Spatial Econometrics Problem Set R

Lin Yu Chen

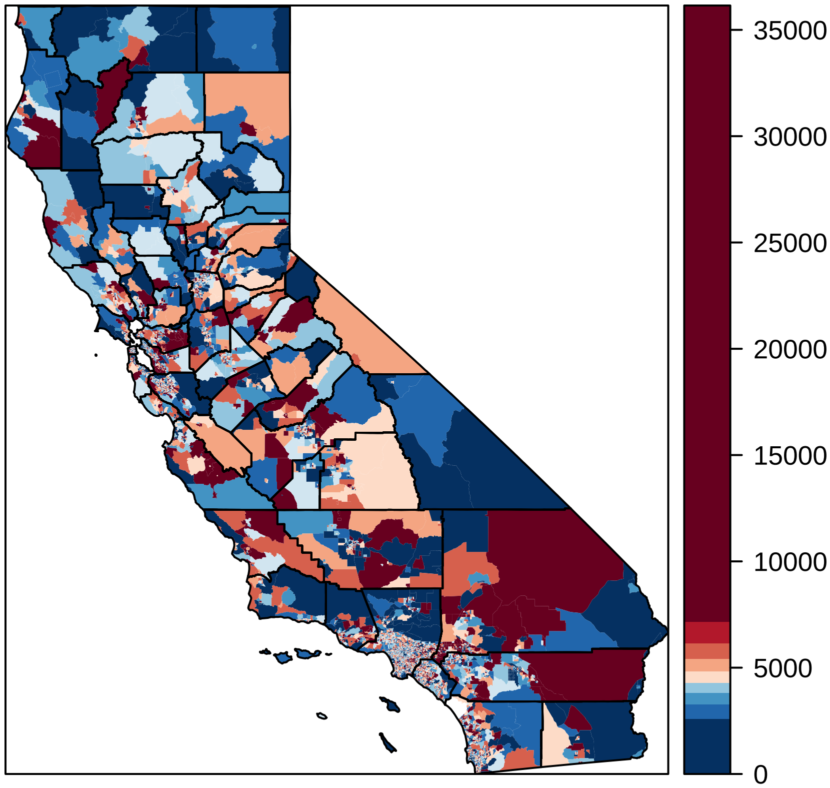
8/7/2020

```r
load("~/Desktop/Waseda/Sophomore Spring Semester/Spatial Econometrics/h.RData")
library(RColorBrewer)
library(latticeExtra)
```

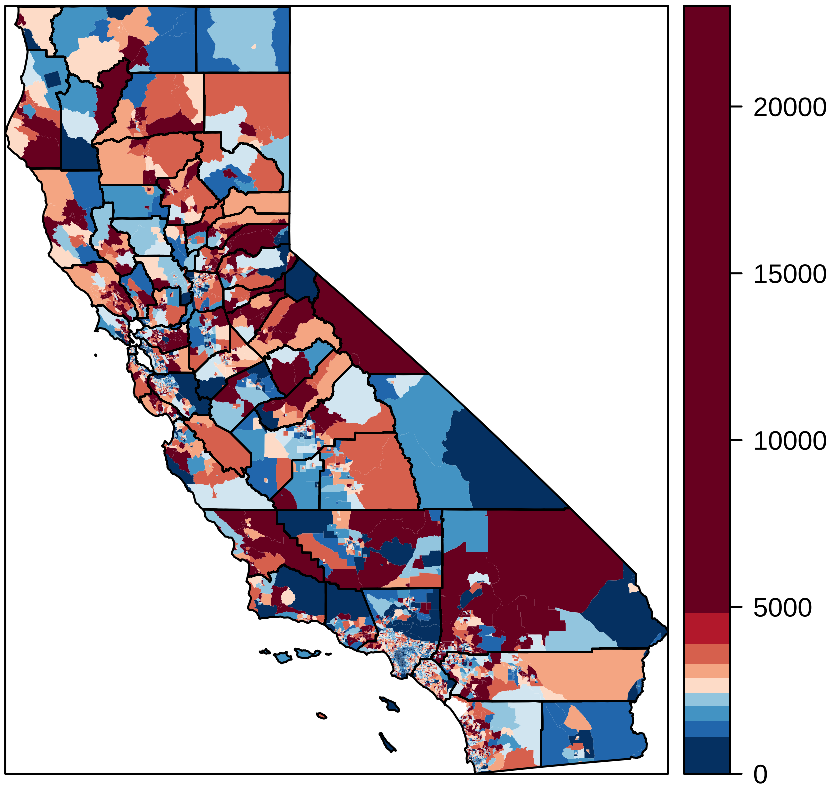
1. Charge the data in R.

```r
#Total Population
grps <- 10
bk1 <- quantile(h$Population, 0:(grps-1)/(grps-1), na.rm=TRUE)
p <- spplot(h, "Population", at=bk1, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
p + layer(sp.polygons(hh))
```

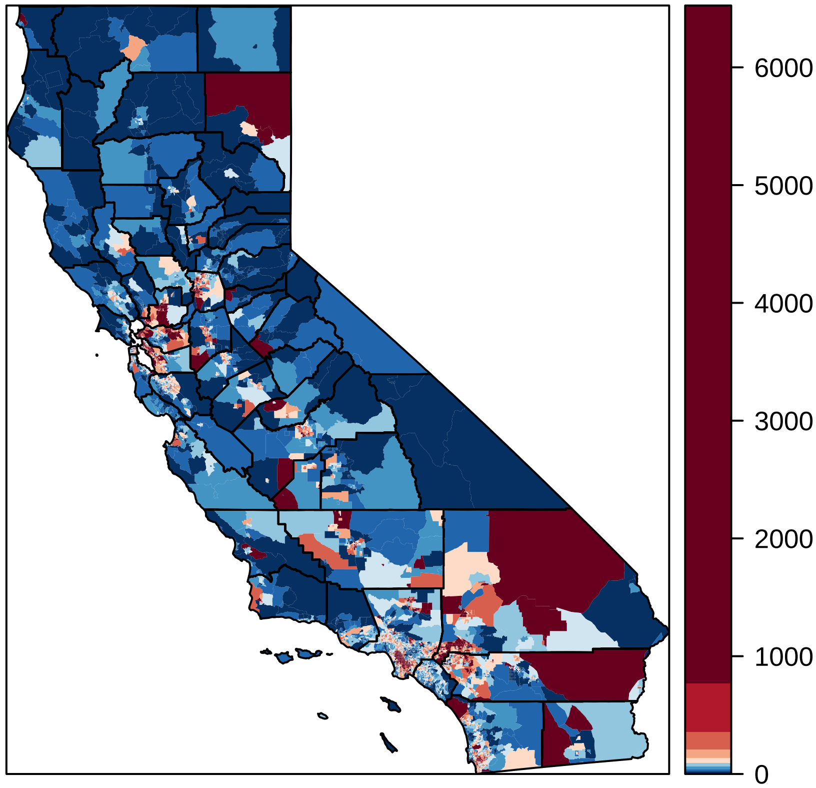
2. In the program, line 40 to line 43, you create the graphic to observe the distribution of the Value of house on the map. Create similar maps for: total population, number of White, African-american, Asian, and Hispanic, additionally do it for median family size and median household income. Analyze your results by comparing your maps.



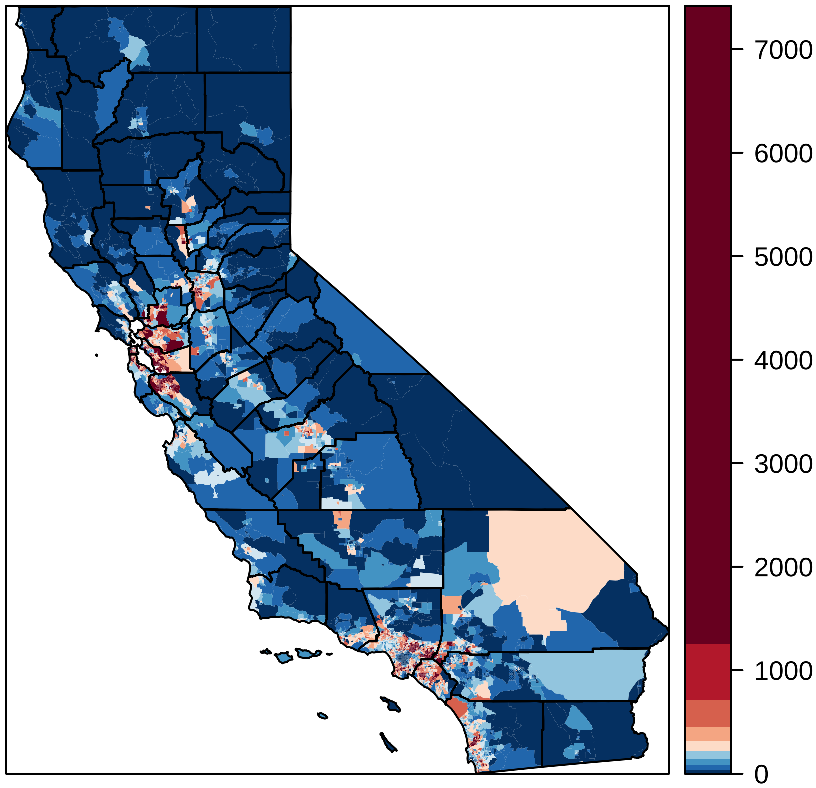

```r
#White
bk2 <- quantile(h$White, 0:(grps-1)/(grps-1), na.rm=TRUE)
p2 <- spplot(h, "White", at=bk2, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
p2 + layer(sp.polygons(hh))
```



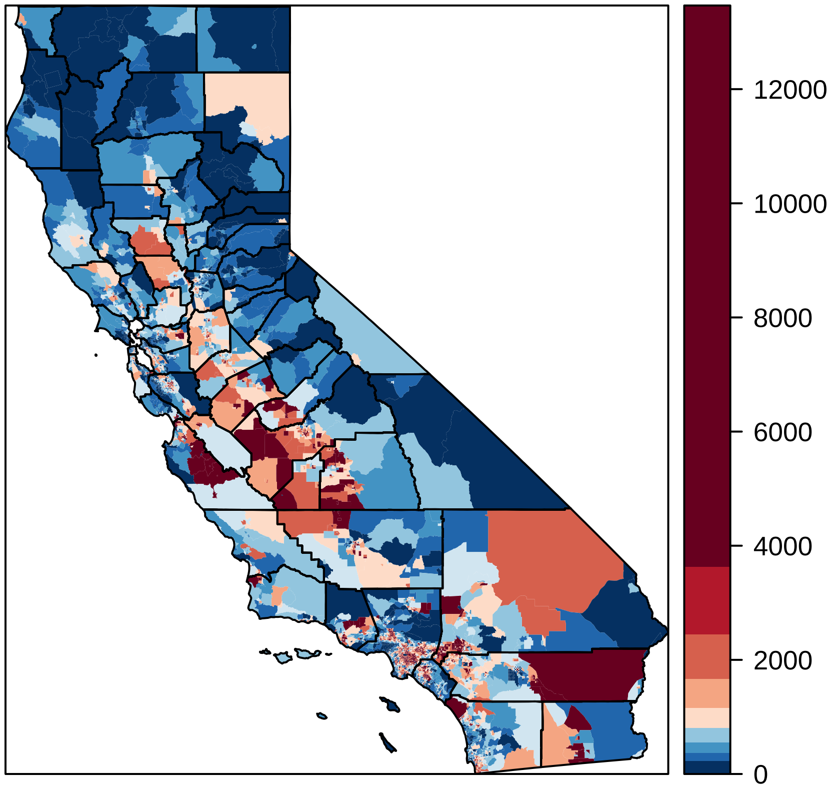

```r
#African-american
bk3 <- quantile(h$Black, 0:(grps-1)/(grps-1), na.rm=TRUE)
p3 <- spplot(h, "Black", at=bk3, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
p3 + layer(sp.polygons(hh))
```



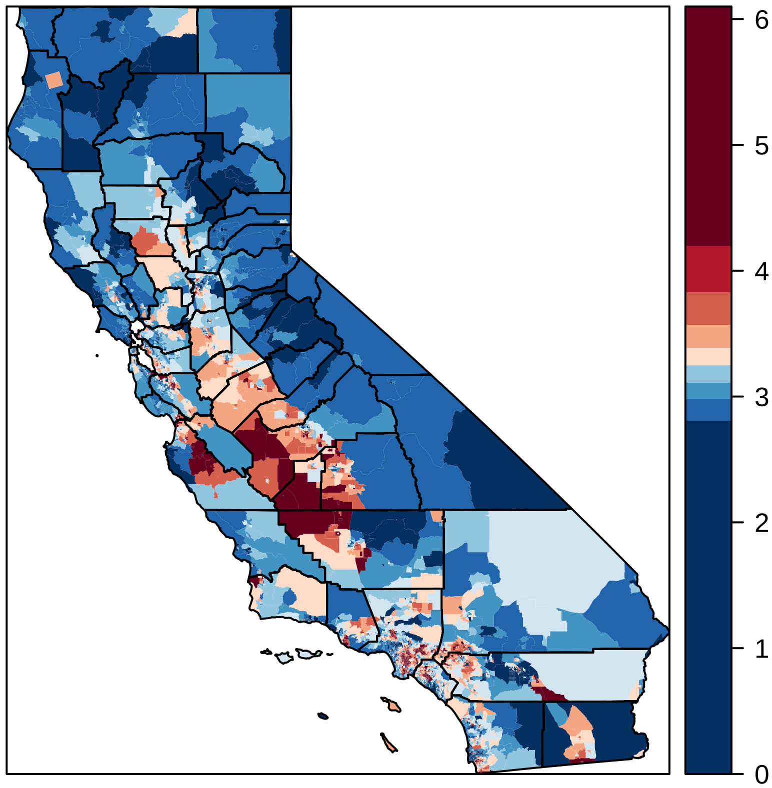

```
#Asian
bk4 <- quantile(h$Asian, 0:(grps-1)/(grps-1), na.rm=TRUE)
p4 <- spplot(h, "Asian", at=bk4, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
p4 + layer(sp.polygons(hh))
```



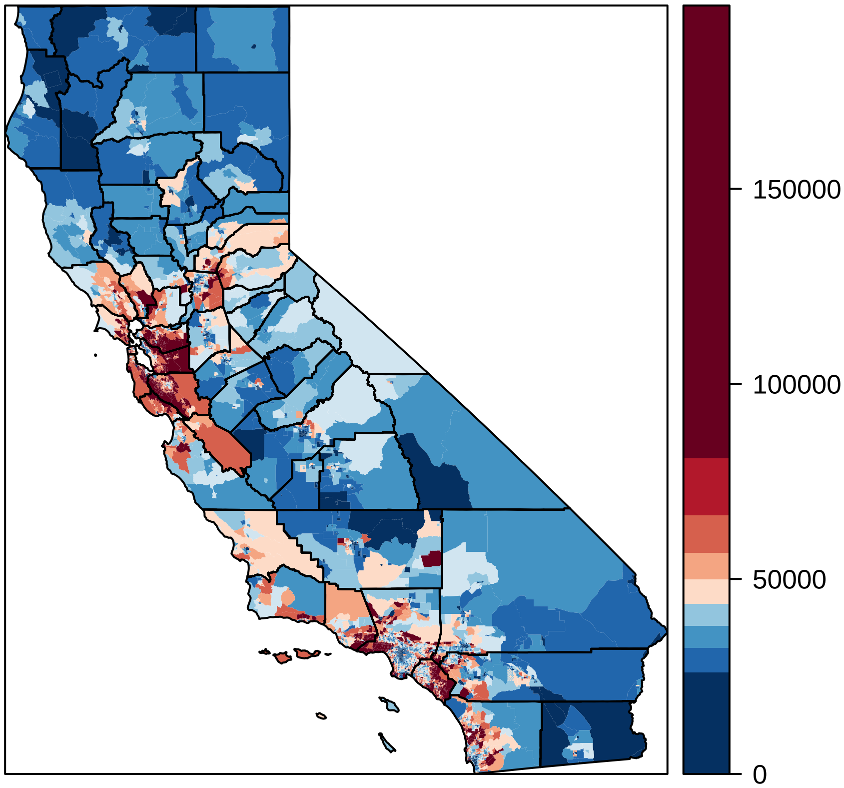

```r
#Hispanic
bk5 <- quantile(h$Hispanic, 0:(grps-1)/(grps-1), na.rm=TRUE)
p5 <- spplot(h, "Hispanic", at=bk5, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
p5 + layer(sp.polygons(hh))
```




```
#familysize
bk6 <- quantile(h$familySize, 0:(grps-1)/(grps-1), na.rm=TRUE)
p6 <- spplot(h, "familySize", at=bk6, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
p6 + layer(sp.polygons(hh))
```




```r
#medHHinc
bk7 <- quantile(h$medHHinc, 0:(grps-1)/(grps-1), na.rm=TRUE)
p7 <- spplot(h, "medHHinc", at=bk7, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
p7 + layer(sp.polygons(hh))
```



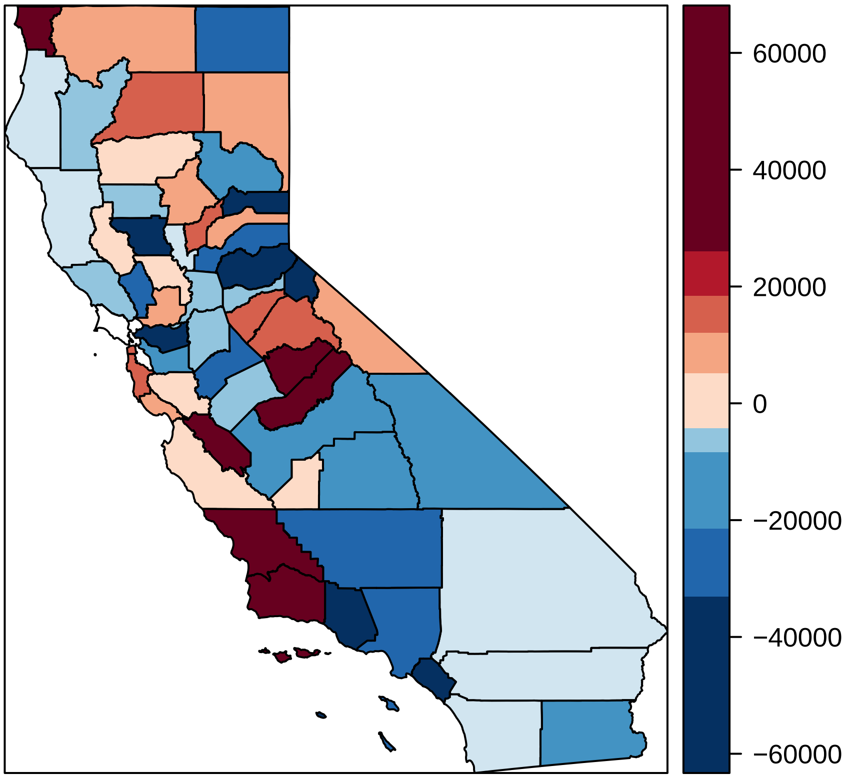
Based on the map, there seems to be a high clustering among people with the same race. As shown in the maps for White, Black, Asian, and Hispanic residents of California, areas with high population of a certain race are located next to each other. Looking at the maps closer, I also noticed that White population are widely distributed across each counties, especially compared to minorities in the state. Meanwhile, the Black residents in California are mostly located near the urban and coastal areas like Los Angeles and San Francisco. There also some Black residents remotely located from rural. Similar to African-Americans, most Asian residents are located near urban areas like Los Angeles and San Francisco. Suprisingly, unlike other minorities in the state, Hispanics are mostly located in the central part of the state with some noticeable presence in the southern region of the state. Areas with high Hispanic population seem to also have large family size since most areas with large family size are also located in the central part of the state. As for income level, residents with high median income are mostly located at Los Angeles and San Francisco and their surrounding counties.

```r
#Create new variables
hh$fAsian <- hh$Asian / hh$Population
hh$fBlack <- hh$Black / hh$Population
hh$age <- 2000 - hh$yearBuilt

#Equation to be estimated
f1 <- houseValue ~ age + nBedrooms + medHHinc + fAsian + fBlack

##OLS regression
m1 <- lm(f1, data=hh)
summary(m1)
```

3. In line 54 you have the equation that might help us to estimate the value of house. Modify the equation by using your results or observations in previous question and estimate the model by OLS. Interpret your results.

```
##
## Call:
## lm(formula = f1, data = hh)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -63294 -16639   2885  16989  68112
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.142e+05  8.419e+04   1.357  0.18067
## age          3.550e+03  1.031e+03   3.443  0.00114 **
## nBedrooms   -1.661e+05  3.046e+04  -5.452 1.39e-06 ***
## medHHinc     9.337e+00  4.981e-01  18.745  < 2e-16 ***
## fAsian      -1.850e+05  1.130e+05  -1.636  0.10780
## fBlack      -4.769e+05  1.369e+05  -3.485  0.00101 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30360 on 52 degrees of freedom
## Multiple R-squared:  0.9343, Adjusted R-squared:  0.928
## F-statistic: 147.9 on 5 and 52 DF,  p-value: < 2.2e-16
```

```r
hh$residuals <- residuals(m1)
bk8 <- quantile(hh$residuals, 0:(grps-1)/(grps-1), na.rm=TRUE)
spplot(hh, "residuals", at=bk8, col.regions=rev(brewer.pal(grps, "RdBu")), col="black")
```




```
library("spdep")
nb <- poly2nb(hh)
nb[[21]] <- sort(as.integer(c(nb[[21]], 38)))
nb[[38]] <- sort(as.integer(c(21, nb[[38]])))
nb
```

```
## Neighbour list object:
## Number of regions: 58
## Number of nonzero links: 278
## Percentage nonzero weights: 8.263971
## Average number of links: 4.793103
```

```
par(mai=c(0,0,0,0))
plot(hh)
plot(nb, coordinates(hh), col='red', lwd=2, add=TRUE)
```

```r
#list of neighbors
lw <- nb2listw(nb)

#MORANS indicator
moran.mc(hh$residuals, lw, 999)
```

```
##
##  Monte-Carlo simulation of Moran I
##
## data:  hh$residuals
## weights: lw
## number of simulations + 1: 1000
##
## statistic = 0.053516, observed rank = 793, p-value = 0.207
## alternative hypothesis: greater
```

Based on the regression table, high presence of African-American residents in an area can cause a negative impact to the price level of houses in the area, specifically lowering the value by -$476,914. This variable is also statistically significant and has the highest impact to the house price among all the variables included in the model. Surprisingly, the presence of Asian residents is not statistically significant to house price, but the explanatory variable still holds a considerable impact to the house level with a decrease in house price of -$184,970. Unlike the regression result shown during the lecture, the number of Bedrooms seem to have a negative effect to the house prices after controlling for Asian, African-American, and median income. Age of the property remains to be a statistically significant factor in the house price. Based on Moran's I, there seems to be no error caused by the presence of spatial autocorrelation since the p-value is greater than 0.05.

```r
m1s = lagsarlm(f1, data=hh, lw, tol.solve=1.0e-30)
summary(m1s)
```

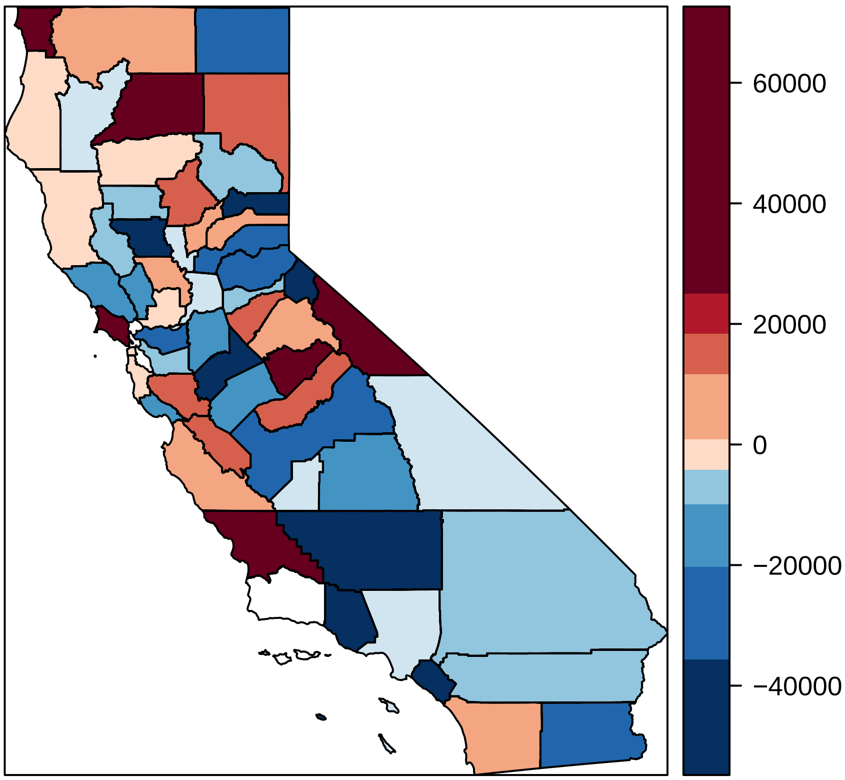
4. From line 83 you estimate the Spatial lag model by using your previous equation. Estimate your spatial model and compare with your results in the previous point.

```
##
## Call:lagsarlm(formula = f1, data = hh, listw = lw, tol.solve = 1e-30)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -54823 -15574  -1963  16551  72644
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##                Estimate  Std. Error z value  Pr(>|z|)
## (Intercept)  6.2174e+04  7.6009e+04  0.8180  0.413373
## age          2.8846e+03  9.0684e+02  3.1809  0.001468
## nBedrooms   -1.3354e+05  2.8763e+04 -4.6427 3.439e-06
## medHHinc     8.1130e+00  6.0205e-01 13.4757 < 2.2e-16
## fAsian      -1.8818e+05  9.8596e+04 -1.9086  0.056308
## fBlack      -5.1157e+05  1.2102e+05 -4.2273 2.365e-05
##
## Rho: 0.24623, LR test value: 8.7399, p-value: 0.0031131
## Asymptotic standard error: 0.080077
##     z-value: 3.0749, p-value: 0.0021054
## Wald statistic: 9.4553, p-value: 0.0021054
##
## Log likelihood: -673.3751 for lag model
## ML residual variance (sigma squared): 701010000, (sigma: 26477)
## Number of observations: 58
## Number of parameters estimated: 8
## AIC: 1362.8, (AIC for lm: 1369.5)
## LM test for residual autocorrelation
## test value: 0.88244, p-value: 0.34753
```

```r
hh$residuals <- residuals(m1s)
moran.mc(hh$residuals, lw, 999)
```

```
##
##  Monte-Carlo simulation of Moran I
##
## data:  hh$residuals
## weights: lw
## number of simulations + 1: 1000
##
## statistic = -0.073533, observed rank = 262, p-value = 0.738
## alternative hypothesis: greater
```

```r
brks <- quantile(hh$residuals, 0:(grps-1)/(grps-1), na.rm=TRUE)
p <- spplot(hh, "residuals", at=brks, col.regions=rev(brewer.pal(grps, "RdBu")), col="transparent")
print( p + layer(sp.polygons(hh)) )
```



Compared to the OLS Model, the Spatial Lag Model that controls for the spatial effect had a noticeable decrease in the impact of non-racial factors such as age, median income and number of bedrooms to the house price. However, the model exhibited an increase in racial factors, particularly in Asian and African-American population. In this model, Asian remains to be statistically insignificant at p-value < 0.05 in explaining the house price in California. There is also a statistically significant positive spatial effect (0.25) in the model based on Likelihood Ratio Test and Wald Statistic. Lastly, the Moran's I also suggest that the model has successfully controlled for the spatial effect since the p-value is greater than 0.05.

```r
m1e = errorsarlm(f1, data=hh, lw, tol.solve=1.0e-30)
summary(m1e)
```

5. Estimate the model by using a Spatial error model. Compare your results.

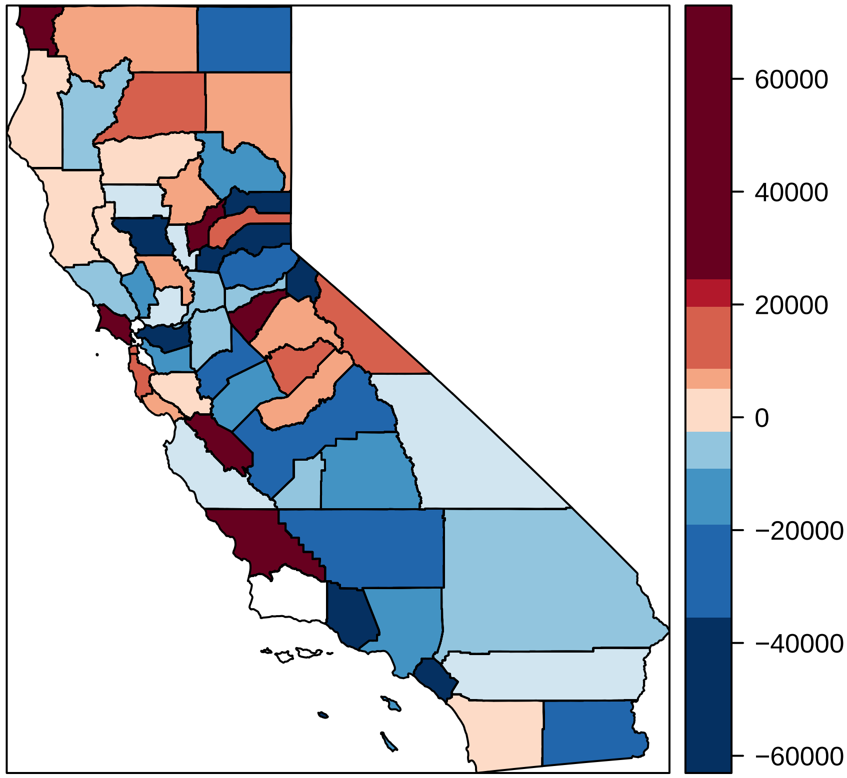
```
##
## Call:errorsarlm(formula = f1, data = hh, listw = lw, tol.solve = 1e-30)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -63052.11 -17241.34    975.04  18652.73  72995.36
##
## Type: error
## Coefficients: (asymptotic standard errors)
##                Estimate  Std. Error z value  Pr(>|z|)
## (Intercept)  1.2979e+05  7.9272e+04  1.6372  0.101585
## age          2.7185e+03  1.0189e+03  2.6680  0.007631
## nBedrooms   -1.6274e+05  2.8836e+04 -5.6437 1.665e-08
## medHHinc     9.2465e+00  5.1585e-01 17.9247 < 2.2e-16
## fAsian      -1.7471e+05  1.0452e+05 -1.6715  0.094626
## fBlack      -3.6772e+05  1.2840e+05 -2.8639  0.004184
##
## Lambda: 0.33701, LR test value: 1.1478, p-value: 0.28401
## Asymptotic standard error: 0.16496
##     z-value: 2.043, p-value: 0.041052
## Wald statistic: 4.1739, p-value: 0.041052
##
## Log likelihood: -677.1712 for error model
## ML residual variance (sigma squared): 788700000, (sigma: 28084)
## Number of observations: 58
## Number of parameters estimated: 8
## AIC: 1370.3, (AIC for lm: 1369.5)
```

```
hh$residuals_sem <- residuals(m1e)
moran.mc(hh$residuals_sem, lw, 999)
```

```
##
##  Monte-Carlo simulation of Moran I
##
## data:  hh$residuals_sem
## weights: lw
## number of simulations + 1: 1000
##
## statistic = -0.0048002, observed rank = 563, p-value = 0.437
## alternative hypothesis: greater
```

```
brks10 <- quantile(hh$residuals_sem, 0:(grps-1)/(grps-1), na.rm=TRUE)
p9 <- spplot(hh, "residuals_sem", at=brks10, col.regions=rev(brewer.pal(grps, "RdBu")),
            col="transparent")
print( p9 + layer(sp.polygons(hh)) )
```



In this model, there is a noticeable decrease in the impact of age, Asian population, and African-American population to the house price after controlling for spatial effects in the error. The effects of number of bedrooms and median income slightly increased. Referring to the LR test and Wald statistic, the 0.34 spatial effect on errors is also statistically significant at < 0.5. The Moran's I also suggest that the model has successfully controlled for the spatial effect since the p-value is greater than 0.05. Comparing the SLM and SEC, the SLM also has a lower Akaike Information Criteria, therefore the SLM model better explains the datasets.