineare per esplorare l'associazione tra il tasso di omicidio e le variabili socio-economiche. Dopo la stima del modello, i residui (differenze tra i valori osservati e quelli stimati) vengono esaminati per individuare pattern spaziali non catturati dal modello. Si effettua quindi un test di Moran sui residui, per valutare l'autocorrelazione spaziale residua, seguito da una visualizzazione grafica che aiuta a identificare aree geografiche con comportamenti spaziali anomali.



Regression residuals



Moran	P-value
0.134	0.0145

Usare diverse matrici delle distanze (calcolo sui residui)

Matrice W	Moran I	Dev. Standard	P-value
Contiguità del primo ordine	0.134	2.183	0.015
Contiguità del secondo ordine	0.056	1.377	0.084
Knn = 2	0.092	1.043	0.148
Knn = 4	0.146	2.166	0.015
Knn = 6	0.110	2.100	0.018

Nei modelli di regressione spaziale, la matrice W dei pesi ha una rilevanza fondamentale, quindi ho considerato diverse possibilità scartando quelle non significative.

Modello Spaziale

Si è proceduto alla modellazione spaziale utilizzando vari modelli che incorporano effetti spaziali. Sono stati esplorati diversi tipi di modelli, tra cui Spatial error model, Spatial lag model e Spatial Durbin model. Ma solo il Spatial Lag model è risultato significativo.

```
> #spatial lag model
> m2=lagsarlm(HR90 ~ BLK90 + DV90 + FP89, data=data, listw=nb2listw(data.nb))
> summary(m2)
```

```
Call:lagsarlm(formula = HR90 ~ BLK90 + DV90 + FP89, data = data, listw = nb2listw(data.nb))
```

Residuals:

Min	1Q	Median	3Q	Max
-17.17747	-3.80961	-0.34186	3.13363	14.73202

Type: lag

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.789971	6.381975	0.4372	0.6620
BLK90	0.151863	0.067567	2.2476	0.0246
DV90	-0.345772	0.755192	-0.4579	0.6471
FP89	0.161930	0.148813	1.0881	0.2765

Rho: 0.27795, LR test value: 4.8115, p-value: 0.028271

Asymptotic standard error: 0.12823

z-value: 2.1677, p-value: 0.030185

wald statistic: 4.6987, p-value: 0.030185

Log likelihood: -260.0695 for lag model

ML residual variance (sigma squared): 32.757, (sigma: 5.7234)

Number of observations: 82

Number of parameters estimated: 6

AIC: 532.14, (AIC for lm: 534.95)

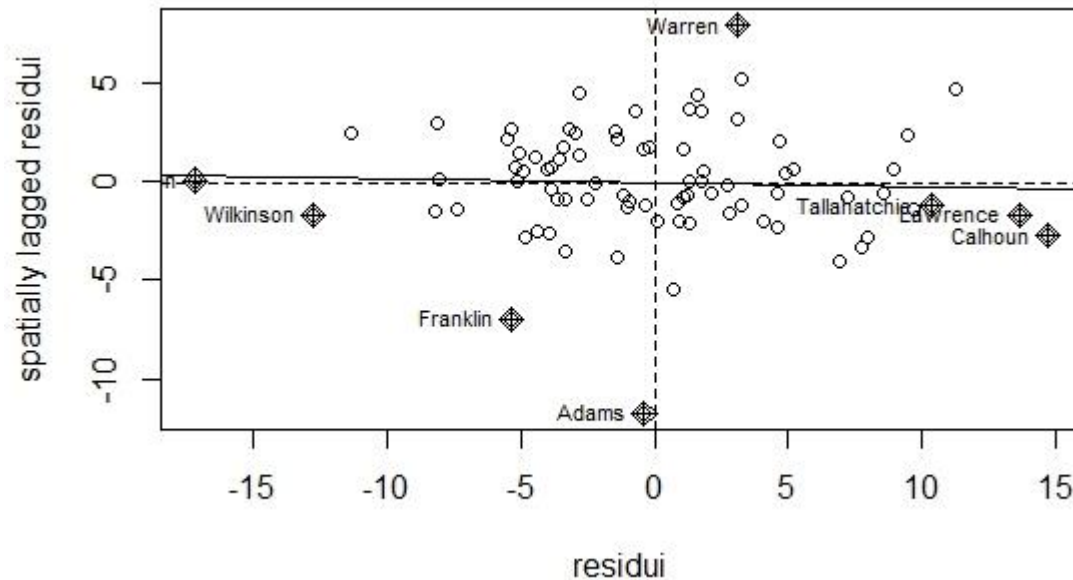
LM test for residual autocorrelation

test value: 0.29282, p-value: 0.58842

Criteri di bontà di adattamento per la selezione della matrice W nel modello SLM

Matrice W	AIC	Log-likelihood
Contiguità del primo ordine	532.14	-260.0695
Knn = 4	532.23	-260.1142
Knn = 6	532.99	-260.4931

Risultati



moran	P-value
-0.0207	0.5497

Il Moran scatterplot quasi piatto e non significativo per i residui indica che il modello ha efficacemente catturato la componente spaziale dei dati. Non ci sono evidenze di autocorrelazione spaziale residua, il che suggerisce che il modello di lag spaziale è appropriato per questi dati.

Conclusioni

Dalle analisi effettuate e dai risultati ottenuti dal modello di lag spaziale (SLM), possiamo trarre le seguenti conclusioni:

- **Significatività di BLK90:** Il coefficiente per BLK90 è statisticamente significativo e positivo, ma il suo valore suggerisce un impatto moderato su HR90 (tasso di omicidio) ma non preponderante.
- **Non Significatività di DV90 e FP89:** Le variabili DV90 e FP89 non mostrano un impatto statisticamente significativo su HR90 nel modello, indicando che, in questo contesto, non possono essere considerate determinanti chiave dei tassi di omicidio.
- **Rilevanza dell'Autocorrelazione Spaziale (Rho):** Il parametro Rho è significativo, confermando l'importanza dell'autocorrelazione spaziale nel fenomeno in esame. Questo suggerisce che i tassi di omicidio in una data area sono influenzati dai tassi nelle aree vicine.
- **Conclusioni Generali sul Modello:** Il modello di lag spaziale appare adeguato per analizzare i tassi di omicidio nel contesto studiato, con particolare enfasi sull'influenza della percentuale di popolazione nera e sull'importanza della prossimità spaziale tra le aree.



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Grazie per l'attenzione!