

Nama : Yudhapratama Nugraha Aryaputra

NIM : 24021302

Homework Spatial Regression

1. Menampilkan Peta East-Coast.

Dengan menggunakan command :

```
subset(nat_east, EAST=="1")
nat_east_only<-subset(nat_east, EAST=="1")
nat_east_only_data <- data.frame(nat_east_only)
plot(nat_east_only, main = "East Coast", axes = TRUE)
summary(nat_east_only)
summary(nat_east_only_data)
head(nat_east_only_data)
```



Gambar 1. Visualisasi Peta County-County Pantai Timur (East-Coast)

Visualisasi 5 data pertama summary statistics :

```
> head(nat_east_only_data)
  NAME STATE_NAME STATE_FIPS CNTY_FIPS  FIPS STFIPS COFIPS FIPSNO SOUTH  HR60  HR70  HR80  HR90
97  Aroostook     Maine      23      003 23003    23     3  23003    0 2.514205 1.0629478 2.554810 1.1502715
159 Somerset     Maine      23      025 23025    23    25  23025    0 0.000000 0.8210787 5.922241 0.6697879
160 Piscataquis  Maine      23      021 23021    23    21  23021    0 1.918024 4.0937468 7.561151 5.3610679
168 Penobscot    Maine      23      019 23019    23    19  23019    0 1.319129 2.6583089 3.892518 2.0463708
228 Washington   Maine      23      029 23029    23    29  23029    0 6.077550 6.6981480 2.860166 4.7203655
229 Franklin     Maine      23      007 23007    23     7  23007    0 1.660936 1.4851779 2.460206 3.4473249
```

2. Membuat bobot spasial dengan metode k-nearest neighbors dengan nilai $k=5$ & 10 . Pastikan bahwa matriks neighbors adalah symmetric. Tampilkan peta matrik tetangga baik untuk yang $k=5$ dan $k=10$.



Gambar 2. Peta Matriks untuk $k = 5$



Gambar 3. Peta Matriks untuk $k = 10$

3. Dengan menggunakan $knn=10$, hitung statistik Moran's I untuk HR60, HR70, HR80, dan HR90. Diskusikan hasil dari perhitungan statistik Moran's I!

Perhitungan Morans'I Angka Pembunuhan di Pantai Timur Tahun 1960

Moran I test under normality

data: nat_east_only\$HR60
weights: nat_east_nnb10_wl

Moran I statistic standard deviate = 20.006, p-value < 2.2e-16
alternative hypothesis: two.sided

sample estimates:

Moran I statistic	Expectation	Variance
0.333206938	-0.001536098	0.000279974

Perhitungan Morans'I Angka Pembunuhan di Pantai Timur Tahun 1970

Moran I test under normality

data: nat_east_only\$HR70
weights: nat_east_nnb10_wl

Moran I statistic standard deviate = 21.71, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:

Moran I statistic	Expectation	Variance
0.361720794	-0.001536098	0.000279974

Perhitungan Morans'I Angka Pembunuhan di Pantai Timur Tahun 1980

Moran I test under normality

data: nat_east_only\$HR80
weights: nat_east_nnb10_wl

Moran I statistic standard deviate = 21.971, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:

Moran I statistic	Expectation	Variance
0.366097496	-0.001536098	0.000279974

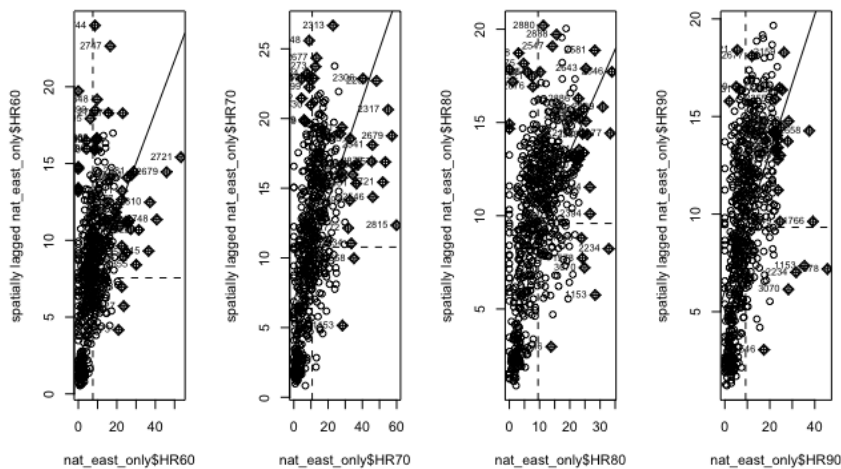
Perhitungan Morans'I Angka Pembunuhan di Pantai Timur Tahun 1990

Moran I test under normality

data: nat_east_only\$HR90
weights: nat_east_nnb10_wl

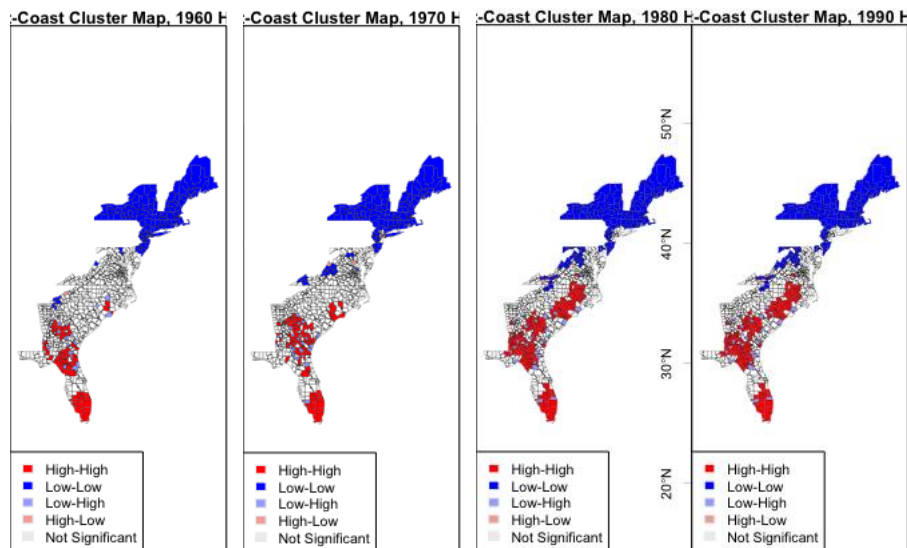
Moran I statistic standard deviate = 21.254, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:

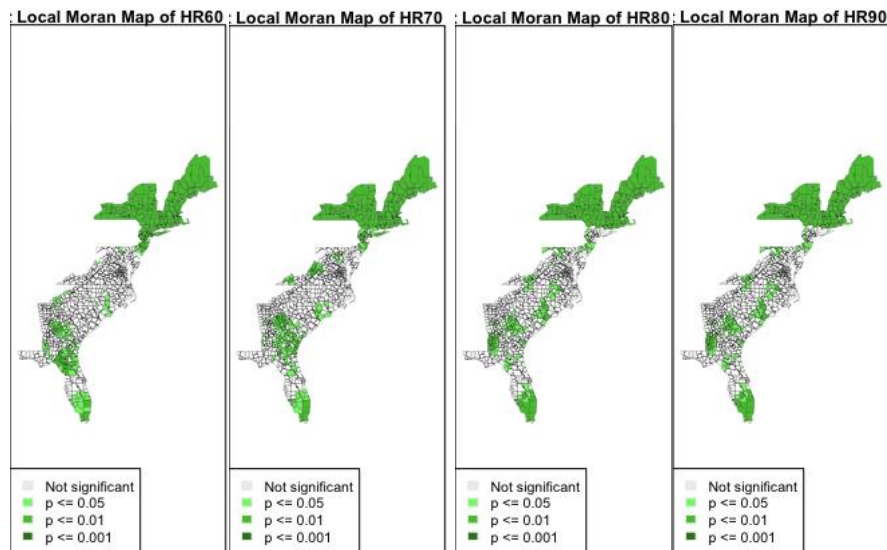
Moran I statistic	Expectation	Variance
0.354092697	-0.001536098	0.000279974



Hasil Nilai Moran's I pada Tahun 1960 adalah 0.33. Pada tahun setelahnya, Nilai Moran's I cenderung tidak berubah. Nilai Moran's I pada tahun 1970, 1980 dan 1990 adalah 0.36, 0.35 dan 0.36. Artinya, dari angka tersebut disimpulkan bahwa 35% dari variasi data mengindikasikan suatu korelasi. Jadi, dapat diartikan terjadi *Spatial Correlation*.

- Hitung Local Moran's I untuk HR60, HR70, HR80, dan HR90. Apakah selama 4 dekade terjadi perubahan pola clusterisasi? Diskusikan!





Berdasarkan hasil plot berdasarkan kluster LISA selama 4 dekade, tidak ada perubahan kluster yang signifikan. Cluster Low-Low ada di Bagian Utara East-Coast sedangkan untuk High-High ada di bagian selatan. Selain itu, terjadi penambahan cluster High-High di Sebagian wilayah selatan East-Coast dari dekade ke dekade.

5. Dengan menggunakan dataframe *nat* lakukan estimasi model OLS untuk HR60, HR70, HR80, dan HR90. Variabel independent yang digunakan sama dengan model OLS yang digunakan oleh Baller et al, tanpa menggunakan dummy variabel regions (SOUTH atau EAST). Diskusikan hasilnya!

Hasil Regresi OLS, pada tahun 1960, data yang dihasilkan menunjukkan bahwa semua signifikan. Semakin tinggi angka Resource Deprivation maka akan meningkatkan tingkat pembunuhan sebesar 2.4 Poin. Sedangkan semakin meningkatnya usia berbanding terbalik dengan tingkat pembunuhan. Semakin bertambahnya Perceraian semakin bertambah pembunuhan. Data selama 4 Dekade menunjukkan Tren ini tidak berubah. Hanya yang menjadi pertanyaan adalah angka pengangguran yang meningkat maka dampaknya angka pembunuhan berkurang. Hal ini akan dikaji lebih lanjut. (Hasil Regresi masing-masing dekade ada pada Halaman berikutnya)

Hasil OLS 1960

Dependent variable:	
Tingkat Pembunuhan	
Resource Deprivation	2.451*** (0.289)
Population Structure	-0.310 (0.296)
Median Age	-0.212*** (0.066)
Percent of males divorced	2.205*** (0.338)
Percent of Male Unemployed	-0.481*** (0.113)
Constant	11.599*** (1.808)
Observations	652
R2	0.285
Adjusted R2	0.279
Residual Std. Error	5.557 (df = 646)
F Statistic	51.400*** (df = 5; 646)
Note:	*p<0.1; **p<0.05; ***p<0.01

Hasil OLS 1980

Dependent variable:	
Tingkat Pembunuhan	
Resource Deprivation	4.200*** (0.211)
Population Structure	0.768*** (0.238)
Median Age	-0.091* (0.054)
Percent of males divorced	0.862*** (0.158)
Percent of Male Unemployed	-0.466*** (0.092)
Constant	9.580*** (1.900)
Observations	652
R2	0.432
Adjusted R2	0.427
Residual Std. Error	4.899 (df = 646)
F Statistic	98.195*** (df = 5; 646)
Note:	*p<0.1; **p<0.05; ***p<0.01

Hasil OLS 1970

Dependent variable:	
Tingkat Pembunuhan	
Resource Deprivation	4.489*** (0.346)
Population Structure	0.295 (0.381)
Median Age	-0.159** (0.074)
Percent of males divorced	0.990*** (0.301)
Percent of Male Unemployed	-0.680*** (0.192)
Constant	14.113*** (2.207)
Observations	652
R2	0.302
Adjusted R2	0.296
Residual Std. Error	7.395 (df = 646)
F Statistic	55.832*** (df = 5; 646)
Note:	*p<0.1; **p<0.05; ***p<0.01

Hasil OLS 1990

Dependent variable:	
Tingkat Pembunuhan	
Resource Deprivation	2.451*** (0.289)
Population Structure	-0.310 (0.296)
Median Age	-0.212*** (0.066)
Percent of males divorced	2.205*** (0.338)
Percent of Male Unemployed	-0.481*** (0.113)
Constant	11.599*** (1.808)
Observations	652
R2	0.285
Adjusted R2	0.279
Residual Std. Error	5.557 (df = 646)
F Statistic	51.400*** (df = 5; 646)
Note:	*p<0.1; **p<0.05; ***p<0.01

Hasil Regresi OLS, pada tahun 1960, data yang dihasilkan menunjukkan bahwa semua signifikan. Semakin tinggi angka Resource Deprivation maka akan meningkatkan tingkat pembunuhan sebesar 2.4 Poin. Sedangkan semakin meningkatnya usia berbanding terbalik dengan tingkat pembunuhan. Semakin bertambahnya Perceraian semakin bertambah pembunuhan. Data selama 4 Dekade menunjukkan Tren ini tidak berubah. Hanya yang menjadi pertanyaan adalah angka pengangguran yang meningkat maka dampaknya angka pembunuhan berkurang. Hal ini akan dikaji lebih lanjut.

6. Dengan menggunakan matrik bobot spasial knn=10, lakukan uji diagnostic LM untuk keempat model diatas. Diskusikan hasilnya, dan tentukan model yang tepat untuk setiap model (HR60,HR70, HR80, dan HR90)

7.

```

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR90 ~ RD90 + PS90 + MA90 + DV90 + UE90, data = data_nat90)
weights: nat_east_symnbn10_wl

LMerr = 24.163, df = 1, p-value = 8.852e-07

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR90 ~ RD90 + PS90 + MA90 + DV90 + UE90, data = data_nat90)
weights: nat_east_symnbn10_wl

LMlag = 38.585, df = 1, p-value = 5.242e-10

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR90 ~ RD90 + PS90 + MA90 + DV90 + UE90, data = data_nat90)
weights: nat_east_symnbn10_wl

RLMerr = 0.72071, df = 1, p-value = 0.3959

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR90 ~ RD90 + PS90 + MA90 + DV90 + UE90, data = data_nat90)
weights: nat_east_symnbn10_wl

RLMlag = 15.143, df = 1, p-value = 9.968e-05

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR90 ~ RD90 + PS90 + MA90 + DV90 + UE90, data = data_nat90)
weights: nat_east_symnbn10_wl

SARMA = 39.306, df = 2, p-value = 2.917e-09

```

Figure 1. LMtest Tahun 1990


```

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR80 ~ RD80 + PS80 + MA80 + DV80 + UE80, data = data_nat80)
weights: nat_east_symmnb10_wl
LMerr = 76.418, df = 1, p-value < 2.2e-16

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR80 ~ RD80 + PS80 + MA80 + DV80 + UE80, data = data_nat80)
weights: nat_east_symmnb10_wl
LMlag = 69.743, df = 1, p-value < 2.2e-16

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR80 ~ RD80 + PS80 + MA80 + DV80 + UE80, data = data_nat80)
weights: nat_east_symmnb10_wl
RLMerr = 15.54, df = 1, p-value = 8.078e-05

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR80 ~ RD80 + PS80 + MA80 + DV80 + UE80, data = data_nat80)
weights: nat_east_symmnb10_wl
RLMlag = 8.8643, df = 1, p-value = 0.002908

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR80 ~ RD80 + PS80 + MA80 + DV80 + UE80, data = data_nat80)
weights: nat_east_symmnb10_wl
SARMA = 85.283, df = 2, p-value < 2.2e-16

```

Figure 2. LMtest Tahun 1980

```

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR70 ~ RD70 + PS70 + MA70 + DV70 + UE70, data = data_nat70)
weights: nat_east_symmnb10_wl
LMerr = 73.787, df = 1, p-value < 2.2e-16

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR70 ~ RD70 + PS70 + MA70 + DV70 + UE70, data = data_nat70)
weights: nat_east_symmnb10_wl
LMlag = 74.932, df = 1, p-value < 2.2e-16

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR70 ~ RD70 + PS70 + MA70 + DV70 + UE70, data = data_nat70)
weights: nat_east_symmnb10_wl
RLMerr = 6.4431, df = 1, p-value = 0.01114

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR70 ~ RD70 + PS70 + MA70 + DV70 + UE70, data = data_nat70)
weights: nat_east_symmnb10_wl
RLMlag = 7.5876, df = 1, p-value = 0.005877

Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = HR70 ~ RD70 + PS70 + MA70 + DV70 + UE70, data = data_nat70)
weights: nat_east_symmnb10_wl
SARMA = 81.375, df = 2, p-value < 2.2e-16

```

Figure 3. LMtest Tahun 1970


```

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR60 ~ RD60 + PS60 + MA60 + DV60 + UE60, data = data_nat60)
weights: nat_east_symnbn10_wl

LMerr = 36.874, df = 1, p-value = 1.26e-09

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR60 ~ RD60 + PS60 + MA60 + DV60 + UE60, data = data_nat60)
weights: nat_east_symnbn10_wl

LMlag = 51.957, df = 1, p-value = 5.674e-13

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR60 ~ RD60 + PS60 + MA60 + DV60 + UE60, data = data_nat60)
weights: nat_east_symnbn10_wl

RLMerr = 0.0002838, df = 1, p-value = 0.9866

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR60 ~ RD60 + PS60 + MA60 + DV60 + UE60, data = data_nat60)
weights: nat_east_symnbn10_wl

RLMlag = 15.083, df = 1, p-value = 0.0001029

Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = HR60 ~ RD60 + PS60 + MA60 + DV60 + UE60, data = data_nat60)
weights: nat_east_symnbn10_wl

SARMA = 51.957, df = 2, p-value = 5.22e-12

```

Figure 4. LMtest Tahun 1990

Dalam menginterpretasi LMtest, dari 4 dekade, tidak ada perubahan antara korelasi variable dependen dan variable independent. penulis juga melihat nilai LMerr, LMlag, RLMerr dan RLMlag. Apabila :

LMerr Signifikan **sedangkan** LMlag Tidak Signifikan. maka; menggunakan SEM
 LMerr Tidak Signifikan **sedangkan** LMlag Signifikan. maka; menggunakan SLM
 Apabila keduanya signifikan, maka dilihat lebih lanjut nilai dari RLMerr dan RLMlag
 Apabila RLMerr signifikan **sedangkan** RMLag tidak signifikan. Maka; menggunakan SEM
 Apabila RLMerr Tidak signifikan **sedangkan** RMLag signifikan. Maka; menggunakan SLM

Apabila semuanya signifikan, maka menggunakan SDM.

Apabila nilai SARMA, dan Robust keduanya signifikan, maka menggunakan SDM.

Hasil observasi dari LM test tahun 1960, LMerr dan LMlag signifikan sedangkan RMLerr tidak signifikan dan RLMLag signifikan. Maka lebih cocok menggunakan model SLM. Untuk tahun 1970 dan 1980 lebih cocok menggunakan model SDM. Sedangkan untuk tahun 1990 lebih cocok menggunakan SLM karena nilai RLMerr tidak signifikan dan nilai lainnya signifikan.

8. Dari hasil point 6, tentukan 1 model baseline OLS dimana anda dapat melanjutkan untuk melakukan pemodelan dengan menggunakan Spatial Error Model (SEM) dan Spatial Lag Model (SLM) secara bersamaan. Lakukan pemodelan dengan model SEM dan SLM. Perhatikan hasilnya, apakah penggunaan kedua model tersebut telah menyelesaikan persoalan spatial autocorrelation?

Penulis tidak menemukan satu model baseline yang cocok untuk melanjutkan pemodelan menggunakan SEM dan SLM secara bersamaan. Maka dari itu penulis hanya akan memilih regresi tahun 1960 dikarenakan menurut hasil LM test cocok menggunakan SLM.

Hasil dari Spatial Lag Model, pada nilai “rho” menunjukan bahwa Efek dari tetangga berperan signifikan terhadap meningkatnya tingkat pembunuhan. Namun, model ini belum menyelesaikan persoalan autocorrelation.

9. Lakukan pemodelan dengan menggunakan Spatial Durbin model. Lakukan estimasi impact (Direct, Indirect, dan Total Impacts) untuk interpretasi estimasi parameter dengan lebih tepat. Interpretasikan hasil estimasi impact tersebut. Diskusikan, apakah anda menemukan bukti terjadinya spillover?

Setelah menghitung direct dan indirect effect serta mengkoreksi Standar Error, terbukti ada temuan yang meyakinkan bahwa terjadi Spillover.

(Hasil ada di halaman berikutnya)

```
Call:lagsarlm(formula = HR90 ~ RD90 + PS90 + MA90 + DV90 + UE90, data = data_nat90,
  listw = nat_east_symnb10_wl, type = "mixed", method = "Matrix",
  tol.solve = 1e-12, trs = T, control = list())
```

Residuals:

	Min	1Q	Median	3Q	Max
	-14.05749	-2.58739	-0.46933	2.01138	23.86591

Type: mixed

Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	11.049772	3.200926	3.4521	0.0005563
RD90	4.183165	0.460748	9.0791	< 2.2e-16
PS90	1.242361	0.366578	3.3891	0.0007013
MA90	-0.023880	0.072112	-0.3312	0.7405253
DV90	0.755146	0.201642	3.7450	0.0001804
UE90	-0.083217	0.173214	-0.4804	0.6309212
lag.RD90	1.538978	0.515063	2.9879	0.0028086
lag.PS90	1.460414	0.559993	2.6079	0.0091096
lag.MA90	-0.065803	0.119372	-0.5512	0.5814647
lag.DV90	0.155879	0.250968	0.6211	0.5345274
lag.UA90	-1.402412	0.232037	-6.0439	1.504e-09

Rho: 0.17722, LR test value: 4.672, p-value: 0.030658

Asymptotic standard error: 0.082513

z-value: 2.1478, p-value: 0.031734

Wald statistic: 4.6128, p-value: 0.031734

Log likelihood: -1924.517 for mixed model

ML residual variance (sigma squared): 21.383, (sigma: 4.6242)

Number of observations: 652

Number of parameters estimated: 13

AIC: 3875, (AIC for lm: 3877.7)

LM test for residual autocorrelation

test value: 7.7235, p-value: 0.0054507

Dengan adanya data indirect, maka dapat dibuktikan bahwa ada pengaruh spasial.

Indirect:

Iterations = 1:1000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1000

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
RD90	2.77621	0.5610	0.017741	0.018940
PS90	2.02451	0.6569	0.020774	0.020774
MA90	-0.08284	0.1339	0.004233	0.004233
DV90	0.34256	0.2763	0.008737	0.008294
UE90	-1.71077	0.2706	0.008557	0.008557

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
RD90	1.6643	2.4006	2.76953	3.144830	3.9505
PS90	0.8087	1.5953	2.03334	2.439881	3.3450
MA90	-0.3535	-0.1726	-0.07729	0.001697	0.1732
DV90	-0.1922	0.1566	0.33296	0.517533	0.8941
UE90	-2.2430	-1.8902	-1.70833	-1.528911	-1.1943

=====
Total:

Iterations = 1:1000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1000

10. Lakukan partisi impacts kedalam 4 orde tetangga. Tentukan dari effects untuk variabel independent yang terbukti memiliki efek spillover, tentukan sifat spillover tersebut apakah bersifat local ataukah global. Diskusikan!

Setelah berhasil mempartisi impact kedalam 4 Orde tetangga, penulis mendapat kesulitan dalam proses visualisasi. Keterbatasan atas hal ini membuat penulis kurang dapat gambaran secara utuh dalam menjawab apakah spillover bersifat global atau local.