Drug Abuse

Oscar Briones Ramirez

2023-04-09

Reading in and Handling Shapefiles

1. Read in the drug abuse shapefile.

```
myShp <- st_read("DrugAbuse.shp")
```

```
## Reading layer `DrugAbuse' from data source
##   `/Users/oscar/Desktop/Winter2023/STAT469/DrugAbuse/DrugAbuse.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 49 features and 6 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7328 ymin: 24.95638 xmax: -66.96927 ymax: 49.37173
## Geodetic CRS:  NAD83
```

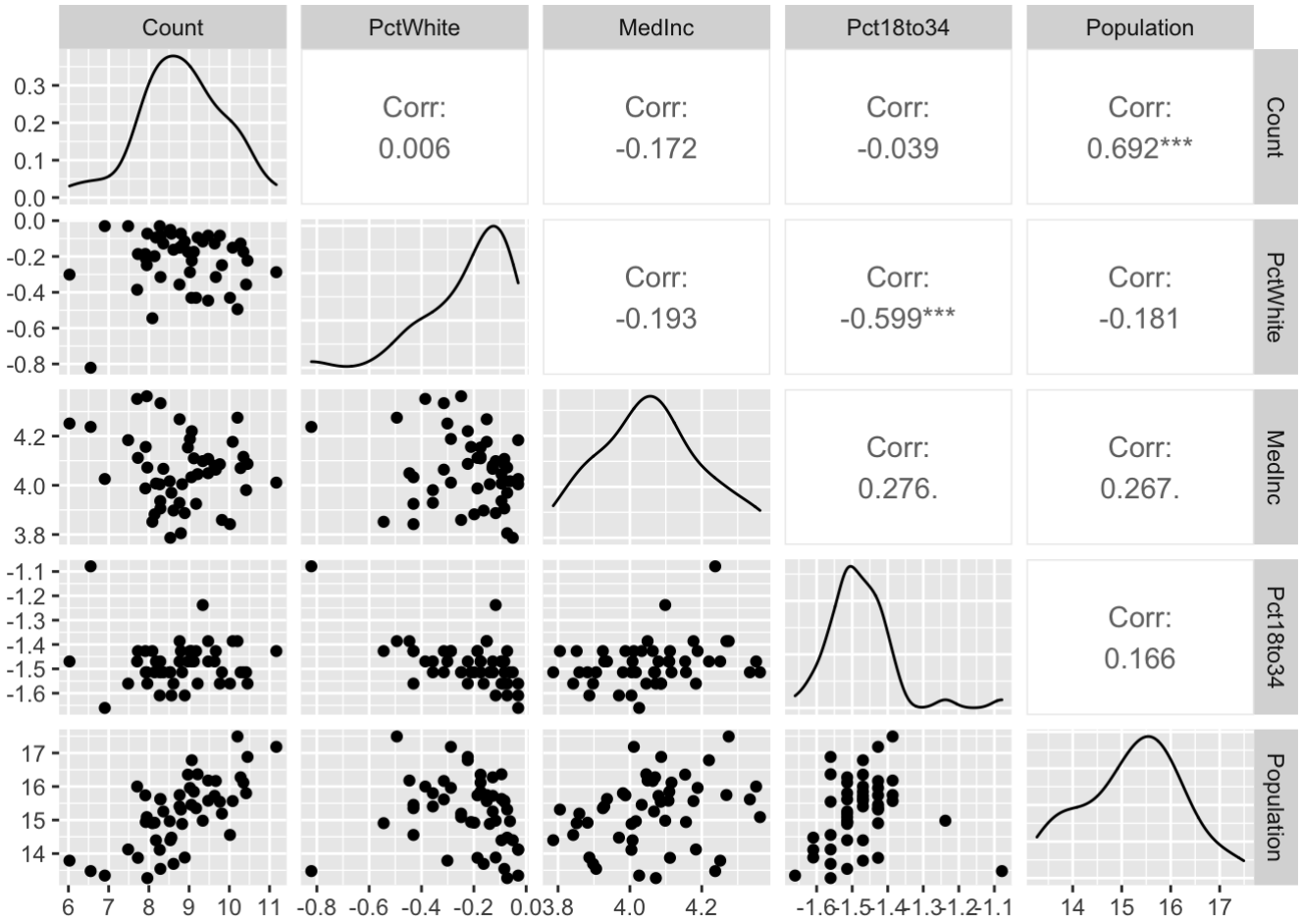
2. Create a dataframe from the shapefile that exclude the polygon information.

```
myShpDF <- data.frame(myShp)%>%
  dplyr::select(-geometry)
```

Exploratory Data Analysis

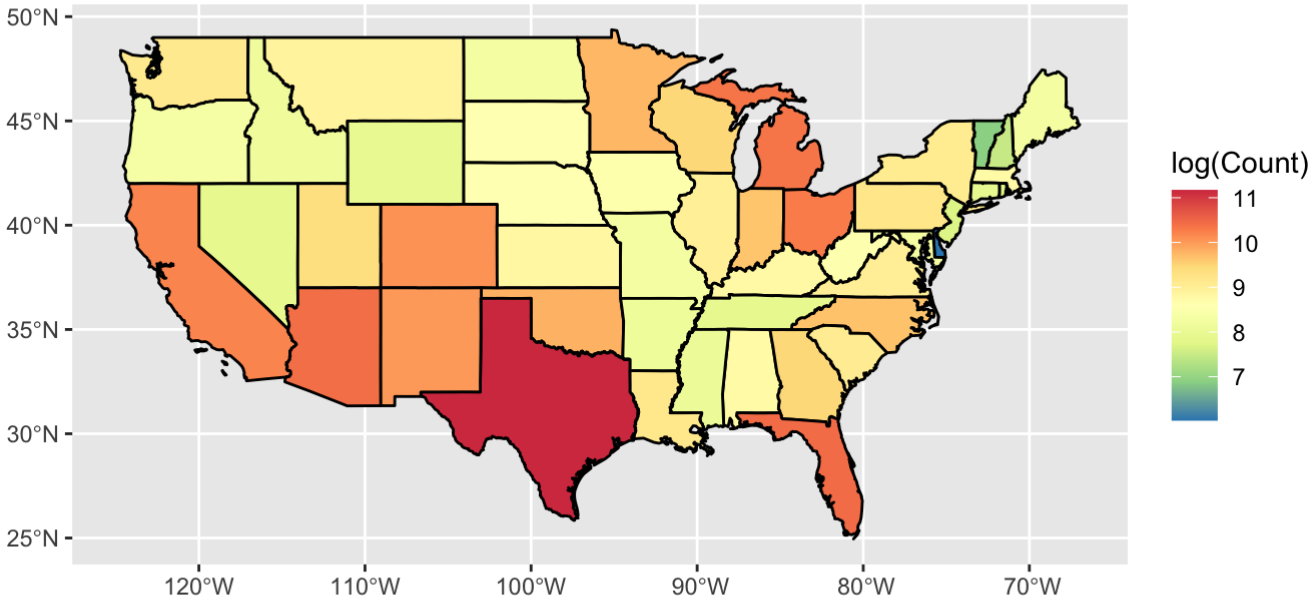
1.Create a pairs plot (ggpairs()) to assess the relationship between log(Count) and the explanatory variables (note we are using log(Count) here as the response because Poisson regression is log-linear).

```
ggpairs(log(myShpDF[,-1]))
```



2. Create a choropleth map of log(Count).

```
ggplot(data=myShp) +  geom_sf(mapping=aes(fill=log(Count)), color="black") +  scale_fill
_distiller(palette="Spectral")
```



3.Fit a lm() of log(Count) using Population, PctWhite, MedInc and Pct18to34 as explanatory variables. Perform a Moran's I test on the residuals to see if there is spatial correlation in the residuals.

```
drug.lm <- lm(formula=log(Count)~Population + PctWhite + MedInc + Pct18to34, data=myShpD
F)

moran.test(x=drug.lm$residuals, listw=nb2listw(poly2nb(myShp)))
```

```
##
##  Moran I test under randomisation
##
## data:  drug.lm$residuals
## weights: nb2listw(poly2nb(myShp))
##
## Moran I statistic standard deviate = 2.5921, p-value = 0.004769
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##       0.230238791      -0.020833333       0.009381812
```

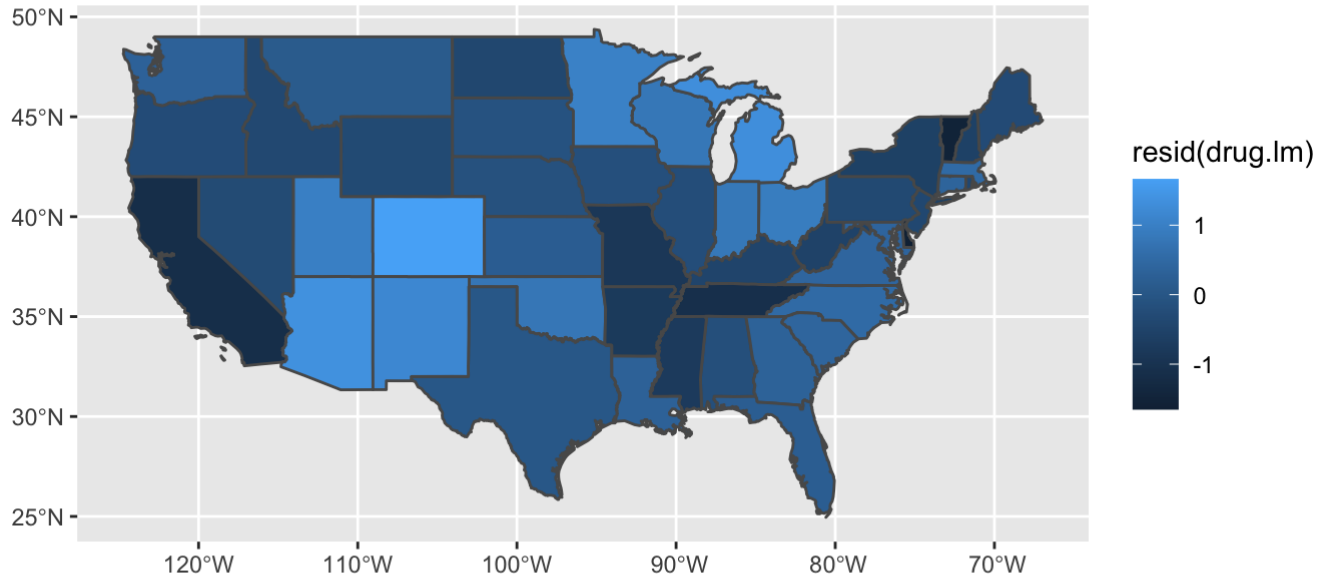
4. Perform a Geary's C test on your residuals from #3 above to double check if there is spatial correlation in the residuals.

```
geary.test(x=drug.lm$residuals, listw=nb2listw(poly2nb(myShp)))
```

```
##
##  Geary C test under randomisation
##
## data:  drug.lm$residuals
## weights: nb2listw(poly2nb(myShp))
##
## Geary C statistic standard deviate = 2.5474, p-value = 0.005427
## alternative hypothesis: Expectation greater than statistic
## sample estimates:
## Geary C statistic       Expectation          Variance
##        0.74378512         1.00000000        0.01011625
```

5. Map the residuals from the lm() fit to see if there is spatial correlation.

```
ggplot(data=myShp) +
  geom_sf(mapping=aes(fill=resid(drug.lm)))
```



Defining Spatial Basis Functions

1. Create the adjacency matrix.

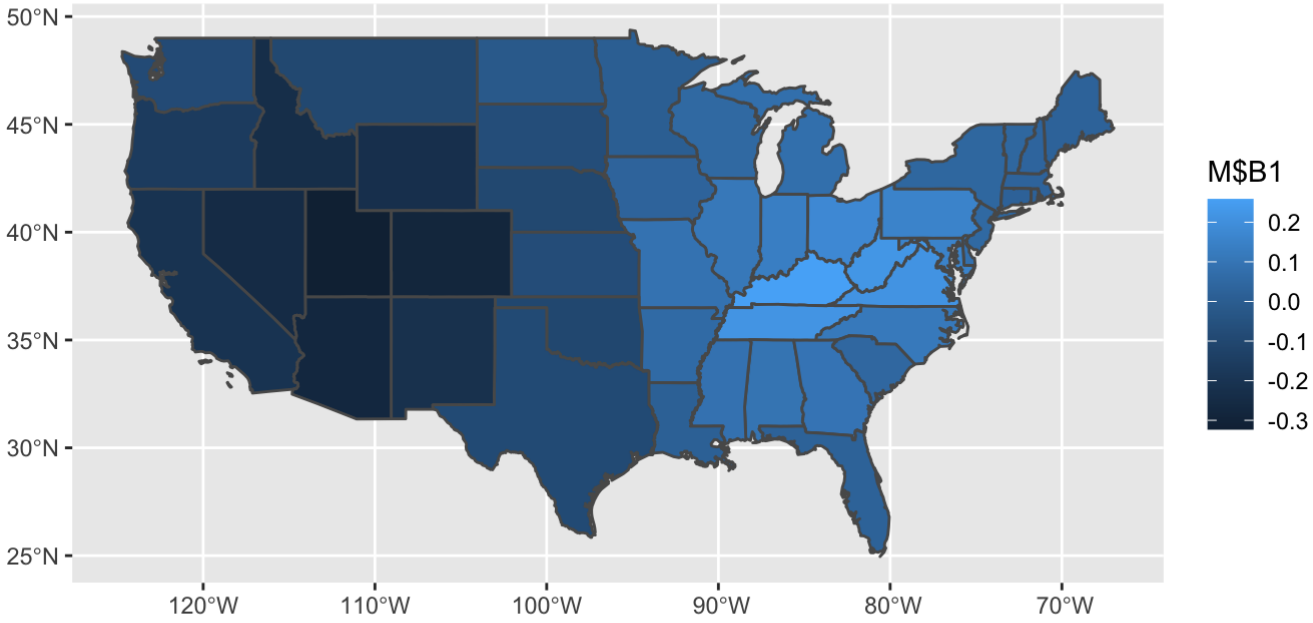
```
A <- nb2mat(poly2nb(myShp), style="B")
```

2. Create the Moran spatial basis and plot the first basis in a chloropleth map.

```
X <- model.matrix(object=log(Count)~Population + PctWhite + MedInc + Pct18to34, data=myS
hpDF)

M <- moranBasis(X, A, tol=0.95)

ggplot(data=myShp) +
  geom_sf(mapping=aes(fill=M$B1))
```



3. Merge the Moran spatial bases into your myShpDF data frame for use in fitting models later.

```
myShpDF <- bind_cols(myShpDF, M)
```

Spatial GLM Model Fitting

1. Fit a spatial GLM model with Count as the response and using PctWhite, MedInc, Population, Pct18to34 AND your spatial bases as explanatory variables. Print a summary() of the model to see your coefficient table.

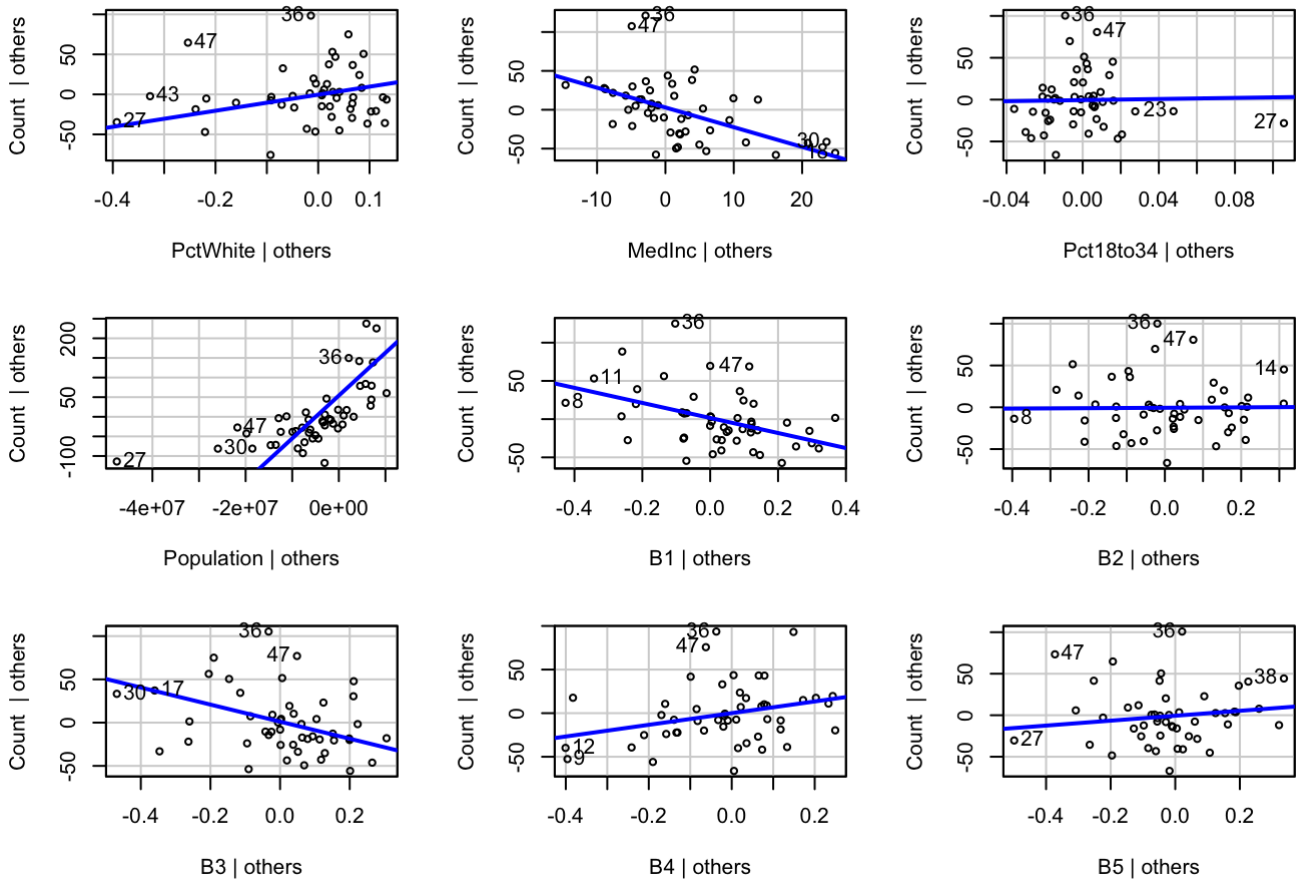
```
drug.glm <- glm(formula=Count~. - State, data=myShpDF, family=poisson)
summary(drug.glm)
```

```
##
## Call:
## glm(formula = Count ~ . - State, family = poisson, data = myShpDF)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -71.238  -23.869   -1.849   11.556   90.700
##
## Coefficients:
##               Estimate Std. Error  z value Pr(>|z|)
## (Intercept)  9.856e+00  2.699e-02  365.161  < 2e-16 ***
## PctWhite     6.351e-01  1.798e-02   35.327  < 2e-16 ***
## MedInc      -3.299e-02  2.491e-04 -132.425  < 2e-16 ***
## Pct18to34   -1.863e-01  9.051e-02   -2.059  0.03954 *
## Population   8.025e-08  1.877e-10  427.517  < 2e-16 ***
## B1          -1.202e+00  1.098e-02 -109.469  < 2e-16 ***
## B2          -1.085e-02  1.188e-02   -0.914  0.36089
## B3          -9.081e-01  1.017e-02  -89.267  < 2e-16 ***
## B4           9.765e-01  1.374e-02   71.047  < 2e-16 ***
## B5           1.921e-01  1.062e-02   18.084  < 2e-16 ***
## B6           1.199e-02  1.143e-02    1.049  0.29437
## B7           3.149e+00  1.123e-02  280.304  < 2e-16 ***
## B8          -3.388e-02  1.232e-02   -2.750  0.00595 **
## B9           3.500e-01  1.195e-02   29.282  < 2e-16 ***
## B10          1.164e+00  1.010e-02  115.231  < 2e-16 ***
## B11         -1.620e-01  1.207e-02  -13.423  < 2e-16 ***
## B12          1.128e-01  1.173e-02    9.619  < 2e-16 ***
## B13         -1.264e+00  1.258e-02 -100.454  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 496777  on 48  degrees of freedom
## Residual deviance:  53329  on 31  degrees of freedom
## AIC: 53887
##
## Number of Fisher Scoring iterations: 5
```

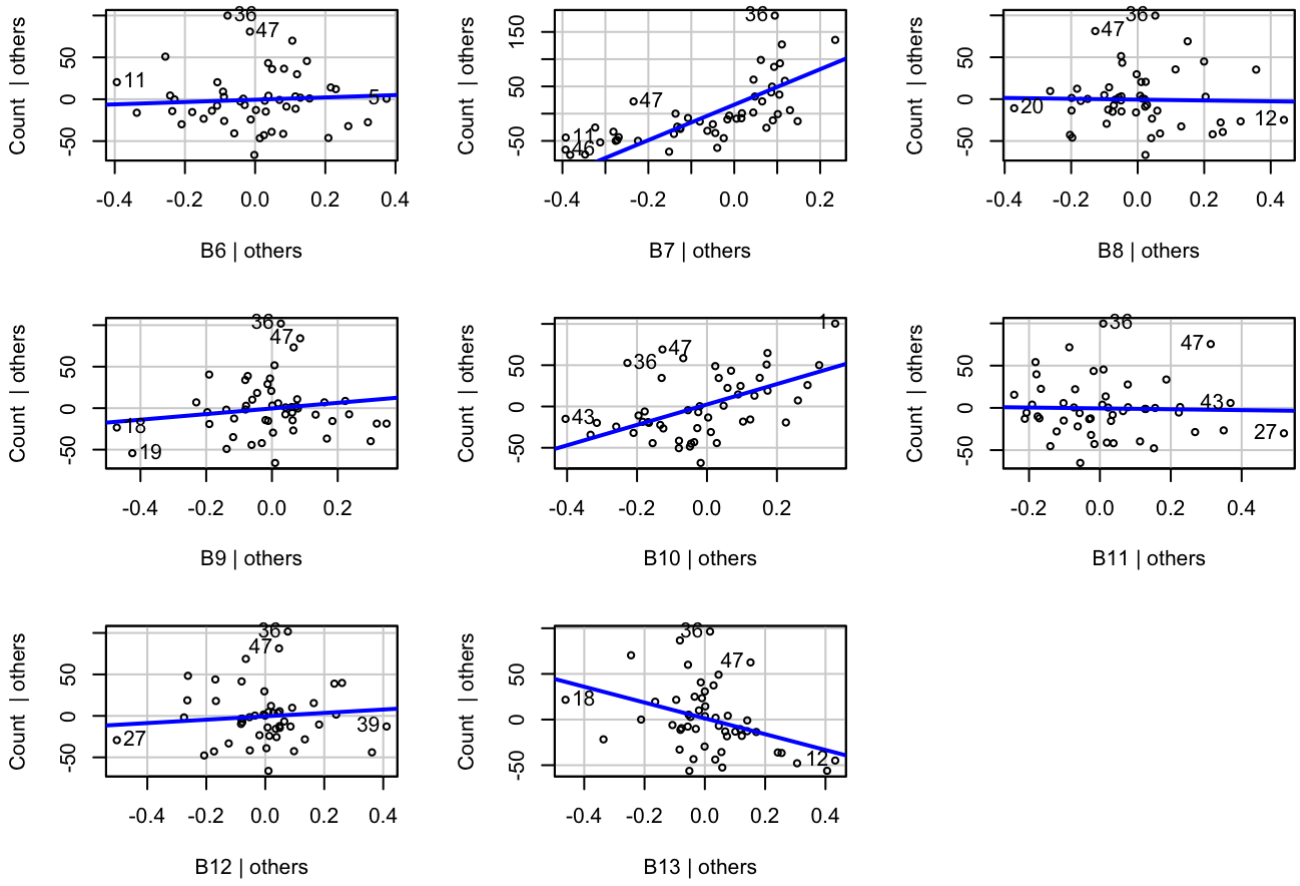
Validating Spatial MLR Model Assumptions and Predictions

1.Check the assumption of linearity using added-variable plots.

```
avPlots(drug.glm, ask=FALSE)
```



Added-Variable Plots



2. Check the assumption of independence by decorrelating residuals and performing Moran's I or Geary's C tests to make sure there is no more spatial correlation.

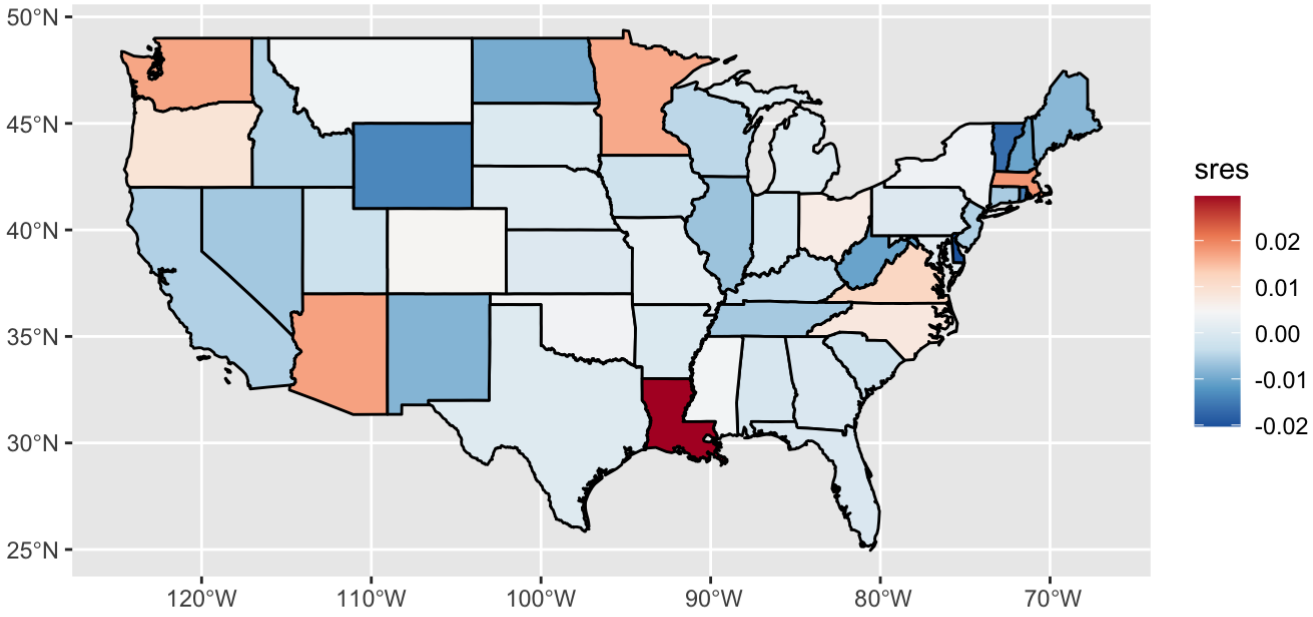
```
source("stdres.gls.R")
sres <- stdres.gls(drug.glm)

moran.test(x=sres, listw=nb2listw(poly2nb(myShp)))
```

```
##
##  Moran I test under randomisation
##
## data:  sres
## weights: nb2listw(poly2nb(myShp))
##
## Moran I statistic standard deviate = -1.2132, p-value = 0.8875
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      -0.136699956      -0.020833333       0.009120747
```

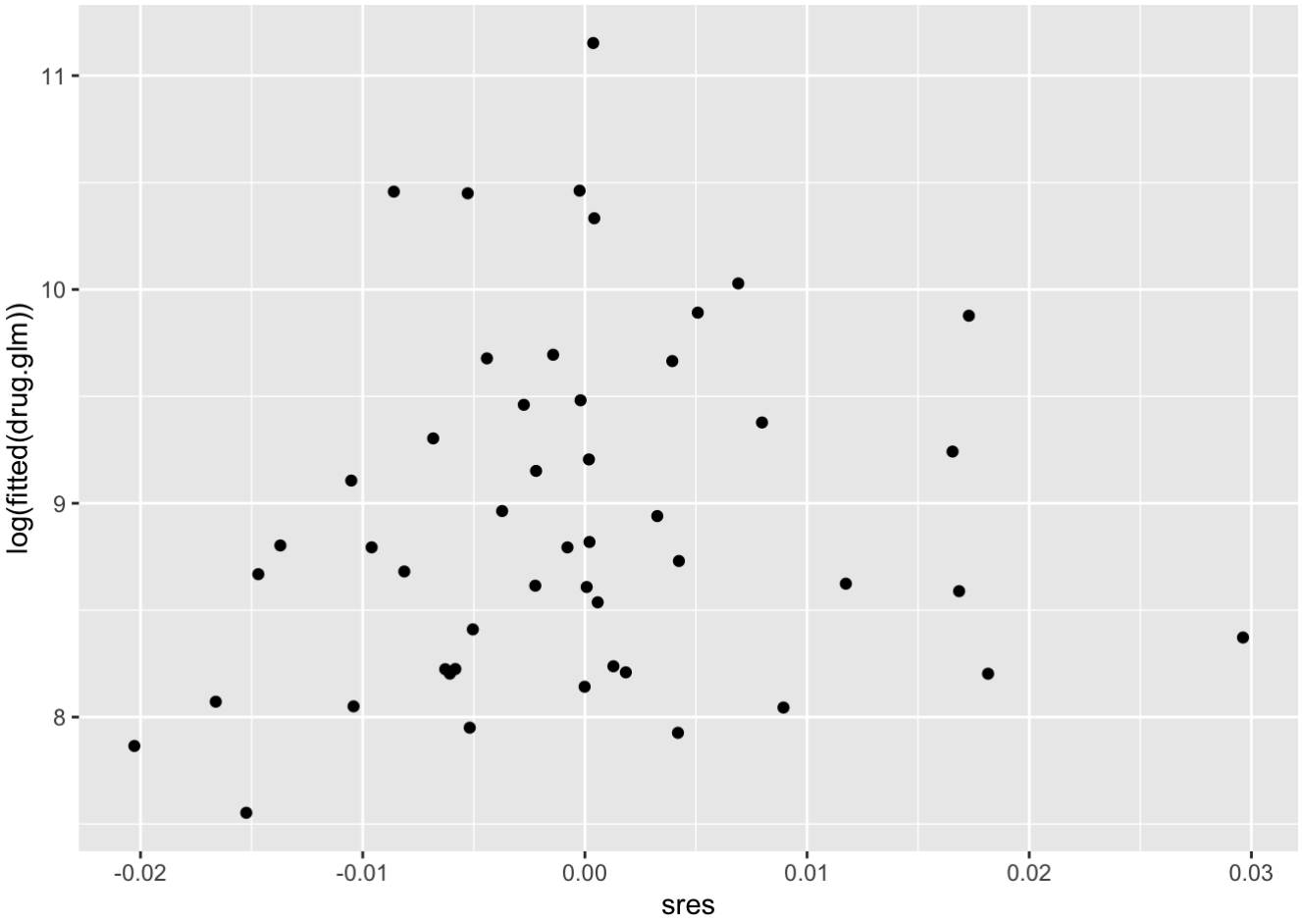
3.Draw a choropleth map of the standardized and decorrelated residuals to visually verify that the residuals are no longer spatially correlated.

```
ggplot(data=myShp) +
   geom_sf(mapping=aes(fill=sres), color="black") +  scale_fill_distiller(palette="RdB
u")
```



4. Check the assumption of equal variance by plotting the standardized and decorrelated residuals vs. the log(fitted values).

```
ggplot(mapping=aes(y=log(fitted(drug.glm)), x=sres)) + geom_point()
```



Statistical Inference

1. Print out the summary of the GLM model fit and identify the estimates and 95% confidence intervals of your explanatory variables.

```
summary(drug.glm)
```

```
##
## Call:
## glm(formula = Count ~ . - State, family = poisson, data = myShpDF)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -71.238  -23.869   -1.849   11.556   90.700
##
## Coefficients:
##               Estimate Std. Error  z value Pr(>|z|)
## (Intercept)  9.856e+00  2.699e-02  365.161  < 2e-16 ***
## PctWhite     6.351e-01  1.798e-02   35.327  < 2e-16 ***
## MedInc      -3.299e-02  2.491e-04 -132.425  < 2e-16 ***
## Pct18to34   -1.863e-01  9.051e-02   -2.059  0.03954 *
## Population   8.025e-08  1.877e-10  427.517  < 2e-16 ***
## B1          -1.202e+00  1.098e-02 -109.469  < 2e-16 ***
## B2          -1.085e-02  1.188e-02   -0.914  0.36089
## B3          -9.081e-01  1.017e-02  -89.267  < 2e-16 ***
## B4           9.765e-01  1.374e-02   71.047  < 2e-16 ***
## B5           1.921e-01  1.062e-02   18.084  < 2e-16 ***
## B6           1.199e-02  1.143e-02    1.049  0.29437
## B7           3.149e+00  1.123e-02  280.304  < 2e-16 ***
## B8          -3.388e-02  1.232e-02   -2.750  0.00595 **
## B9           3.500e-01  1.195e-02   29.282  < 2e-16 ***
## B10          1.164e+00  1.010e-02  115.231  < 2e-16 ***
## B11         -1.620e-01  1.207e-02  -13.423  < 2e-16 ***
## B12          1.128e-01  1.173e-02    9.619  < 2e-16 ***
## B13         -1.264e+00  1.258e-02 -100.454  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 496777  on 48  degrees of freedom
## Residual deviance:  53329  on 31  degrees of freedom
## AIC: 53887
##
## Number of Fisher Scoring iterations: 5
```

```
confint(drug.glm,level=.95)
```

```
## Waiting for profiling to be done...
```

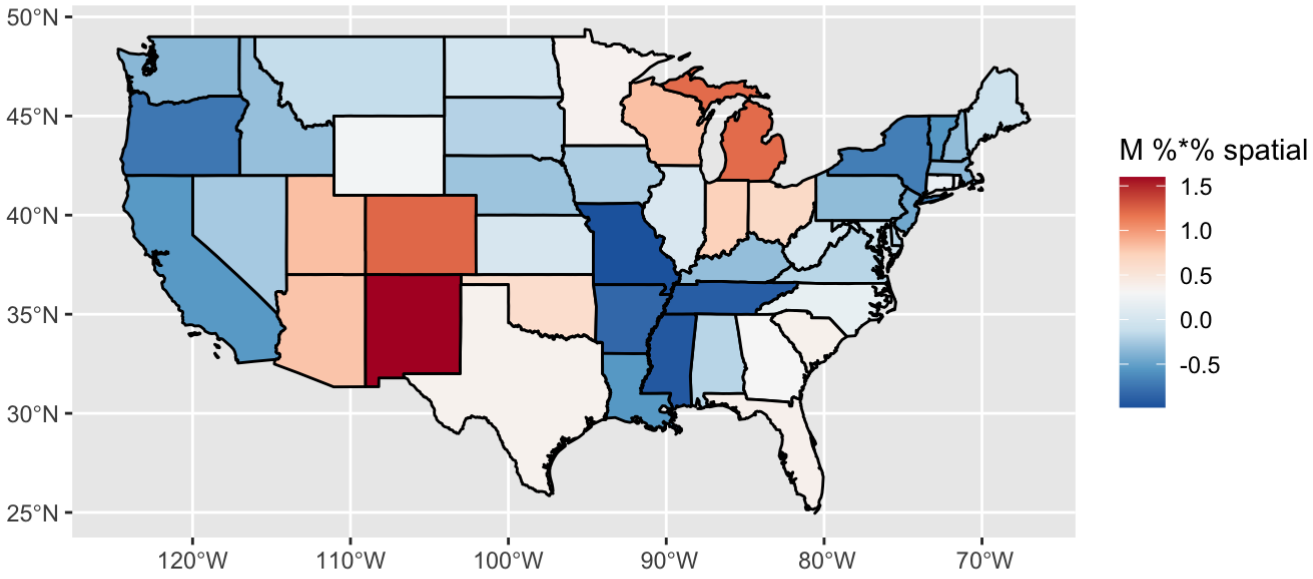
```
##                     2.5 %        97.5 %
## (Intercept)  9.803483e+00  9.909289e+00
## PctWhite     5.999095e-01  6.703836e-01
## MedInc      -3.347494e-02 -3.249851e-02
## Pct18to34   -3.638322e-01 -9.054965e-03
## Population   7.988507e-08  8.062091e-08
## B1          -1.223213e+00 -1.180182e+00
## B2          -3.413037e-02  1.242783e-02
## B3          -9.280516e-01 -8.881737e-01
## B4           9.495625e-01  1.003439e+00
## B5           1.713144e-01  2.129622e-01
## B6          -1.041803e-02  3.439184e-02
## B7           3.127042e+00  3.171080e+00
## B8          -5.803710e-02 -9.744216e-03
## B9           3.265731e-01  3.734271e-01
## B10          1.144093e+00  1.183686e+00
## B11         -1.856086e-01 -1.383116e-01
## B12          8.983116e-02  1.358101e-01
## B13         -1.288454e+00 -1.239138e+00
```

2. Create a chloropleth map of the spatially correlated residuals (just the bʼiϑ part) to identify states that, after accounting for the explanatory variables, have an elevated level of risk.

```
spatial <- as.matrix(coef(drug.glm)[-(1:5)])

M <- as.matrix(M)


ggplot(data=myShp) +
   geom_sf(mapping=aes(fill=M %*% spatial), color="black") +  scale_fill_distiller(palet
te="RdBu")
```


50°N
45°N
40°N
35°N
30°N
25°N
120°W
110°W
100°W
90°W
80°W
70°W
M %*% spatial
1.5
1.0
0.5
0.0
-0.5