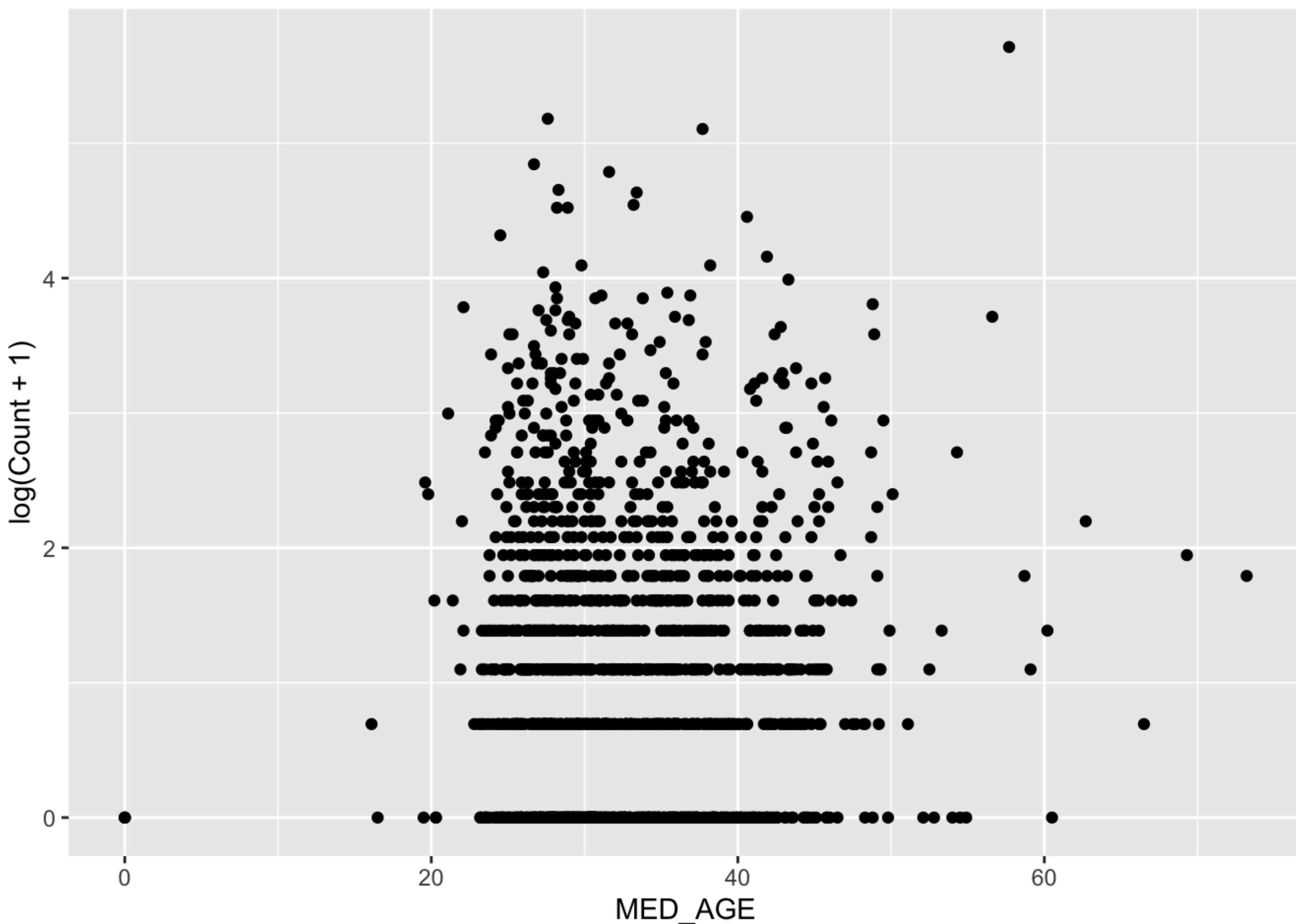
Heat-related Mortality

Oscar Briones Ramirez

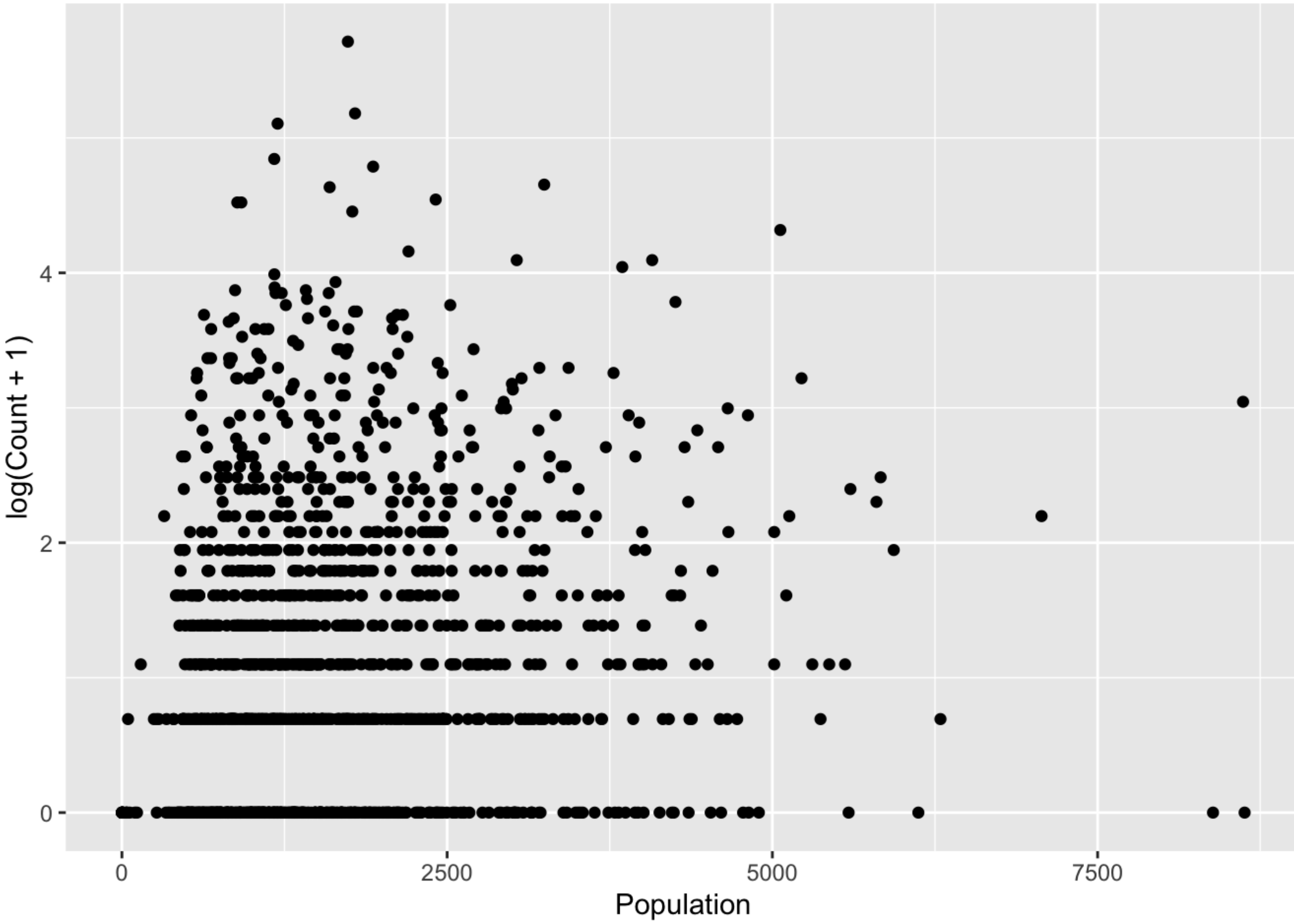
2023-04-12

1. Create exploratory plots of the data by looking at the relationship between log(Count+1) (the response variable) and a few of the explanatory variables. Comment on any general relationships you see from the data. Note that we explore log(Count+1) here because Poisson regression is log-linear and we arbitrarily add one to the counts because log(0) = -Inf .

```
ggplot(data=myShpDF, mapping=aes(x=MED_AGE, y=log(Count+1))) + geom_point()
```

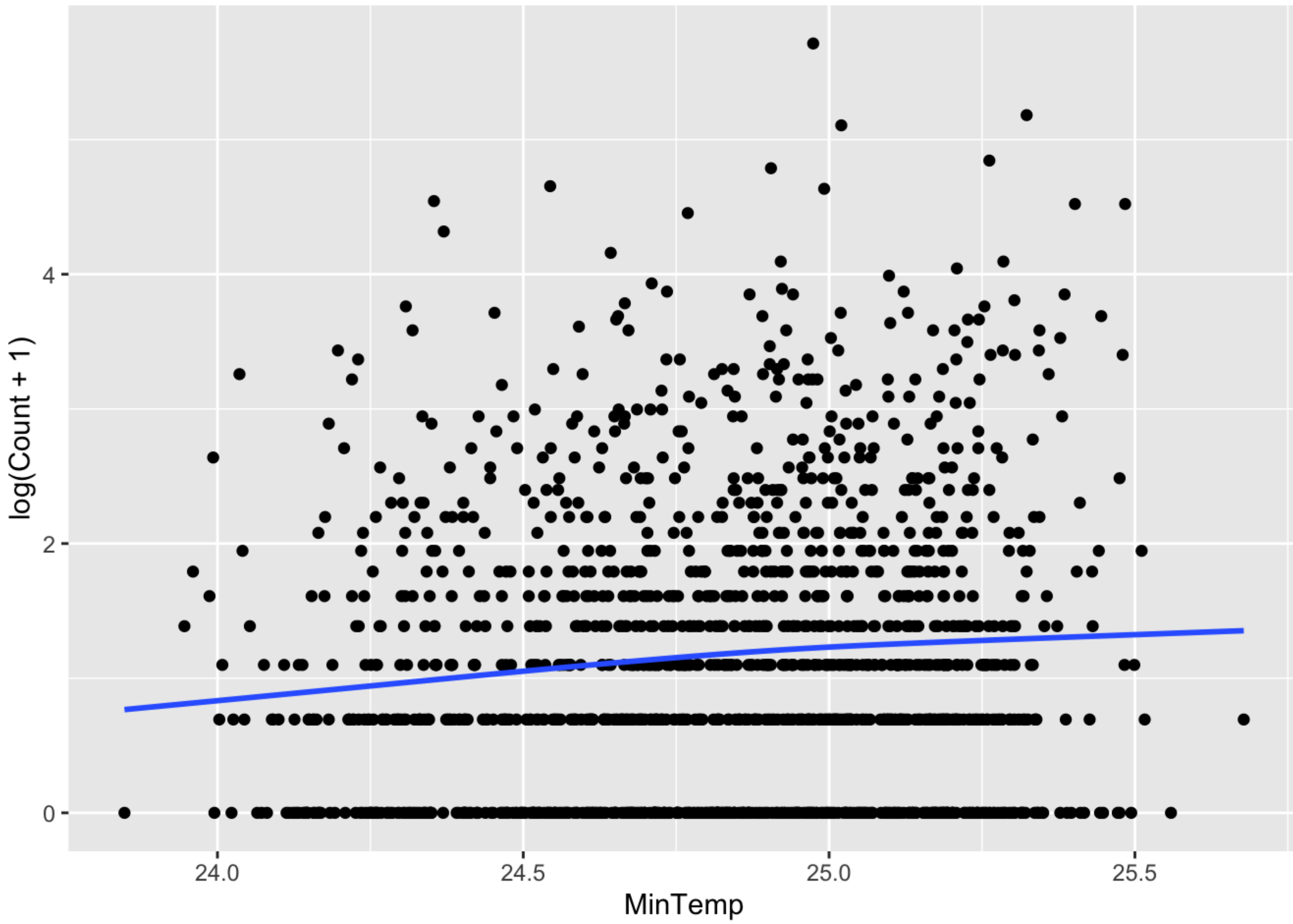



```
ggplot(data=myShpDF, mapping=aes(x=Population, y=log(Count+1))) + geom_point()
```




```
ggplot(data=myShpDF, mapping=aes(x=MinTemp, y=log(Count+1))) + geom_point() + geom_smoot
h(se=FALSE)
```

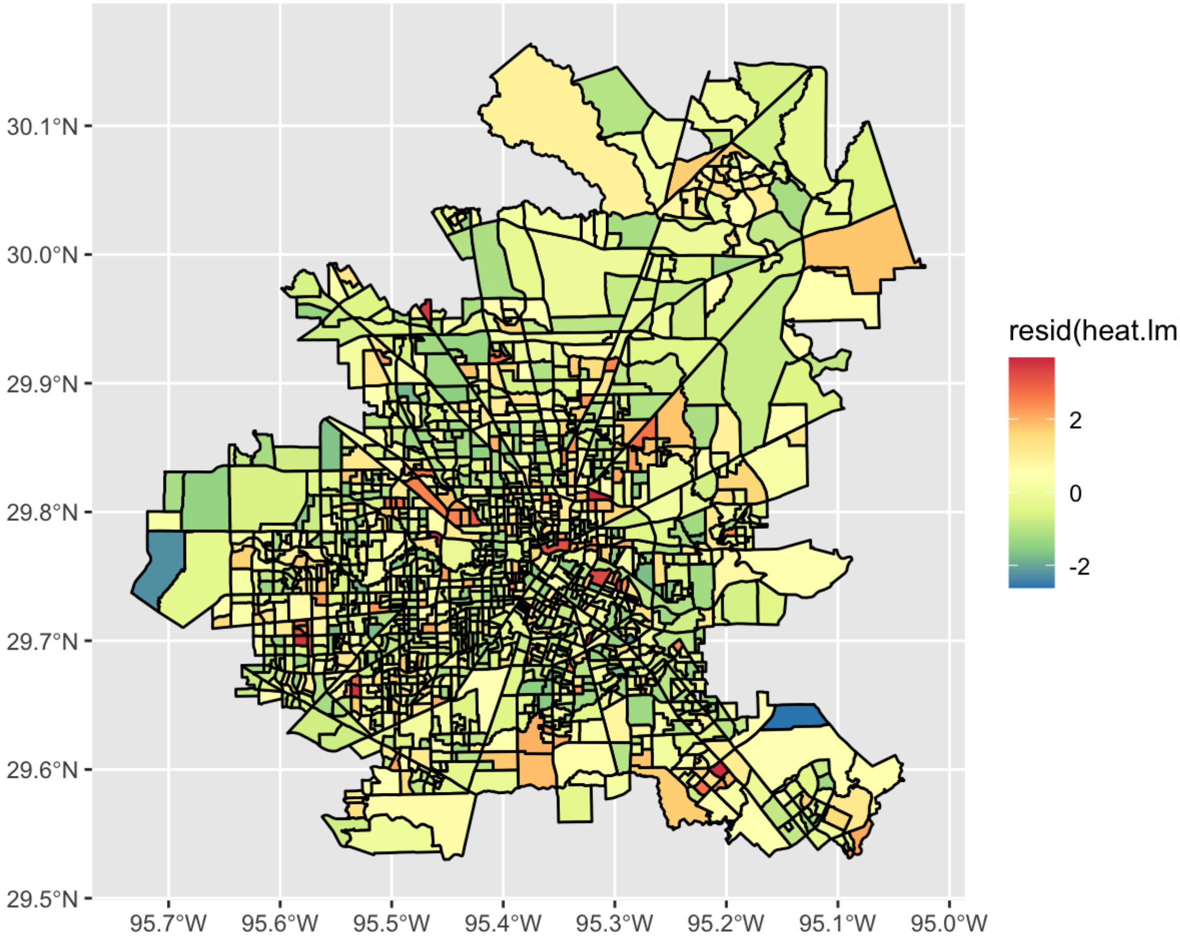
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



- We can see that morbidities seem to be highest around groups where the median age is about 30.
- We can also see that count is high where population is around 1200
- Finally, we can also tell that morbidities increase slightly as the minimum summer temperature increases.

2. Fit an independent MLR model with a linear effect between log(Count+1) and all the explanatory variables. Explore the residuals to see if there is evidence of spatial correlation by mapping them and using a Moran's I or Geary's C test.

```
#LM
heat.lm <- lm(formula=log(Count+1)~., data=myShpDF)

#Residuals
ggplot(data=myShp) +
  geom_sf(mapping=aes(fill=resid(heat.lm)), color="black") +  scale_fill_distiller(palet
te="Spectral")
```




```
#Moran test
moran.test(x=heat.lm$residuals, listw=nb2listw(poly2nb(st_make_valid(myShp))))
```

```
##
##  Moran I test under randomisation
##
## data:  heat.lm$residuals
## weights: nb2listw(poly2nb(st_make_valid(myShp)))
##
## Moran I statistic standard deviate = 15.736, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      0.2338486649     -0.0006729475      0.0002221256
```

-We can see from the map of the residuals that there is correlation algon the areas of Houston.

-From the Moran test we can comfirm there is aerial correlation.

4. Fit your spatial GLM model using Moran basis function (use up to tol=.95 but you may be able to get away with less) and validate any assumptions you made to fit the model.

3. Write out a spatial GLM model for analyzing the mortality data in terms of parameters that includes spatial basis functions to capture the spatial correlation in the data. Explain and interpret at least 1 parameter associated with the model and, in very general terms, explain what the spatial basis functions are doing.

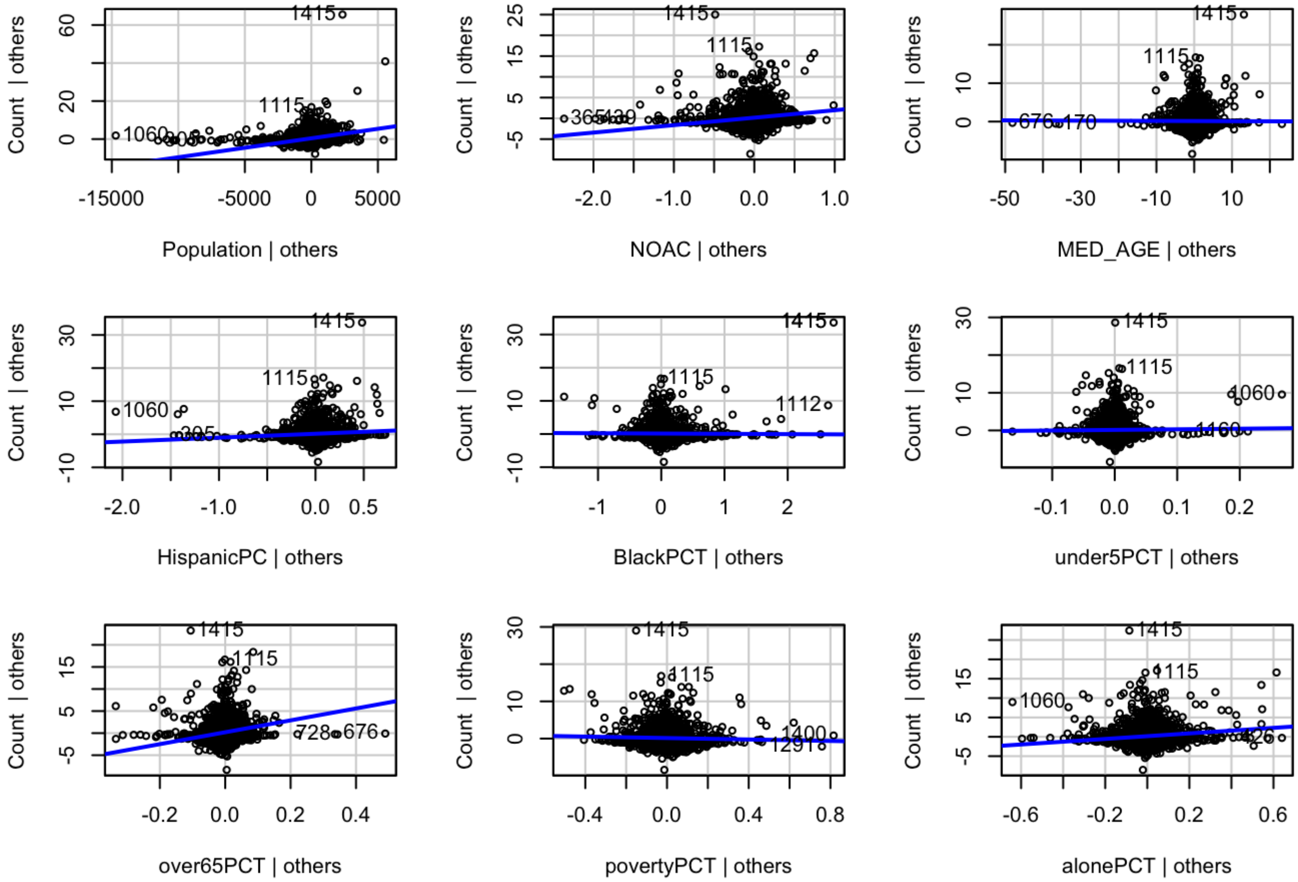
$$Y_i \overset{ind}{\sim} \text{Pois}(\lambda_i)$$
$$\log(\lambda_i) = \underbrace{\mathbf{x}_i'}_{\text{Covariates}} \boldsymbol{\beta} + \underbrace{\mathbf{b}_i'}_{\text{Spatial Bases}} \boldsymbol{\theta}$$

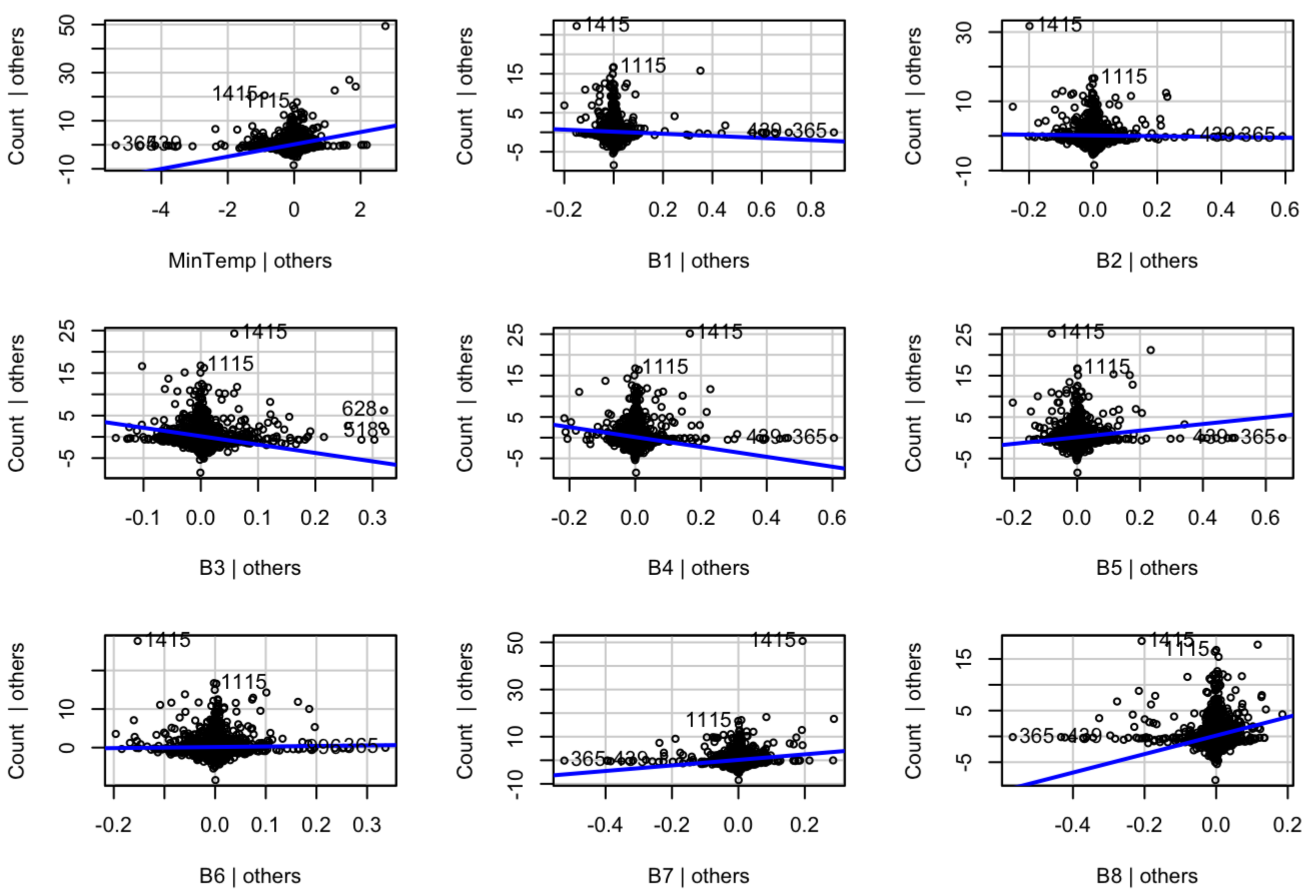
- β is the matrix containing the coefficients associated with each explanatory variable
- The spatial basis functions are additional parameters we insert in the model and tune in order to achieve a specific property that we want. They decrease with distance and tie observations together over space.

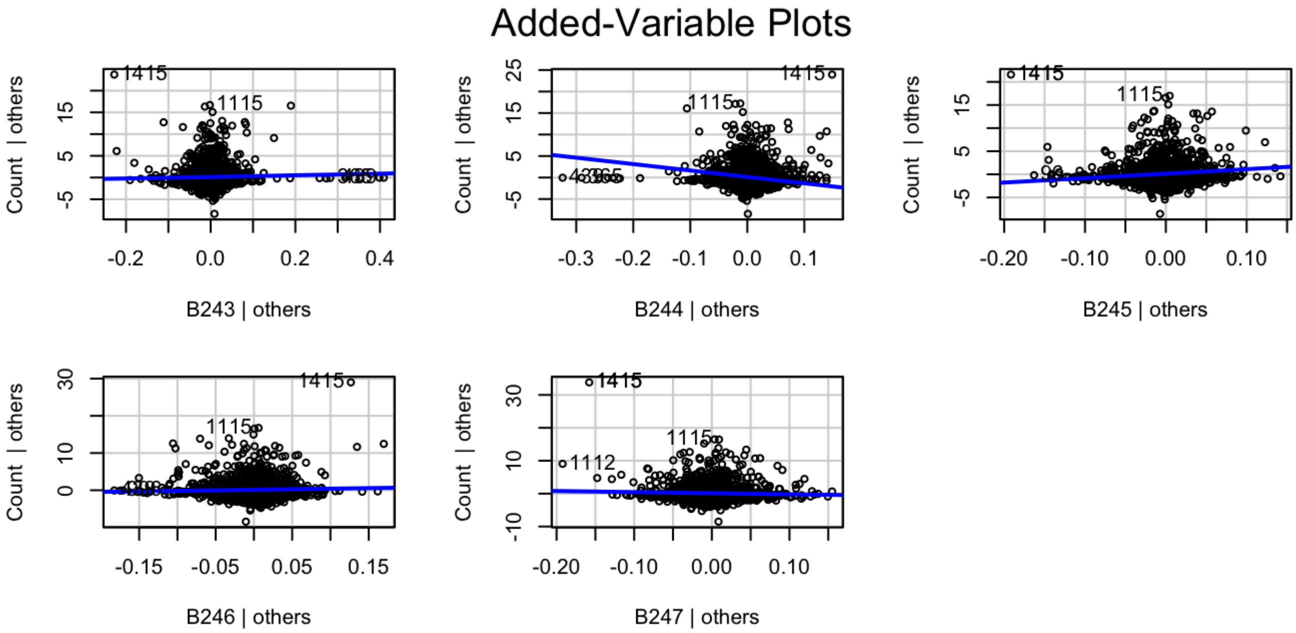
4. Fit your spatial GLM model using Moran basis function (use up to tol=.95 but you may be able to get away with less) and validate any assumptions you made to fit the model.

```r
#Moran basis GLM
A <- nb2mat(poly2nb(st_make_valid(myShp)), style="B")

X <- model.matrix(object=log(Count + 1)~., data=myShpDF)

M <- moranBasis(X, A, tol=.75)

myShpDF <- bind_cols(myShpDF, M)

heat.glm <- glm(formula=Count~., data=myShpDF, family=poisson)
heat.glm$coefficients
```

```
#Assumptions:

#linearity:
avPlots(heat.glm, ask=FALSE)
```


Count | others
Population | others
NOAC | others
MED_AGE | others
HispanicPC | others
BlackPCT | others
under5PCT | others
over65PCT | others
povertyPCT | others
alonePCT | others

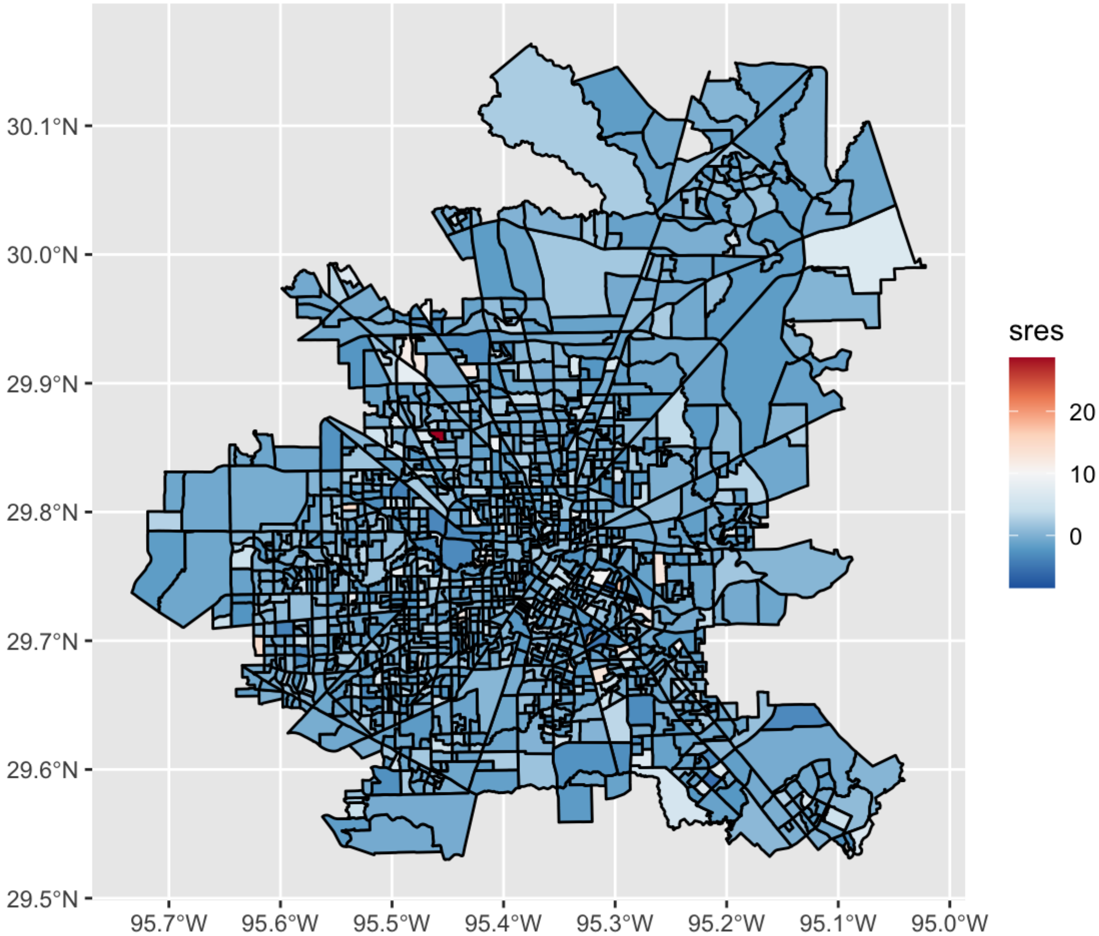

Count | others
MinTemp | others
B1 | others
B2 | others
B3 | others
B4 | others
B5 | others
B6 | others
B7 | others
B8 | others



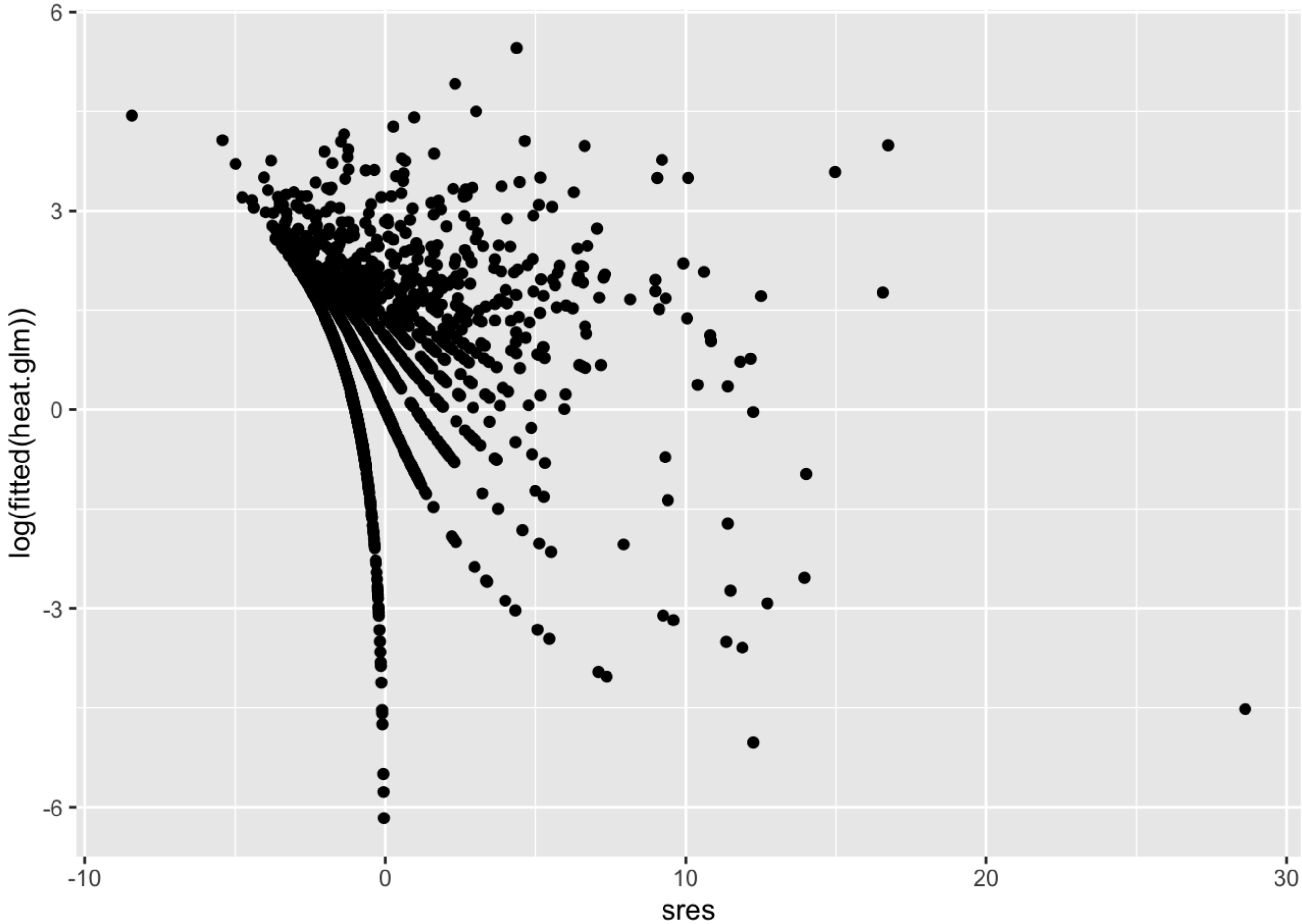

```
#Independence
sres <- resid(heat.glm, "pearson")

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

```
##
##  Moran I test under randomisation
##
## data:  sres
## weights: nb2listw(poly2nb(st_make_valid(myShp)))
##
## Moran I statistic standard deviate = -1.9471, p-value = 0.9742
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##     -0.0295628499     -0.0006729475      0.0002201430
```

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




```
#Equal variance
ggplot(mapping=aes(y=log(fitted(heat.glm)), x=sres)) + geom_point()
```



-After fitting the Moran Basis function GLM, we can see from the added variable plots that the linearity assumption is met. -We can also see from the Moran test that there is no longer correlation, and we have independence.

-The map of the standardized residuals shows that correlation present has been reduced.

-Finally, the plot of the fitted values vs. residuals shows there is equal variance.

5. Calculate confidence intervals for the effect of each explanatory variable included in your model (but NOT the basis functions). Draw conclusions about who is at greatest risk for heat-related mortality based on your estimated effects.

```
varnames <- names(coef(heat.glm))[1:11]

confint(heat.glm, varnames, level=.95)
```

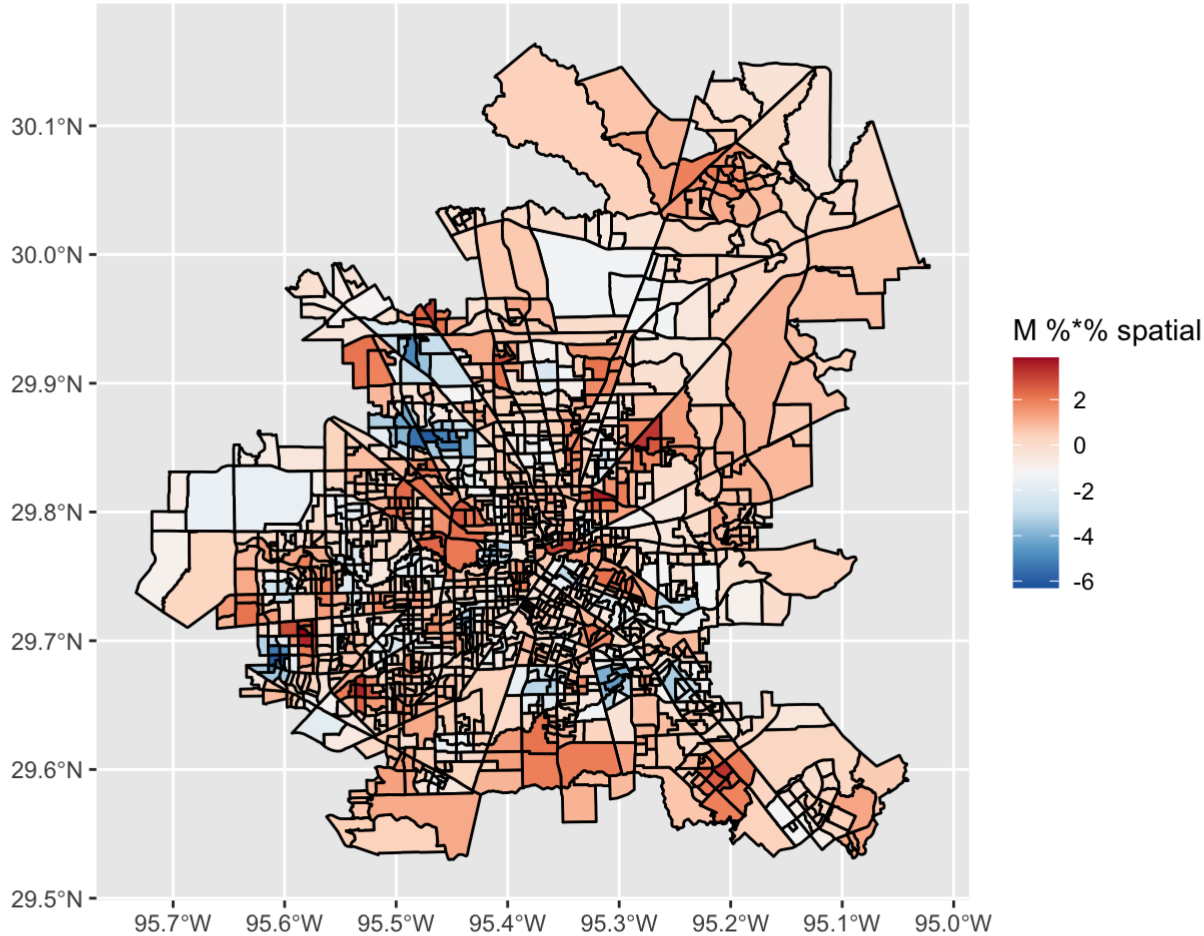
```
## Waiting for profiling to be done...
```

```
##                    2.5 %        97.5 %
## (Intercept) -2.188201e+01 -1.498755e+01
## Population   3.713962e-04  4.260959e-04
## NOAC         4.131714e-01  7.434439e-01
## MED_AGE     -1.616601e-02  7.982046e-03
## HispanicPC   3.814223e-01  8.587494e-01
## BlackPCT    -6.693084e-02  2.932209e-01
## under5PCT   -3.779673e+00 -1.092387e-01
## over65PCT    4.177763e+00  5.958354e+00
## povertyPCT  -5.117010e-01  1.451735e-01
## alonePCT     7.474321e-01  1.439371e+00
## MinTemp      5.891570e-01  8.622939e-01
```

After looking at the intervals of the explanatory variables, it looks like a few demographics are at higer risk of heat related mortality. We can see that places with higher population of Hispanic people, higher percentage of people ovr 65, living alone, no AC, or the minimum temperature is higher, the count of morbidities could increase.

6. Draw a map of the correlated residuals to try and reach conclusions about areas at risk of heat-related mortality not explained by your explanatory variables

```
spatial <- as.matrix(coef(heat.glm)[-(1:11)])

M <- as.matrix(M)

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



From the map of our correlated residuals, we can see that most areas around the metropolitan area have high risk of heat realted mortality due to the higher heat presence, and preventative measures should be considered for those areas.