Homework4

Laha Ale

March 17, 2019

```
library(spdep)
```

```
## Loading required package: sp
```

```
## Loading required package: Matrix
```

```
## Loading required package: spData
```

```
## To access larger datasets in this package, install the spDataLarge
## package with: `install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source')`
```

```
## Loading required package: sf
```

```
## Linking to GEOS 3.7.0, GDAL 2.4.0, PROJ 5.2.0
```

```
library(maps)
library(maptools)
```

```
## Checking rgeos availability: TRUE
```

```
library(classInt)
library(RColorBrewer)
```

Excercise 11

(a)

```
state.sat.scores = read.table("state-sat.dat", header=F)
colnames(state.sat.scores) <- c("STATE","VERBAL","MATH","ELIGIBLE")
head(state.sat.scores)
```

```
##    STATE VERBAL MATH ELIGIBLE
## 1    ala    561  555        9
## 2 alaska    516  514       50
## 3   ariz    524  525       34
## 4    ark    563  556        6
## 5  calif    497  514       49
## 6   colo    536  540       32
```

```
# create listw
usa.state <- map(database="state", fill=TRUE, plot=FALSE)
state.ID <- sapply(strsplit(usa.state$names, ":"),
                   function(x) x[1])

usa.poly <- map2SpatialPolygons(usa.state,
                                IDs=state.ID)
usa.nb <- poly2nb(usa.poly)
usa.listb <- nb2listw(usa.nb, style="B")
usa.listw <- nb2listw(usa.nb, style="W")

# train SAR model
```

```r
x = ((state.sat.scores$STATE=="alaska") |
       (state.sat.scores$STATE=="hawaii") |
       (state.sat.scores$STATE=="us"))

index = c(1:nrow(state.sat.scores))[x]
state.sat.scores = state.sat.scores[-index,]

# binnary weights
stat.sat.sar.b = spautolm(ELIGIBLE~ VERBAL,
                          data=state.sat.scores,
                          family="SAR",
                          listw=usa.listb,
                          zero.policy=TRUE)
summary(stat.sat.sar.b)
```

```
##
## Call: spautolm(formula = ELIGIBLE ~ VERBAL, data = state.sat.scores,
##     listw = usa.listb, family = "SAR", zero.policy = TRUE)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -24.88699  -7.47460   0.97745   6.14293  16.45480
##
## Coefficients:
##               Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 390.605790  30.166352  12.948 < 2.2e-16
## VERBAL       -0.653476   0.058125 -11.243 < 2.2e-16
##
## Lambda: 0.15957 LR test value: 26.465 p-value: 2.6836e-07
## Numerical Hessian standard error of lambda: 0.016179
##
## Log likelihood: -179.7985
## ML residual variance (sigma squared): 75.436, (sigma: 8.6854)
## Number of observations: 49
## Number of parameters estimated: 4
## AIC: 367.6
```

```r
# row-normalized weights
stat.sat.sar.w = spautolm(ELIGIBLE~ VERBAL,
                          data=state.sat.scores,
                          family="SAR",
                          listw=usa.listw,
                          zero.policy=TRUE)
summary(stat.sat.sar.w)
```

```
##
## Call: spautolm(formula = ELIGIBLE ~ VERBAL, data = state.sat.scores,
##     listw = usa.listw, family = "SAR", zero.policy = TRUE)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -23.69268  -5.31293  -0.21455   5.86328  17.06011
##
## Coefficients:
```

```
##               Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 387.625170  30.011970  12.916 < 2.2e-16
## VERBAL       -0.658120   0.055979 -11.757 < 2.2e-16
##
## Lambda: 0.74267 LR test value: 31.772 p-value: 1.7338e-08
## Numerical Hessian standard error of lambda: 0.088661
##
## Log likelihood: -177.145
## ML residual variance (sigma squared): 67.444, (sigma: 8.2124)
## Number of observations: 49
## Number of parameters estimated: 4
## AIC: 362.29
```

As we can see from the above model that both VERBAL and MATH are significant since their P-Value is small enough; therefore, we shall keep both. The binnary weights model is $ELIGIBLE = 390.60579 - 0.653476 * VERBAL$,and The row-normalized weights model is $ELIGIBLE = 387.62517 - 0.65812 * VERBAL$. The ELIGIBLE has a negative relation with VERBAL. A possible reason VERBAL score has a negative relative with ELIGIBLE is that the erea higher verbal score the higher the more competitive in terms of admission for the area.

(b)

```
library(ggplot2)
library(ggmap)
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(mapdata)

states <- map_data("state")
states <- data.frame(states)

states[,"sat_range"] <- 0
names(states)[6] <- "sat"
state.sat.scores$STATE <- unique(states$region)

for(i in 1:nrow(states))
{
  for (j in 1:nrow(state.sat.scores))
  {
    if(grepl(state.sat.scores[j,]$STATE, states[i,]$region))
    {
      #sat <-  (state.sat.scores[j,]$VERBAL+state.sat.scores[j,]$MATH)/2
      sat <- state.sat.scores[j,]$VERBAL
      states[i,]$sat <- sat
      if(sat <= 503)
      {
        states[i,]$sat_range <- "<503"
      }else if(sat > 503 && sat <=525 )
      {
        states[i,]$sat_range <- "504-525"
      }else if(sat > 525 && sat <562 )
      {
```

```
        states[i,]$sat_range <- "526-562"
      }else
      {
        states[i,]$sat_range <- ">563"
      }
    }
  }
}

ggplot(data = states) +
  geom_polygon(aes(x = long,
                   y = lat, fill = sat_range,group=group),color = "white") +
  coord_fixed(1.3)
```