Hw6

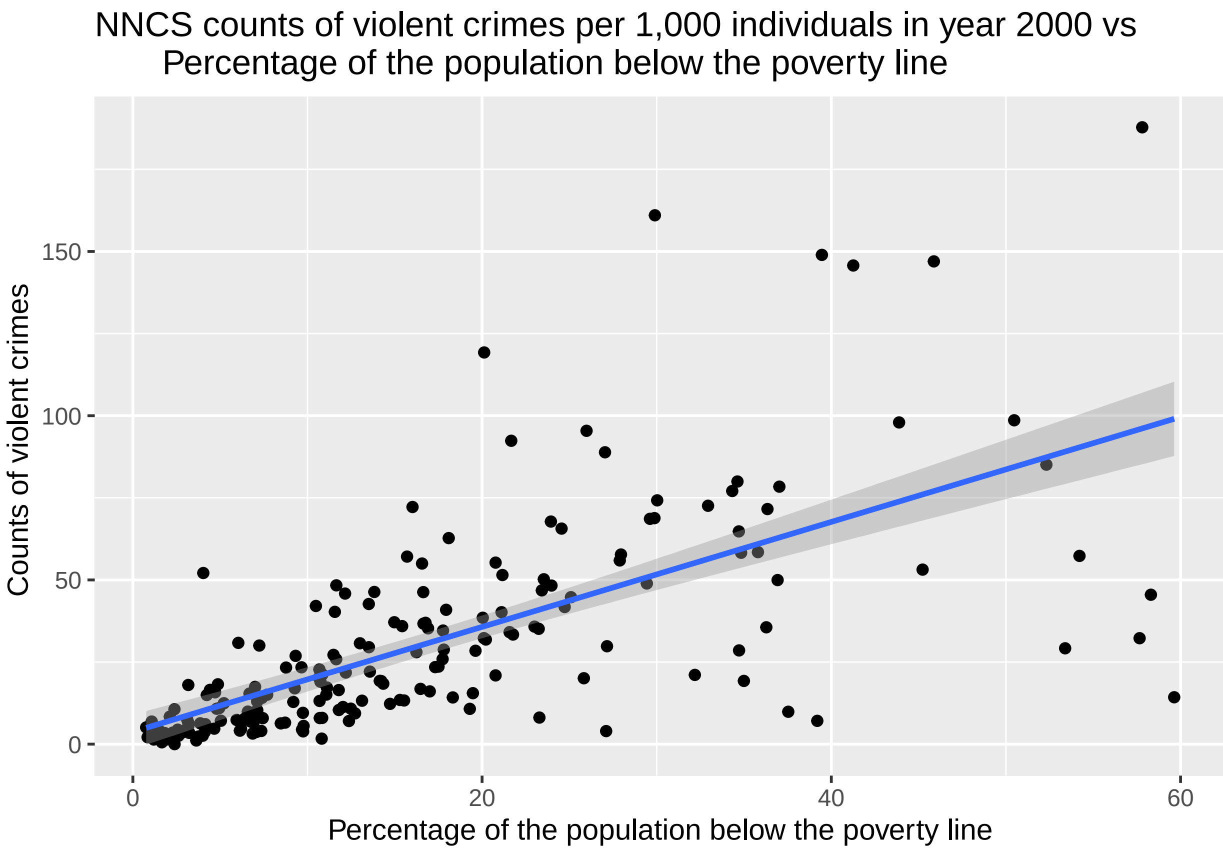
Rafael Espinosa Castaneda

2023-03-06

```
# -------------------------Setting the parameters to be used-----------------------
## load libraries
library(ggplot2)
library(gridExtra)
library(sf)
library(sp)
library(spdep)
library(spatialreg)
load("/Users/rafa/Documents/Master Austin/MAESTRIA_AUSTIN/Advanced Predictive Models/Hw6/Hw6Part1.RData
```

Part a

```
# scatterplot of avg_logPOS vs. logPOS
ggplot(data=crime_dat,
       aes(x = poverty, y = crime)) +
  geom_point() +
  stat_smooth(method = "lm") +
  xlab("Percentage of the population below the poverty line") +
  ylab("Counts of violent crimes ")+
  labs(title="NNCS counts of violent crimes per 1,000 individuals in year 2000 vs
       Percentage of the population below the poverty line")
```



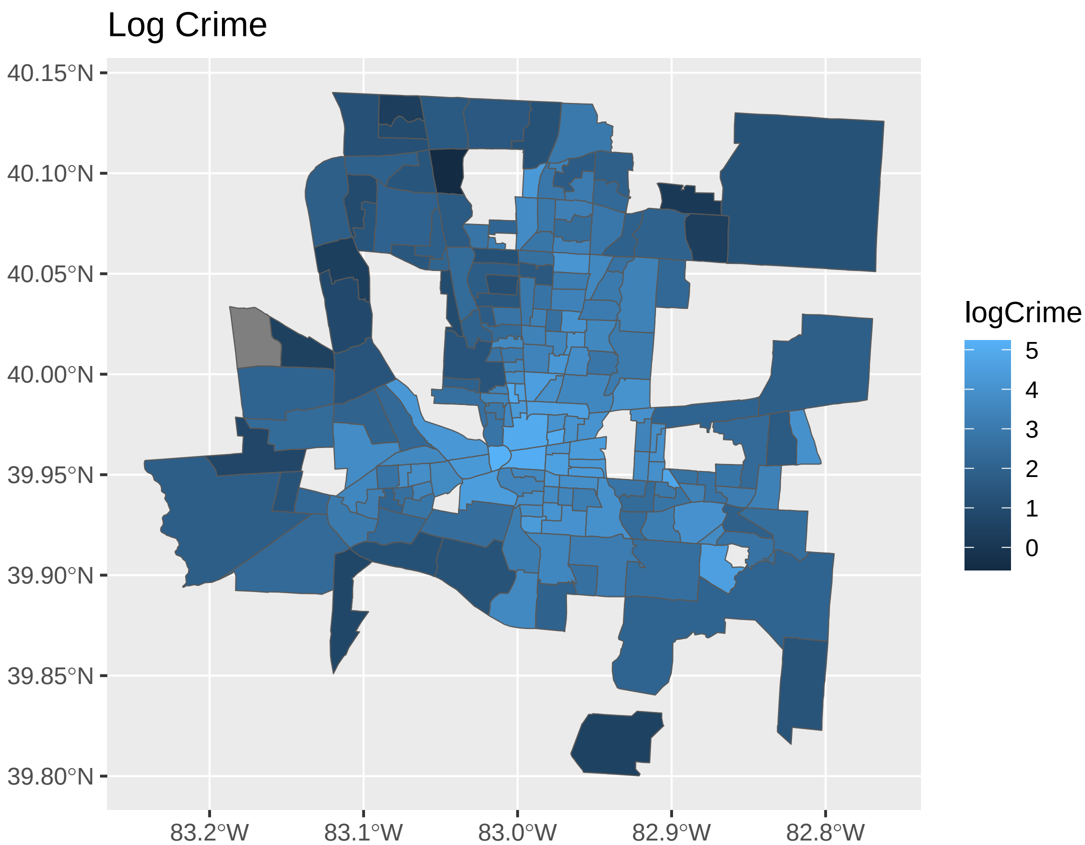
Part b

ANSWER

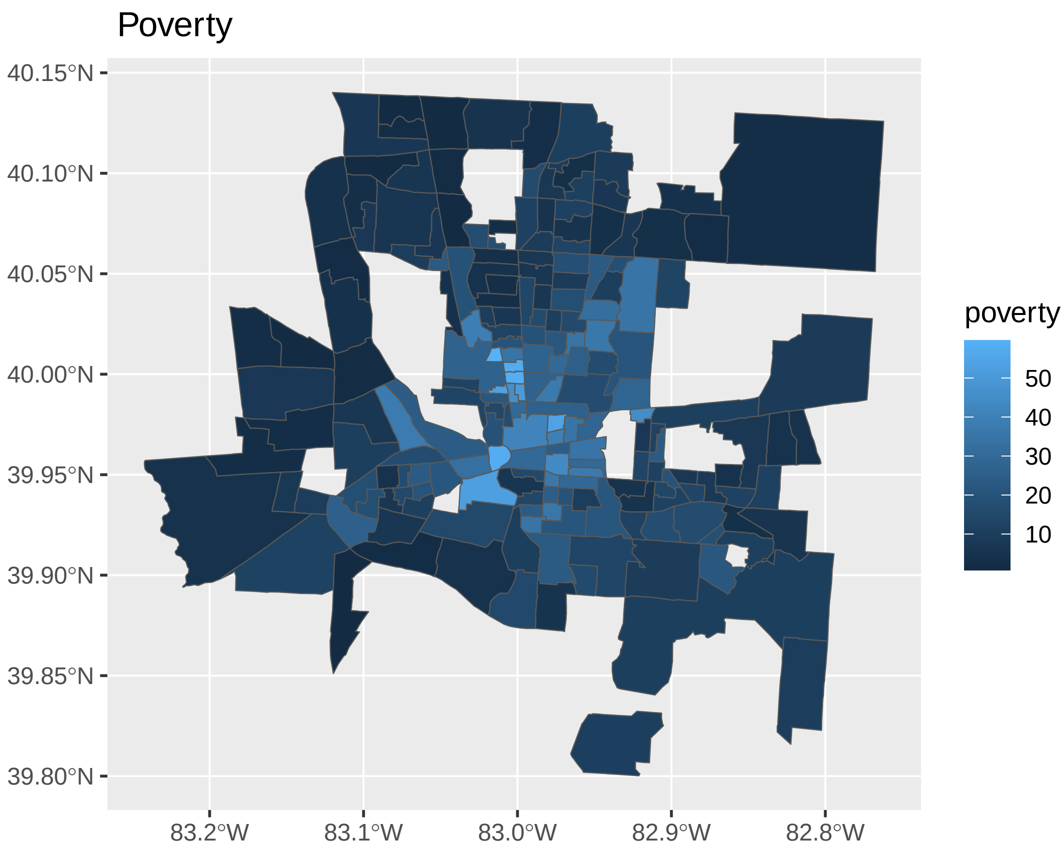
We cannot make a conclusion from aggregate data to individuals.

Part c

```
crime_dat$logCrime <- log(crime_dat$crime)
crime_dat$logPoverty <- log(crime_dat$poverty)







ggplot(data = crime_dat) +
  geom_sf(aes(fill = logCrime))+
  labs(title="Log Crime")
```

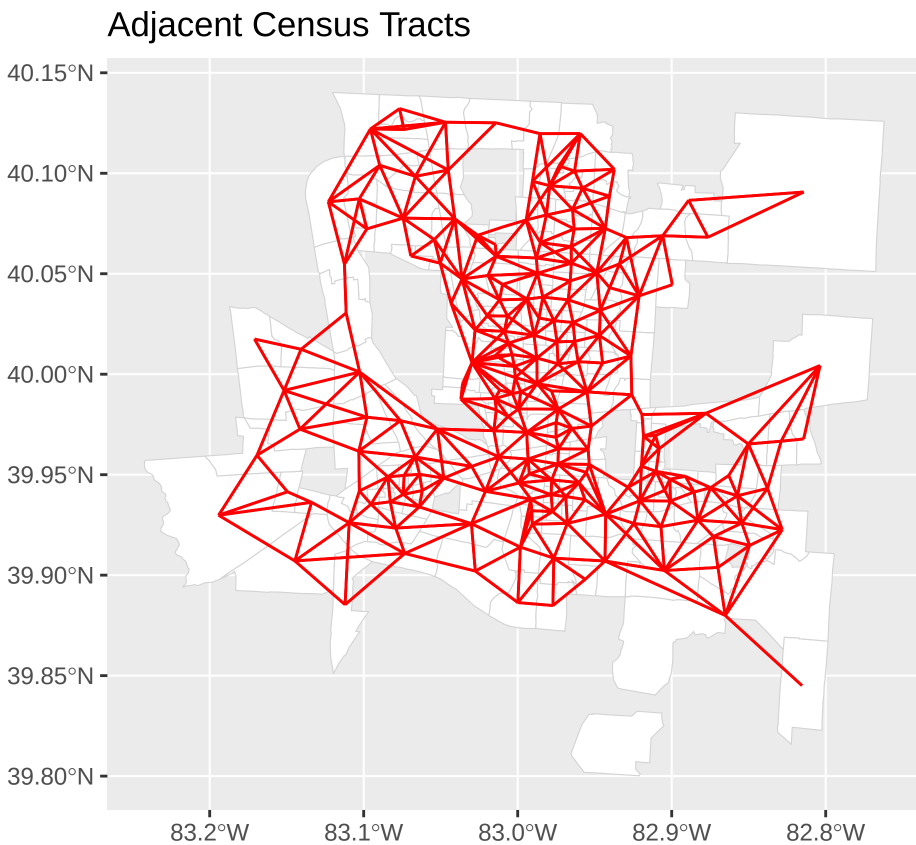



```r
#-------


# map of the log Poverty


ggplot(data = crime_dat) +
  geom_sf(aes(fill = poverty))+
  labs(title=" Poverty")
```



Part d

```
# Construct neighborhood list
nb <- poly2nb(crime_dat)




# -> plot
nb_lines <- nb %>%
  nb2lines(coords = coordinates(as(crime_dat, "Spatial"))) %>%
  as("sf") %>%
  st_set_crs(st_crs(crime_dat))

ggplot(data = crime_dat) +
  geom_sf(fill = "white", color = "lightgrey") +
  geom_sf(data = nb_lines, col = "red") +
  labs(title = "Adjacent Census Tracts")
```



Part e

```
summary(nb)
```

```
## Neighbour list object:
## Number of regions: 196
## Number of nonzero links: 1072
## Percentage nonzero weights: 2.790504
## Average number of links: 5.469388
## 1 region with no links:
## 95.10
## Link number distribution:
##
##  0  1  2  3  4  5  6  7  8  9 10 13
##  1  1  7 16 34 42 43 28 14  6  3  1
## 1 least connected region:
## 94.50 with 1 link
## 1 most connected region:
## 11.20 with 13 links
```

```
crime_dat
```

```
## Simple feature collection with 196 features and 4 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -83.24239 ymin: 39.80016 xmax: -82.7621 ymax: 40.14016
## CRS:           +proj=longlat +datum=WGS84
```

```
## First 10 features:
##          crime poverty                        geometry logCrime logPoverty
## 1.10  3.430532   3.230 POLYGON ((-83.01831 40.0623... 1.232715  1.1724821
## 1.20  5.719733   4.107 POLYGON ((-83.0029 40.04886... 1.743922  1.4126928
## 2.10  2.987056   2.622 POLYGON ((-82.9999 40.04696... 1.094288  0.9639374
## 2.20  4.404145   2.569 POLYGON ((-83.01511 40.0332... 1.482546  0.9435167
## 3.10 19.207317  14.224 POLYGON ((-82.9901 40.03076... 2.955291  2.6549307
## 3.20 15.384615   6.680 POLYGON ((-82.9797 40.03216... 2.733368  1.8991180
## 3.30 27.998328  16.240 POLYGON ((-82.9904 40.02836... 3.332145  2.7874773
## 4.10  5.542359   9.778 POLYGON ((-83.02371 40.0255... 1.712420  2.2801350
## 4.20 16.336789   6.955 POLYGON ((-83.01401 40.0243... 2.793420  1.9394608
## 5    10.798594  12.481 POLYGON ((-83.007 40.01776,... 2.379416  2.5242075
```

```r
coordinates_centroids<-coordinates(as(crime_dat, "Spatial"))

print("One census tract have no neighbors, i.e. one island")
```

```
## [1] "One census tract have no neighbors, i.e. one island"
```

```r
print(paste("coordinates",-82.93626, 39.81671))
```

```
## [1] "coordinates -82.93626 39.81671"
```

ANSWER

One census tract have no neighbors, i.e. one island.

coordinates of the island (-82.93626, 39.81671)

—————————— **Second Homework**——————————

Part f

```r
crime_dat2<-crime_dat[-195,]
crime_filtered<-crime_dat2[crime_dat2$crime!=0,]

nb_filtered <- poly2nb(crime_filtered)





# construct binary adjacency matrix
W_mat <- nb2listw(nb_filtered,
                  style = 'B',
                  zero.policy = T)

crime_filtered$log_Crime <- log(crime_filtered$crime)

# SMA Model

fit_SMA <- spautolm(log_Crime ~ poverty,
                    listw = W_mat,
                    family = "SMA",
                    data = crime_filtered)
summary(fit_SMA)
```

```
##
## Call: spautolm(formula = log_Crime ~ poverty, data = crime_filtered,
##     listw = W_mat, family = "SMA")
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -2.0916860 -0.4282897  0.0065588  0.3374632  1.9290587
##
## Coefficients:
##             Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 1.999317   0.118105 16.9283 < 2.2e-16
## poverty     0.045886   0.004671  9.8236 < 2.2e-16
##
## Lambda: 0.14511 LR test value: 56.527 p-value: 5.54e-14
## Numerical Hessian standard error of lambda: 0.019082
##
## Log likelihood: -210.7505
## ML residual variance (sigma squared): 0.57208, (sigma: 0.75636)
## Number of observations: 194
## Number of parameters estimated: 4
## AIC: 429.5
```

```r
print(paste("The intercept: " ,signif(1.999317 ,3)," with standard error",signif(0.118105,3)))
```

```
## [1] "The intercept:  2  with standard error 0.118"
```

```r
print(paste("The poverty coefficient: " ,signif(0.045886,3)," with standard error",signif(0.004671,3)))
```

```
## [1] "The poverty coefficient:  0.0459  with standard error 0.00467"
```

```r
print("A positive  change in poverty rate is statistically significantly associated with a positive smal
```

```
## [1] "A positive  change in poverty rate is statistically significantly associated with a positive sma
```

ANSWER

```
## [1] "The intercept:  2  with standard error 0.118"
```

```
## [1] "The poverty coefficient:  0.0459  with standard error 0.00467"
```

A positive change in poverty rate is statistically significantly associated with a positive small change in crime rate after adjusting for the spatial structure of crime rate in the city

Part g

```r
# non-spatial linear model
fit_ns <- lm(log_Crime ~ poverty,
             data = crime_filtered)

summary(fit_ns)
```

```
##
## Call:
## lm(formula = log_Crime ~ poverty, data = crime_filtered)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6803 -0.4624  0.1775  0.5833  1.8034
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.918417   0.094889   20.22   <2e-16 ***
## poverty     0.057356   0.004466   12.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8338 on 192 degrees of freedom
## Multiple R-squared:  0.4621, Adjusted R-squared:  0.4593
## F-statistic:   165 on 1 and 192 DF,  p-value: < 2.2e-16
```

ANSWER

```
## [1] "Without three significant figures"
```

```
## [1] "The intercept:  1.918417 3  with standard error 0.094889"
```

```
## [1] "The poverty coefficient:  0.057356  with standard error 0.004466"
```

```
## [1] "Non spatial AIC: 484.028320130582"
```

```
## [1] "SMA AIC :  429.500953871692"
```

```
## [1] "---------------With three significant figures------"
```

```
## [1] "The intercept:  1.92  with standard error 0.0949"
```

```
## [1] "The poverty coefficient:  0.0574  with standard error 0.00447"
```

```
## [1] "Non spatial AIC: 484.028"
```

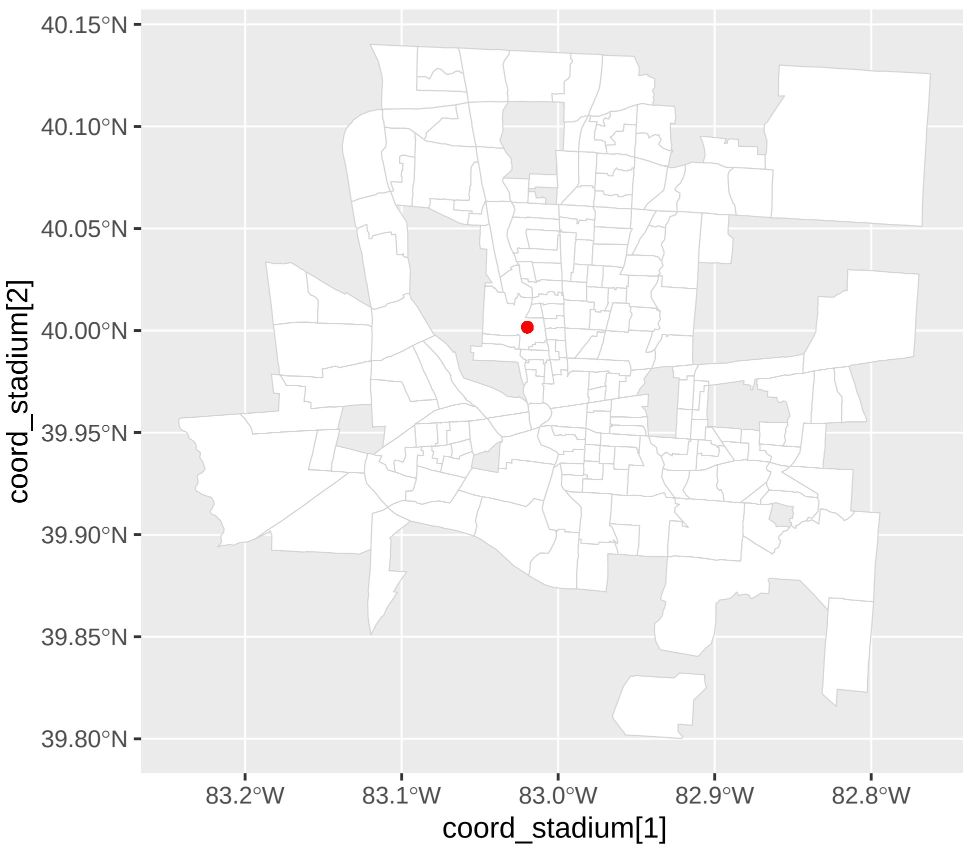
```
## [1] "SMA AIC :  429.501"
```

-The non-spatial model coefficient predicts a greater change in crime rate due to change in poverty rate than the SMA model.

-The model that better fits the data is the spatial moving average model (SMA)

Part h

```
coord_stadium<-c(-83.019707, 40.001633)


st_stadium<-st_point(coord_stadium)
r_within<-st_contains(st_stadium,crime_dat)






ggplot(data = crime_dat) +
  geom_sf(fill = "white", color = "lightgrey") +
  geom_point(aes(x=coord_stadium[1],y=coord_stadium[2]),color="red")
```



Part i

```
crime_filtered$pred_ns <- fit_ns$fitted.values
crime_filtered$pred_sma <- fit_SMA$fit$fitted.values
```

```r
crime_filtered
```

```
## Simple feature collection with 194 features and 7 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -83.24239 ymin: 39.81582 xmax: -82.7621 ymax: 40.14016
## CRS:           +proj=longlat +datum=WGS84
## First 10 features:
##          crime poverty                       geometry logCrime logPoverty
## 1.10  3.430532   3.230 POLYGON ((-83.01831 40.0623... 1.232715  1.1724821
## 1.20  5.719733   4.107 POLYGON ((-83.0029 40.04886... 1.743922  1.4126928
## 2.10  2.987056   2.622 POLYGON ((-82.9999 40.04696... 1.094288  0.9639374
## 2.20  4.404145   2.569 POLYGON ((-83.01511 40.0332... 1.482546  0.9435167
## 3.10 19.207317  14.224 POLYGON ((-82.9901 40.03076... 2.955291  2.6549307
## 3.20 15.384615   6.680 POLYGON ((-82.9797 40.03216... 2.733368  1.8991180
## 3.30 27.998328  16.240 POLYGON ((-82.9904 40.02836... 3.332145  2.7874773
## 4.10  5.542359   9.778 POLYGON ((-83.02371 40.0255... 1.712420  2.2801350
## 4.20 16.336789   6.955 POLYGON ((-83.01401 40.0243... 2.793420  1.9394608
## 5    10.798594  12.481 POLYGON ((-83.007 40.01776,... 2.379416  2.5242075
##      log_Crime  pred_ns pred_sma
## 1.10  1.232715 2.103677 2.096258
## 1.20  1.743922 2.153978 1.675274
## 2.10  1.094288 2.068804 1.898327
## 2.20  1.482546 2.065764 1.836134
## 3.10  2.955291 2.734247 2.491140
## 3.20  2.733368 2.301554 2.526191
## 3.30  3.332145 2.849876 2.928146
## 4.10  1.712420 2.479243 2.055964
## 4.20  2.793420 2.317327 2.200635
## 5     2.379416 2.634276 2.212027
```

ANSWER

```
## [1] "Prediction of log crime NS 2.873"

## [1] "Prediction of log crime SMA:  3.145"

## [1] "Observed log crime 3.602"
```

We can notice that the prediction of the SMA model is closer to the actual log crime value than the non spatial model.