

Fitting CAR and SAR Models

Lecture 20

Dr. Colin Rundel

Fitting areal models

Revised SAR Model

- Formula Model

$$y(s_i) = X_i \cdot \beta + \phi \sum_{j=1}^n D_{jj}^{-1} A_{ij} (y(s_j) - X_j \cdot \beta) + \epsilon_i$$

$$\epsilon \sim N(\mathbf{0}, \sigma^2 D^{-1})$$

- Joint Model

$$y \sim N \left(X\beta, (I - \phi D^{-1} A)^{-1} \sigma^2 D^{-1} \left((I - \phi D^{-1} A)^{-1} \right)^t \right)$$

Revised CAR Model

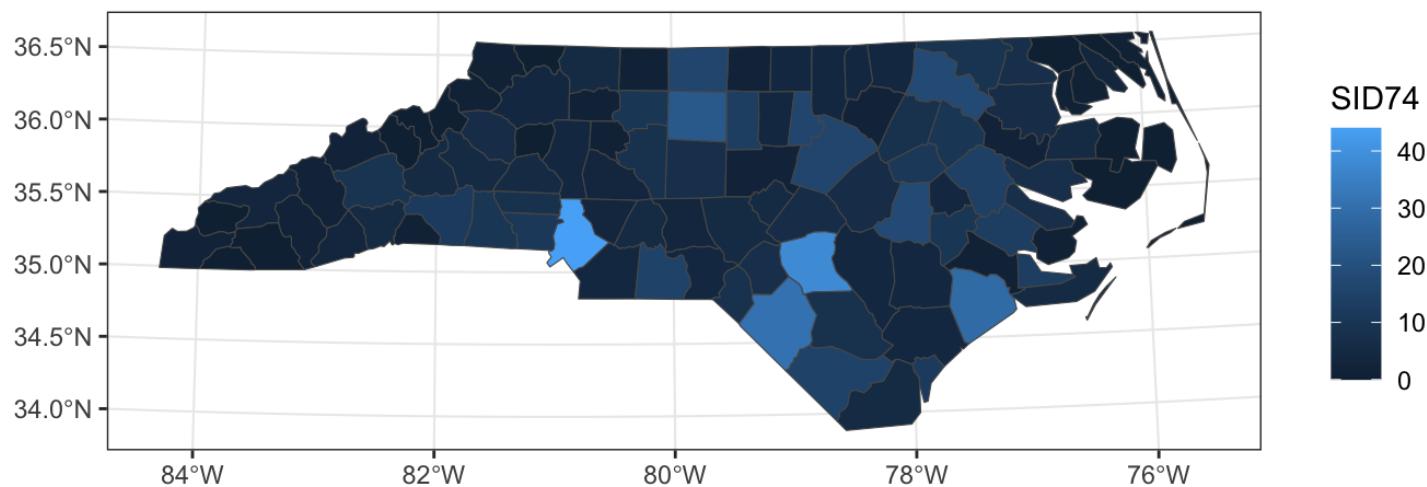
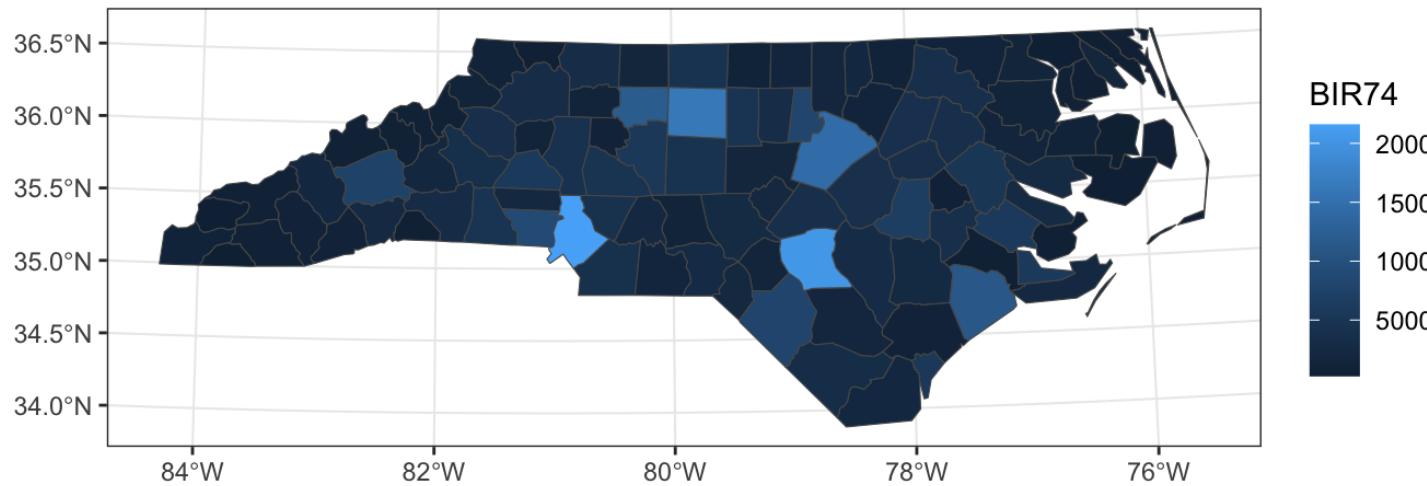
- Conditional Model

$$y(s_i) | y_{-s_i} \sim N \left(X_{\cdot i} \beta + \phi \sum_{j=1}^n \frac{A_{ij}}{D_{ii}} (y(s_j) - X_{\cdot j} \beta), \sigma^2 D_{ii}^{-1} \right)$$

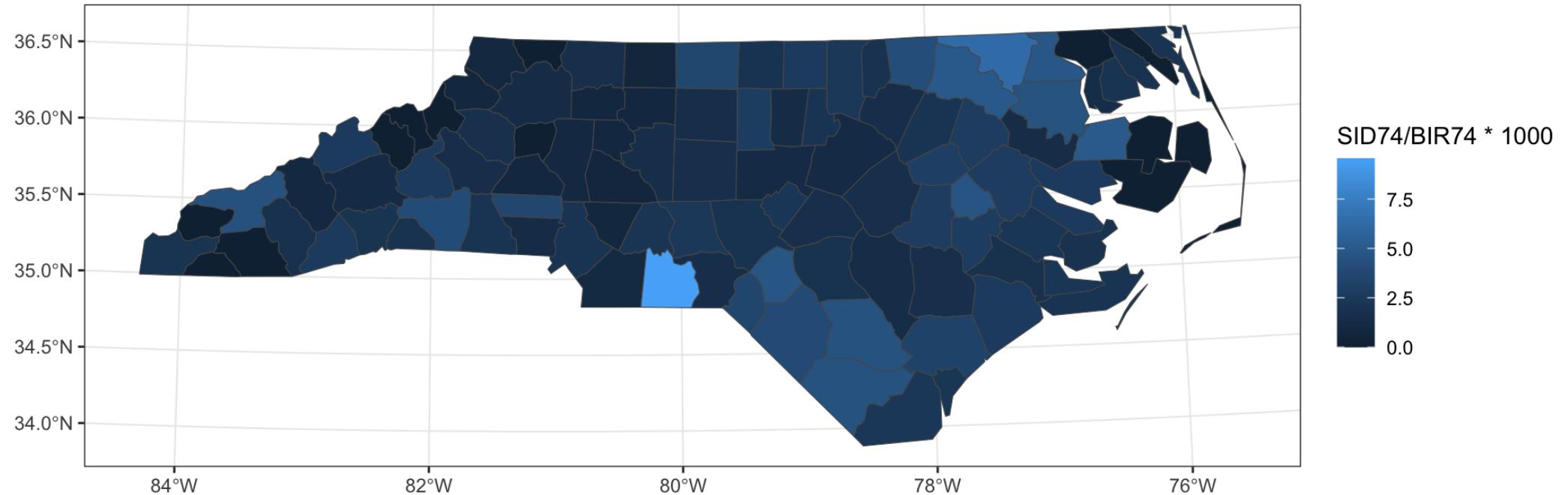
- Joint Model

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{D} - \phi\mathbf{A})^{-1})$$

Example - NC SIDS



```
1 ggplot() + geom_sf(data=nc, aes(fill=SID74/BIR74*1000))
```



Using `spdep` + `spatialreg`

```
1 library(spdep)
2
3 A = st_touches(nc, sparse=FALSE)
4 (listW = spdep::mat2listw(A))
```

Characteristics of weights list object:

Neighbour list object:

Number of regions: 100

Number of nonzero links: 490

Percentage nonzero weights: 4.9

Average number of links: 4.9

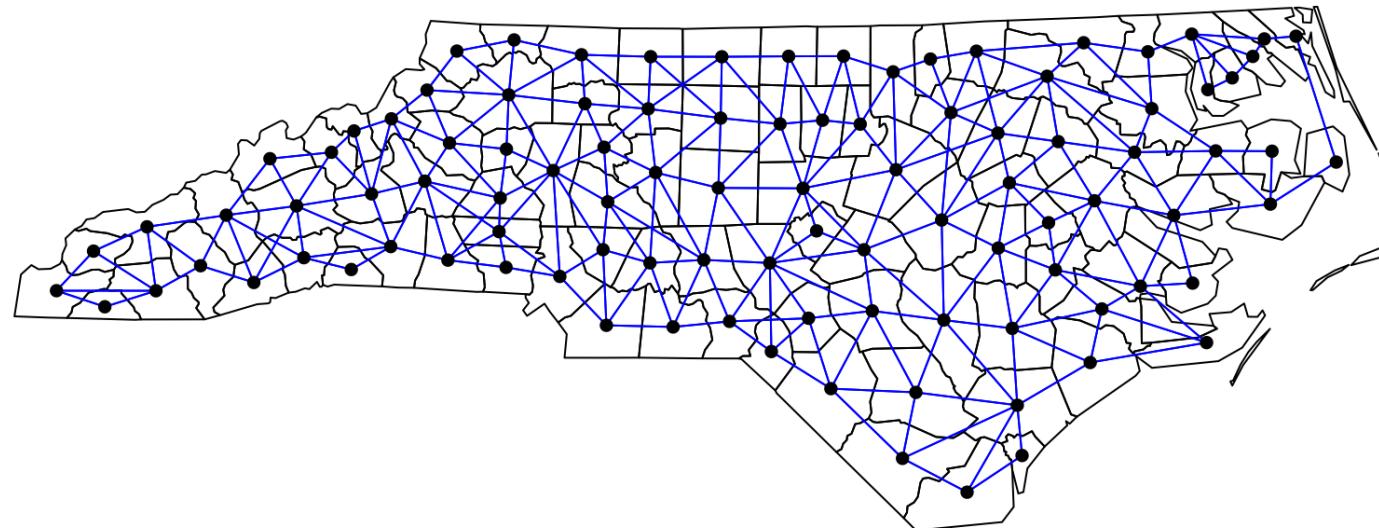
Weights style: M

Weights constants summary:

n	nn	S0	S1	S2	
M	100	10000	490	980	10696

Plotting listw

```
1 nc_coords = nc %>% st_centroid() %>% st_coordinates()  
2  
3 plot(st_geometry(nc))  
4 plot(listW, nc_coords, add=TRUE, col="blue", pch=16)
```



Moran's I

```
1 spdep::moran.test(nc$SID74, listW)
```

Moran I test under randomisation

```
data: nc$SID74  
weights: listW  
  
Moran I statistic standard deviate = 2.1707,  
p-value = 0.01498  
alternative hypothesis: greater  
sample estimates:  
Moran I statistic      Expectation  
0.119089049      -0.010101010  
Variance  
0.003542176
```

```
1 spdep::moran.test(1000*nc$SID74/nc$BIR74, lis
```

Moran I test under randomisation

```
data: 1000 * nc$SID74/nc$BIR74  
weights: listW  
  
Moran I statistic standard deviate = 3.6355,  
p-value = 0.0001387  
alternative hypothesis: greater  
sample estimates:  
Moran I statistic      Expectation  
0.210046454      -0.010101010  
Variance  
0.003666802
```

Geary's C

```
1 spdep::geary.test(nc$SID74, listW)
```

Geary C test under randomisation

```
data: nc$SID74  
weights: listW  
  
Geary C statistic standard deviate =  
0.91949, p-value = 0.1789
```

alternative hypothesis: Expectation greater than
statistic

sample estimates:

Geary C statistic	Expectation
0.88988684	1.00000000

Variance
0.01434105

```
1 spdep::geary.test(1000*nc$SID74/nc$BIR74, lis
```

Geary C test under randomisation

```
data: 1000 * nc$SID74/nc$BIR74  
weights: listW
```

```
Geary C statistic standard deviate = 3.0989,  
p-value = 0.0009711  
alternative hypothesis: Expectation greater than  
statistic
```

sample estimates:

Geary C statistic	Expectation
0.67796679	1.00000000

Variance
0.01079878

CAR Model

```
1 nc_car = spatialreg::spautolm(  
2   formula = 1000*SID74/BIR74 ~ 1, data = nc,  
3   listw = listW, family = "CAR"  
4 )  
5  
6 summary(nc_car)
```

Call:

```
spatialreg::spautolm(formula = 1000 * SID74/BIR74 ~ 1, data = nc,  
listw = listW, family = "CAR")
```

Residuals:

Min	1Q	Median	3Q	Max
-2.13872	-0.83535	-0.22355	0.55014	7.68640

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.00203	0.24272	8.2484	2.22e-16

Lambda: 0.13062 LR test value: 8.6251 p-value: 0.0033157

Numerical Hessian standard error of lambda: 0.030475

SAR Model

```
1 nc_sar = spatialreg::spautolm(  
2   formula = 1000*SID74/BIR74 ~ 1, data = nc,  
3   listw = listW, family = "SAR"  
4 )  
5  
6 summary(nc_sar)
```

Call:

```
spatialreg::spautolm(formula = 1000 * SID74/BIR74 ~ 1, data = nc,  
listw = listW, family = "SAR")
```

Residuals:

Min	1Q	Median	3Q	Max
-2.09307	-0.87039	-0.20274	0.51156	7.62830

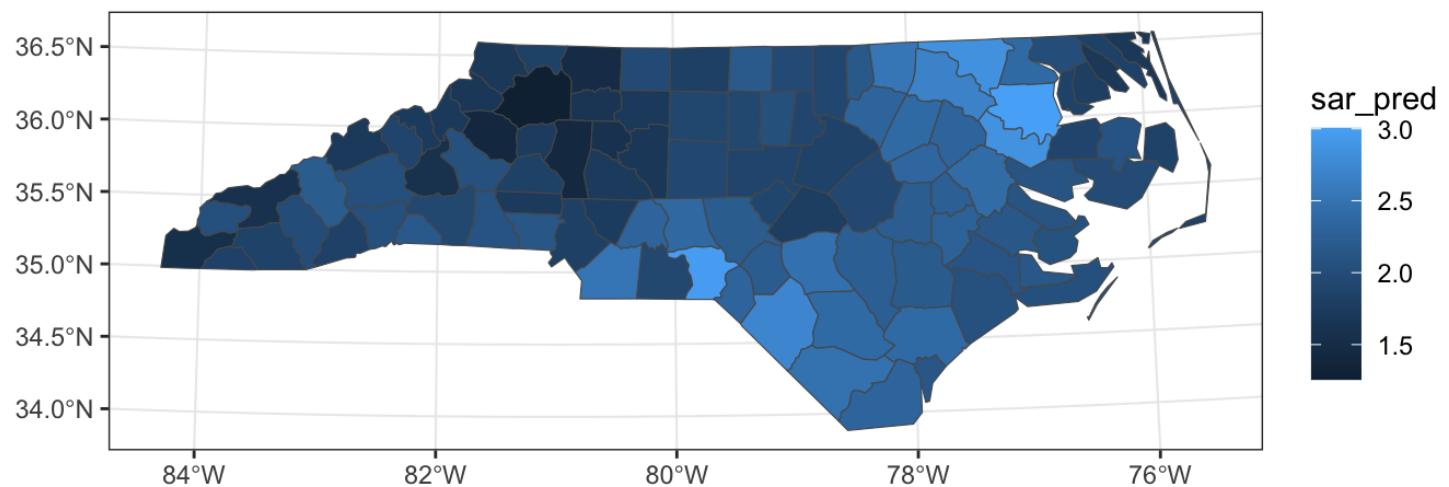
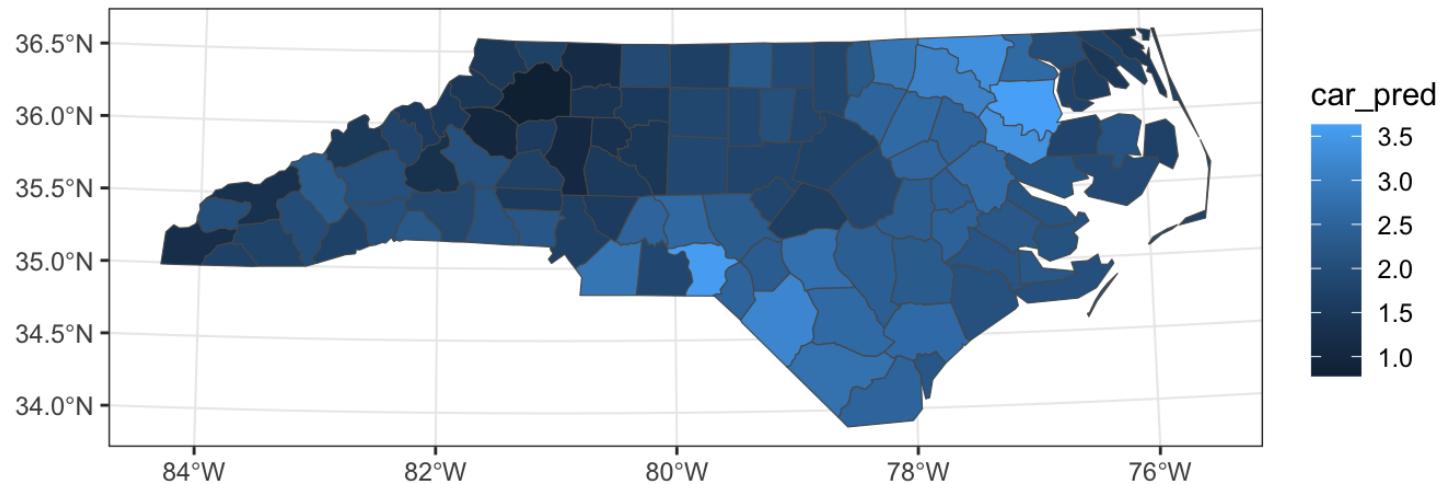
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.01084	0.23622	8.5127	< 2.2e-16

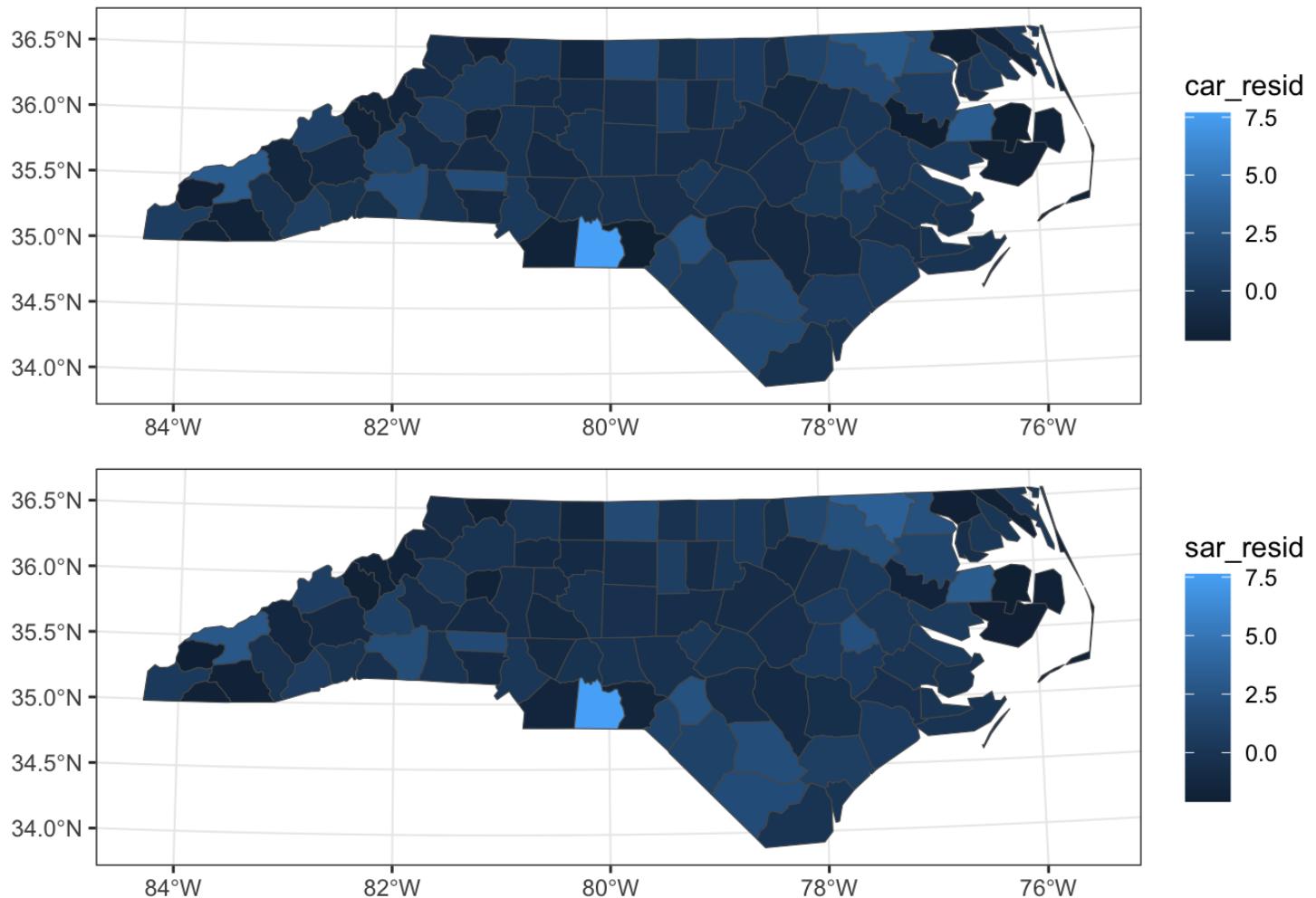
Lambda: 0.079934 LR test value: 8.8911 p-value: 0.0028657

Numerical Hessian standard error of lambda: 0.024599

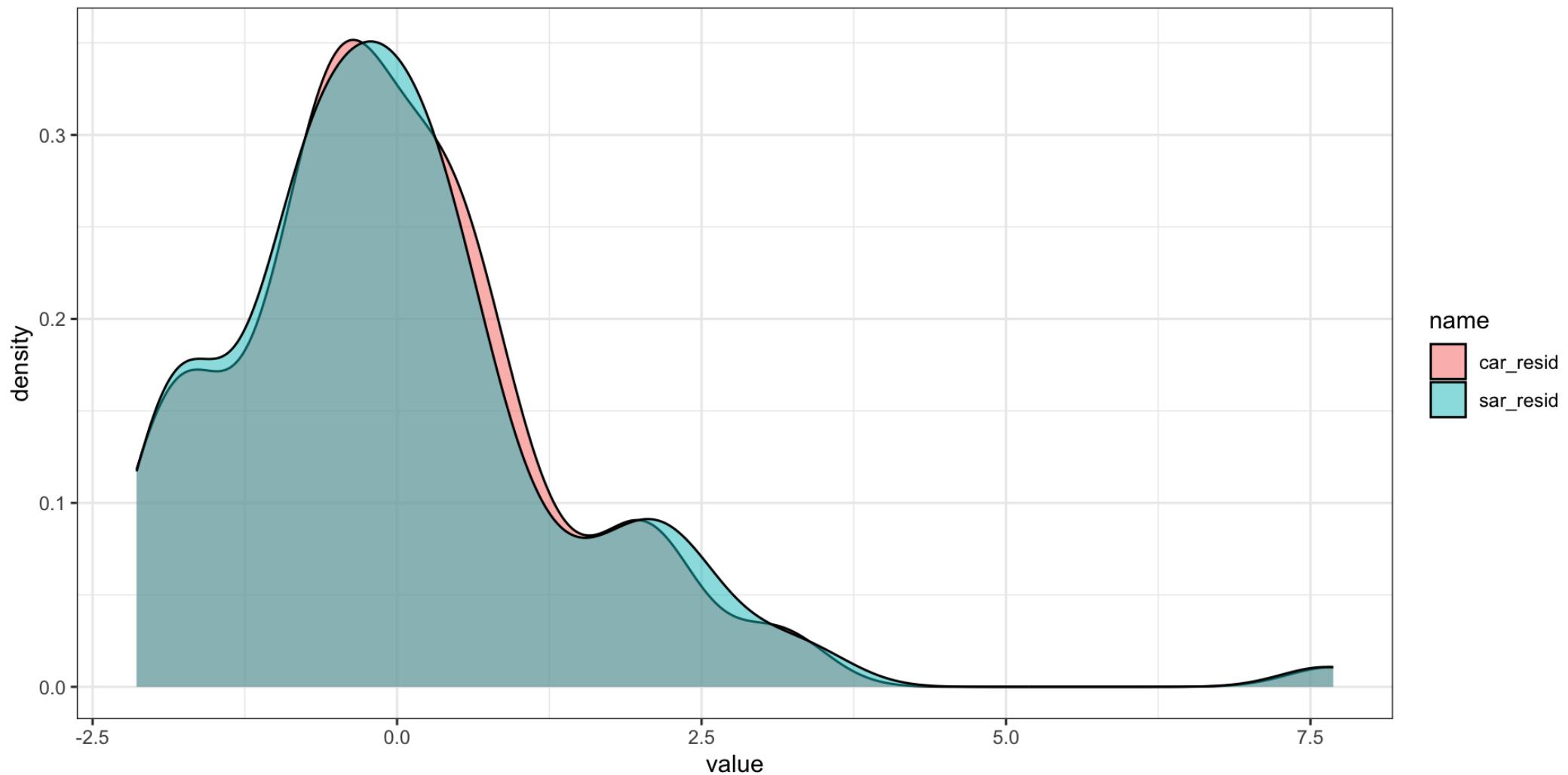
Predictions



Residuals



Residual distributions



Residual autocorrelation

```
1 spdep::moran.test(nc$car_resid, listW)
```

Moran I test under randomisation

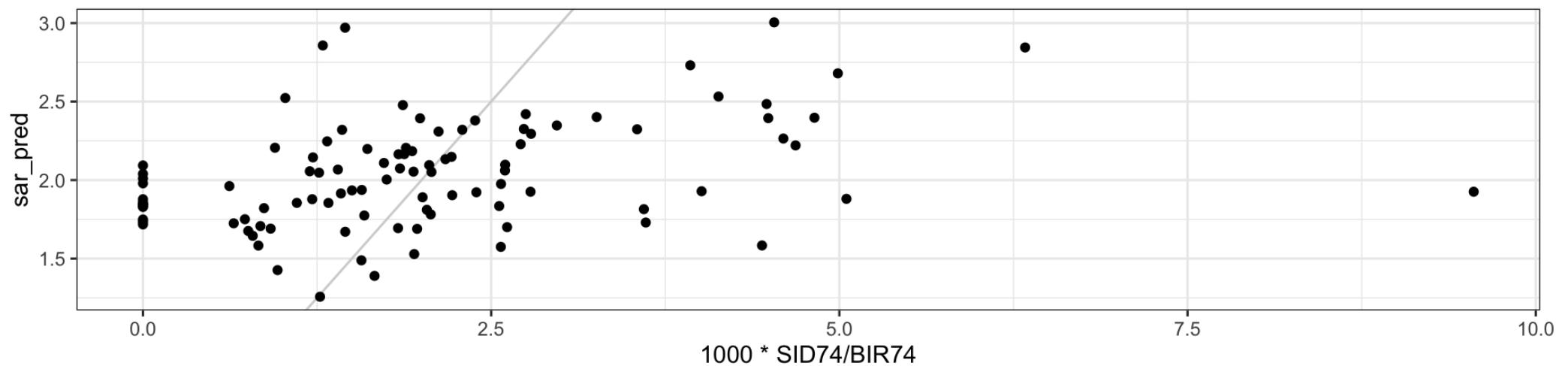
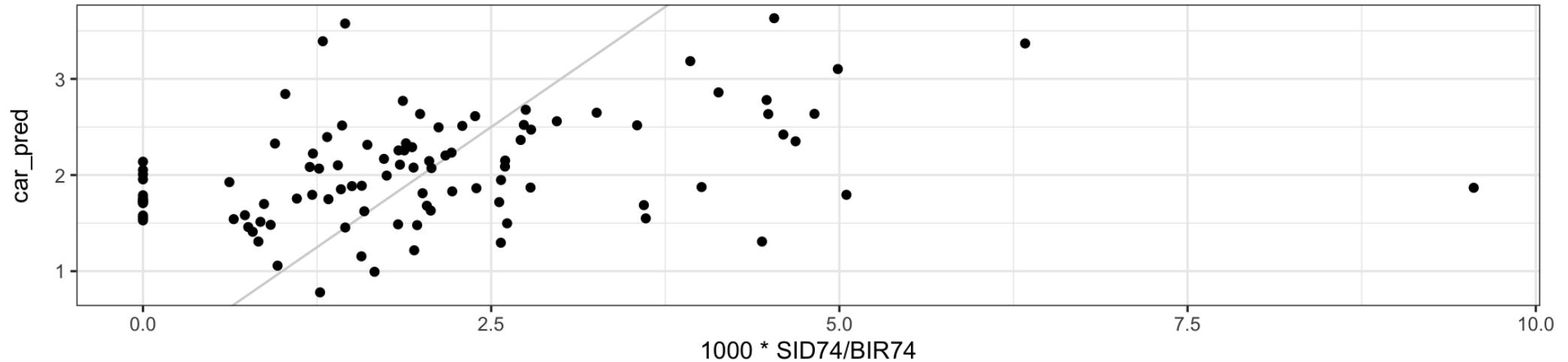
```
data: nc$car_resid  
weights: listW  
  
Moran I statistic standard deviate =  
-1.7952, p-value = 0.9637  
alternative hypothesis: greater  
sample estimates:  
Moran I statistic      Expectation  
-0.117449316      -0.010101010  
Variance  
0.003575538
```

```
1 spdep::moran.test(nc$sar_resid, listW)
```

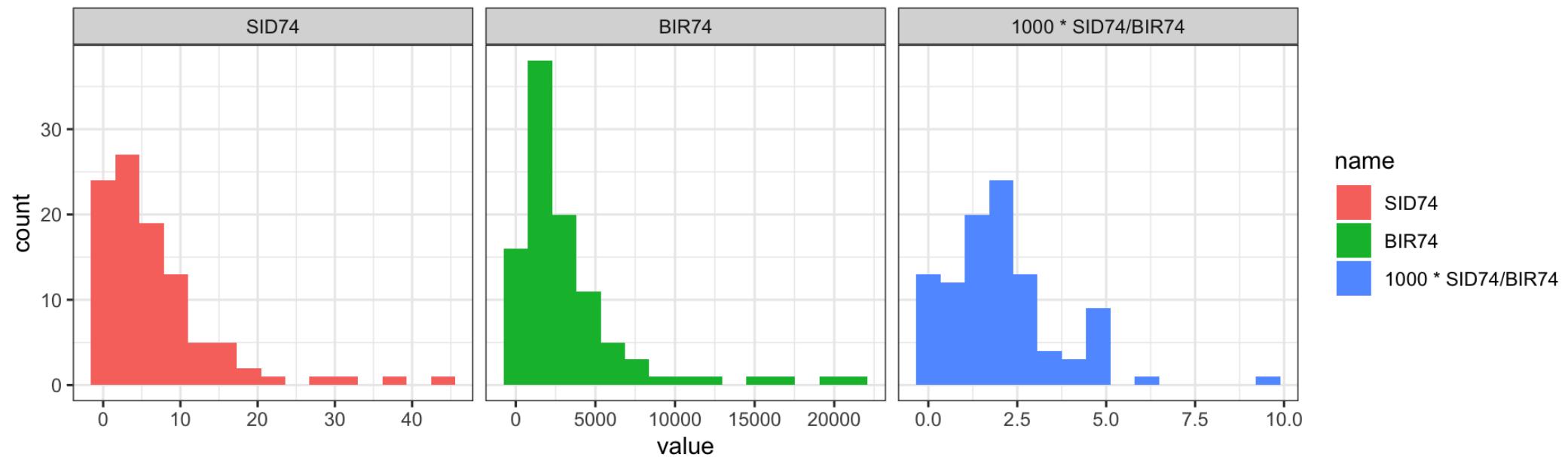
Moran I test under randomisation

```
data: nc$sar_resid  
weights: listW  
  
Moran I statistic standard deviate =  
0.17958, p-value = 0.4287  
alternative hypothesis: greater  
sample estimates:  
Moran I statistic      Expectation  
0.0006769074      -0.010101010  
Variance  
0.0036020941
```

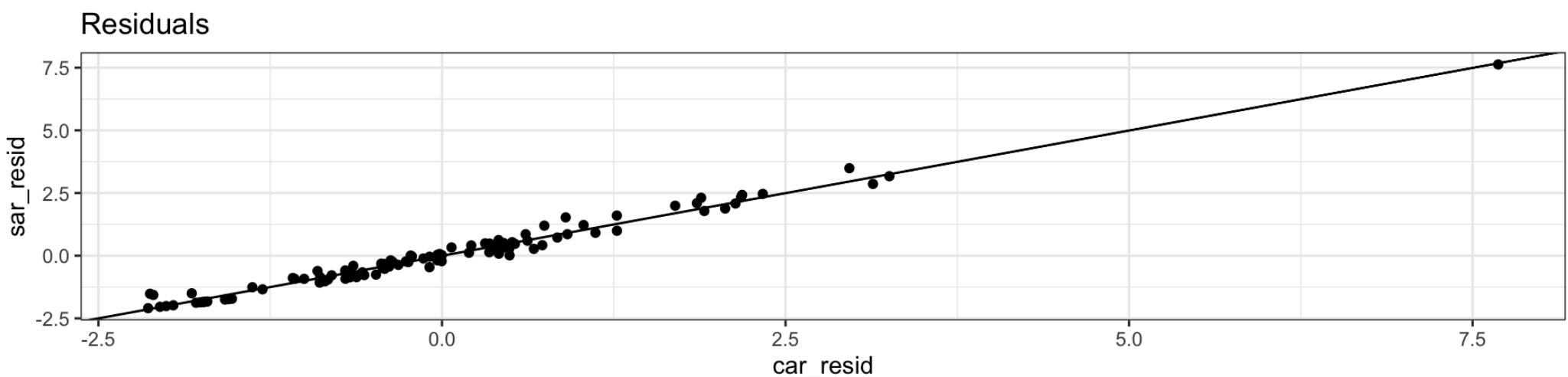
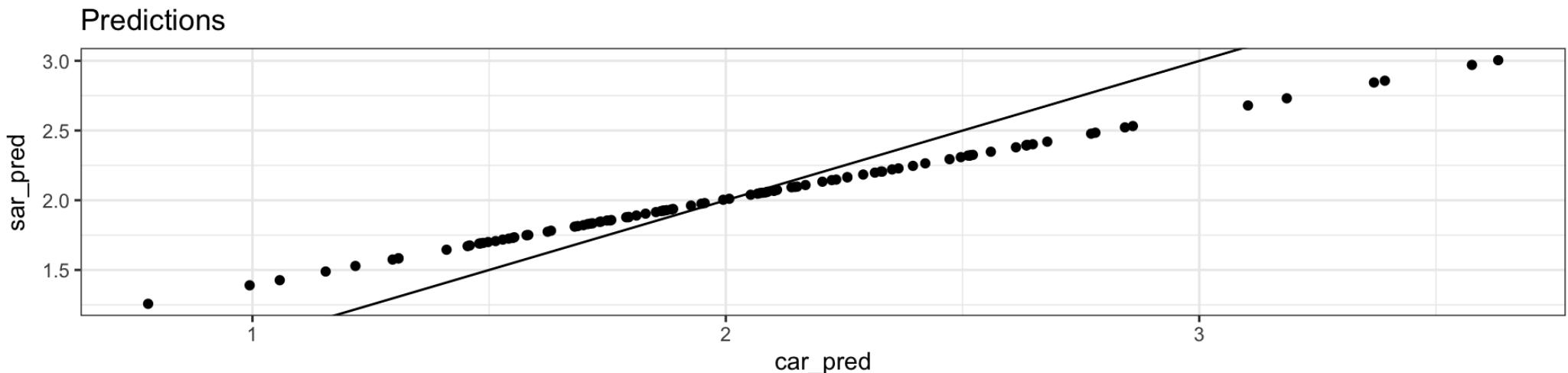
Predicted vs Observed



What's wrong?



Comparing CAR vs SAR.



Transforming the data

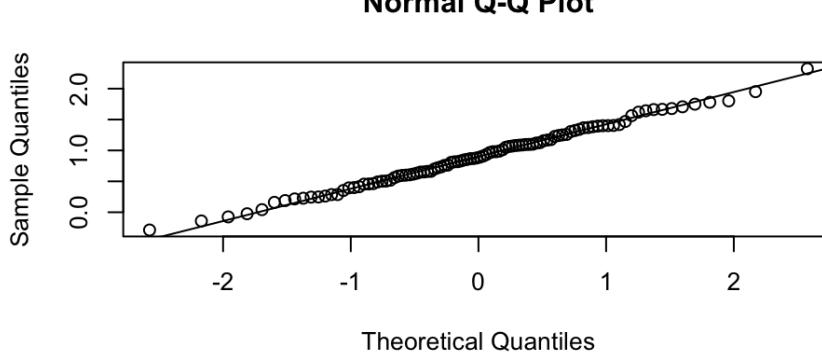
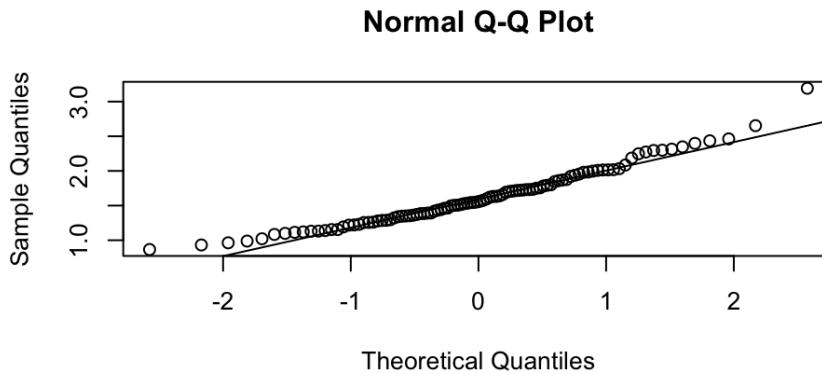
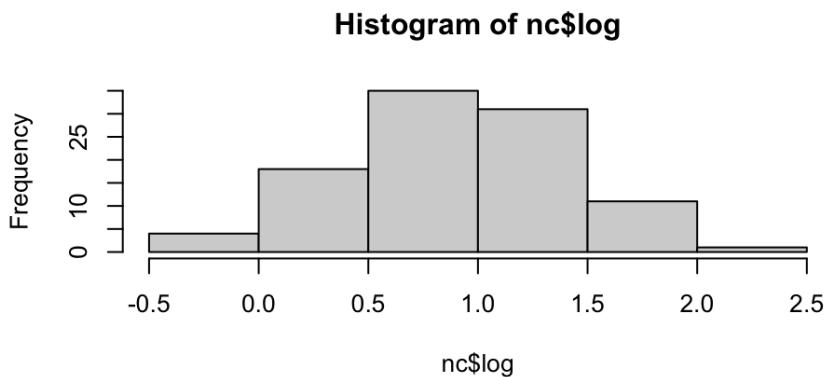
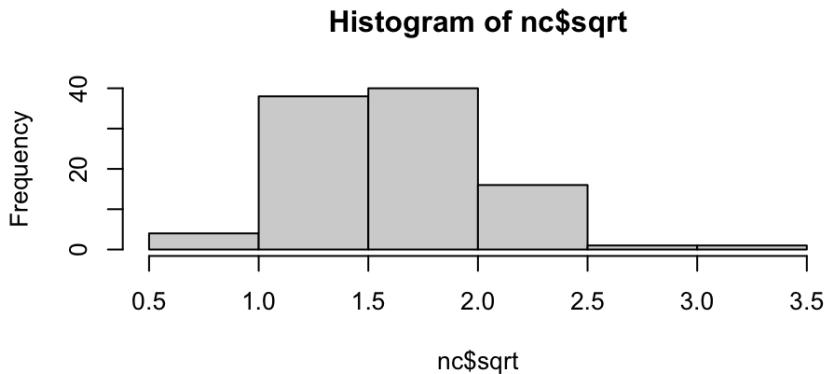
Freeman-Tukey's transformation

This is the transformation used by Cressie and Road in Spatial Data Analysis of Regional Counts (1989).

$$FT = \sqrt{1000} \left(\sqrt{\frac{SID74}{BIR74}} + \sqrt{\frac{SID74 + 1}{BIR74}} \right)$$

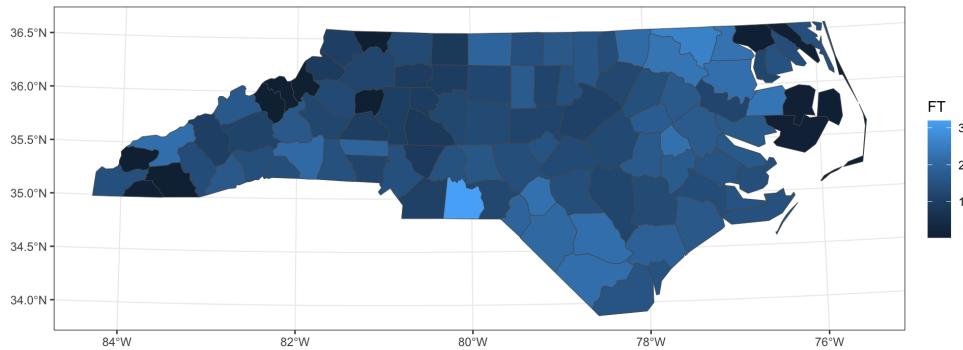
Other possibilities

```
1 nc = mutate(nc,  
2   sqrt = sqrt(1000*(SID74+1)/BIR74),  
3   log  = log(1000*(SID74+1)/BIR74),  
4 )
```



FT transformation

```
1 ggplot(nc) + geom_sf(aes(fill=FT))
```



```
1 spdep::moran.test(nc$FT, listW)
```

Moran I test under randomisation

data: nc\$FT

weights: listW

Moran I statistic standard deviate = 3.664,
p-value = 0.0001242

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation
-------------------	-------------

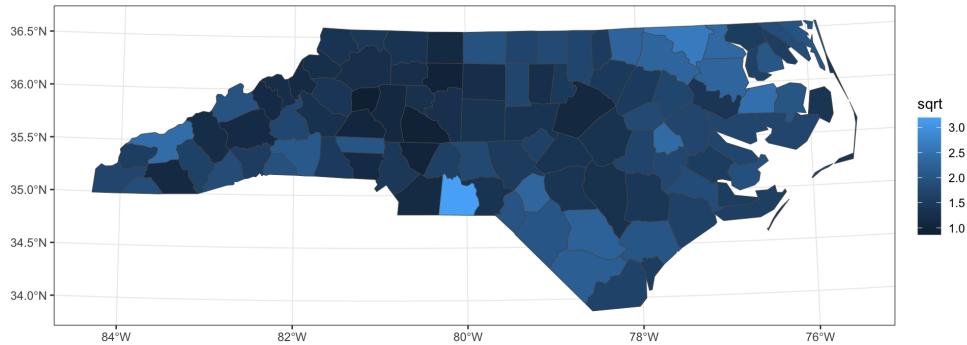
0.216246481	-0.010101010
-------------	--------------

Variance

0.003816298

sqrt transformation

```
1 ggplot(nc) + geom_sf(aes(fill=sqrt))
```



```
1 spdep::moran.test(nc$sqrt, listW)
```

Moran I test under randomisation

data: nc\$sqrt

weights: listW

Moran I statistic standard deviate = 4.5217,

p-value = 3.067e-06

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation
-------------------	-------------

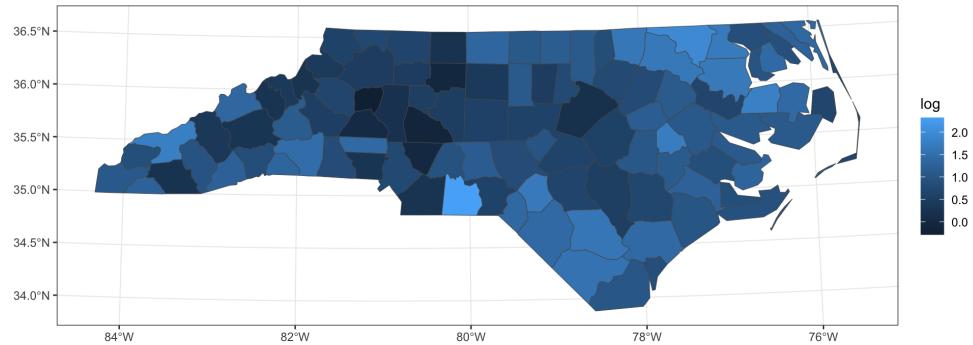
0.268600322	-0.010101010
-------------	--------------

Variance

0.003798988

log transformation

```
1 ggplot(nc) + geom_sf(aes(fill=log))
```



```
1 spdep::moran.test(nc$log, listW)
```

Moran I test under randomisation

data: nc\$log

weights: listW

Moran I statistic standard deviate = 4.9895,

p-value = 3.027e-07

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation
-------------------	-------------

0.299245438	-0.010101010
-------------	--------------

Variance

0.003843927

CAR Models

```
1 nc_car_ft    = spatialreg::spautolm(formula = FT ~ 1,    data = nc, listw = listW, family = "CAR")
2 nc_car_sqrt = spatialreg::spautolm(formula = sqrt ~ 1, data = nc, listw = listW, family = "CAR")
3 nc_car_log   = spatialreg::spautolm(formula = log ~ 1,   data = nc, listw = listW, family = "CAR")
4
5 AIC(nc_car_ft)
```

```
[1] 192.1781
```

```
1 AIC(nc_car_sqrt)
```

```
[1] 100.8898
```

```
1 AIC(nc_car_log)
```

```
[1] 134.644
```

SAR Model

```
1 nc_sar_ft    = spatialreg::spautolm(formula = FT ~ 1,    data = nc, listw = listW, family = "SAR")
2 nc_sar_sqrt = spatialreg::spautolm(formula = sqrt ~ 1, data = nc, listw = listW, family = "SAR")
3 nc_sar_log   = spatialreg::spautolm(formula = log ~ 1,   data = nc, listw = listW, family = "SAR")
4
5 AIC(nc_sar_ft)
```

```
[1] 191.9918
```

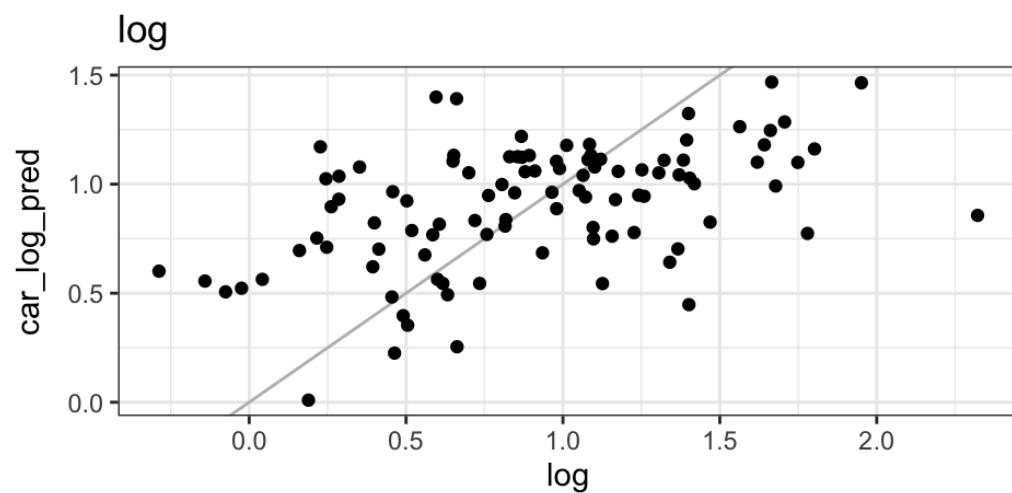
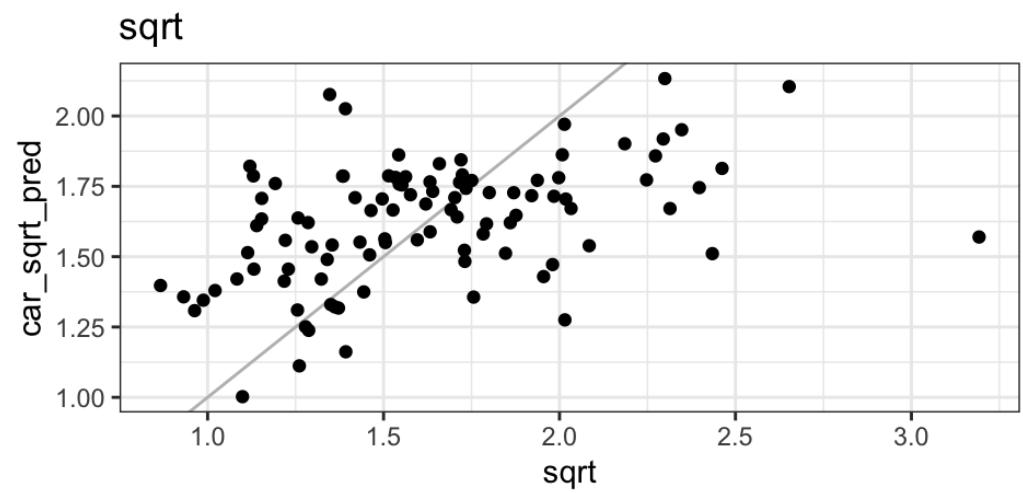
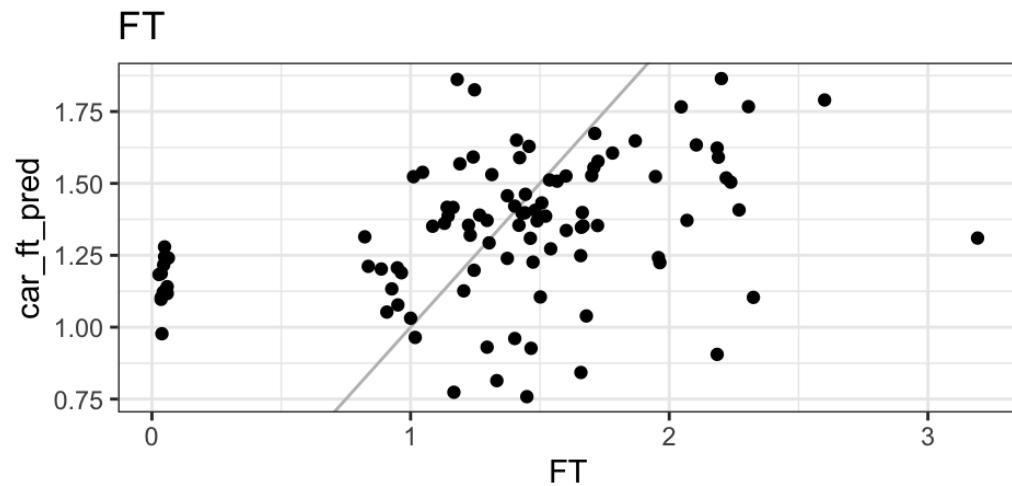
```
1 AIC(nc_sar_sqrt)
```

```
[1] 102.717
```

```
1 AIC(nc_sar_log)
```

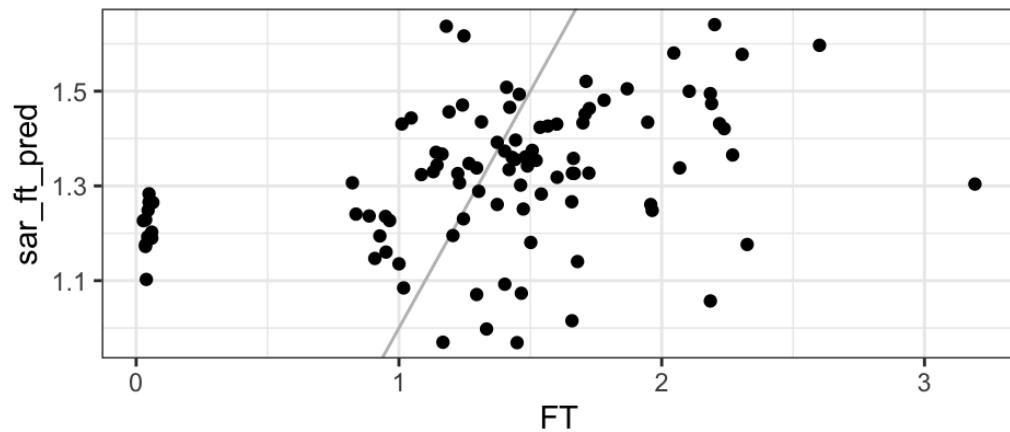
```
[1] 137.4095
```

CAR predictions

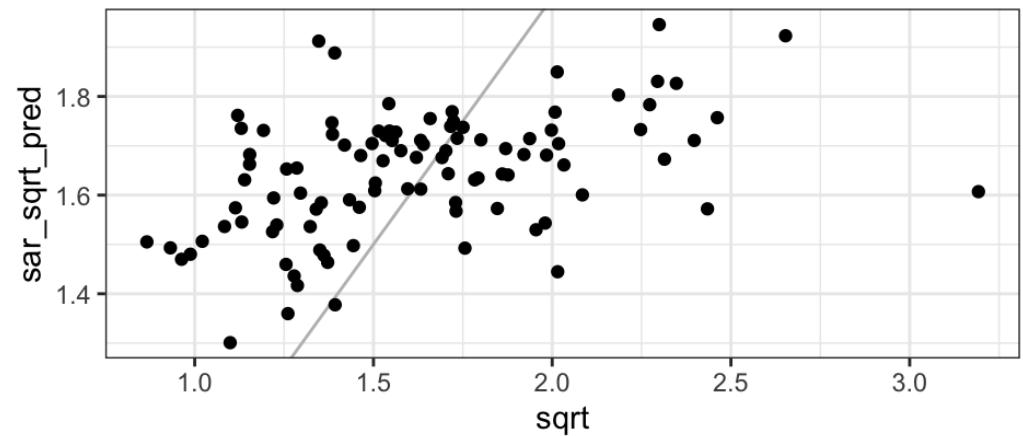


SAR predictions

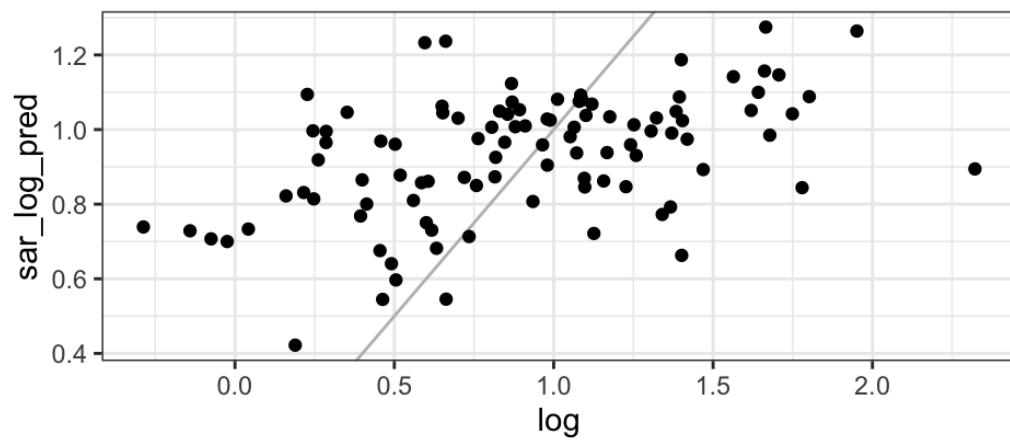
FT



sqrt

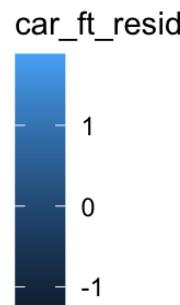
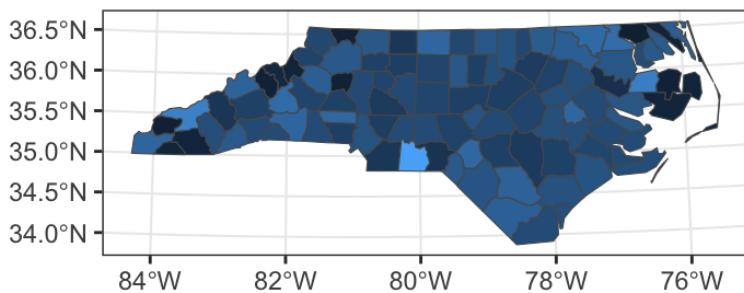


log

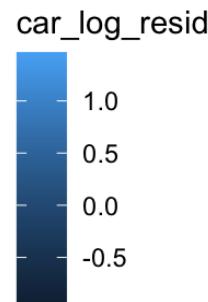
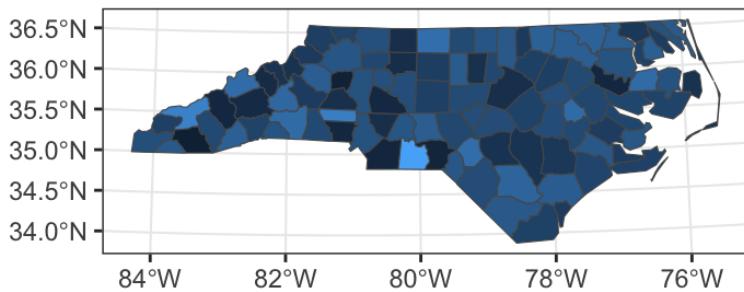


CAR residuals

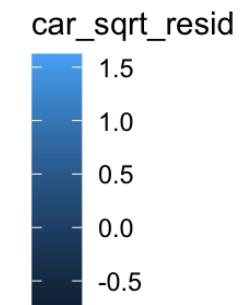
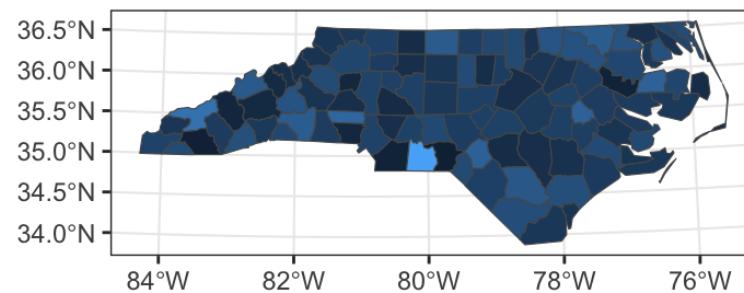
FT



log

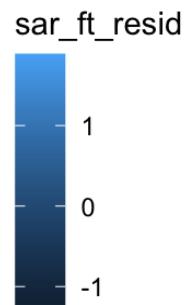
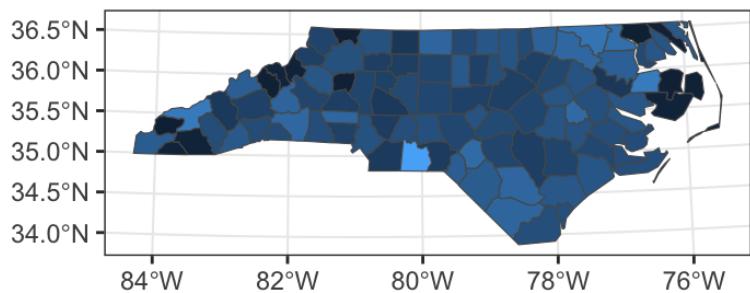


sqrt

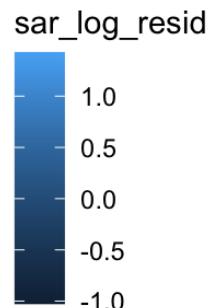
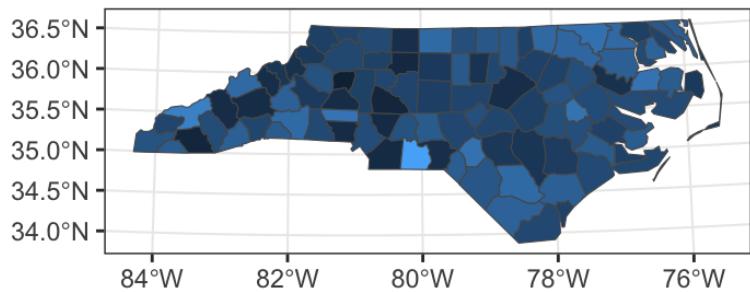


SAR predictions

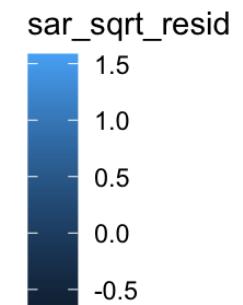
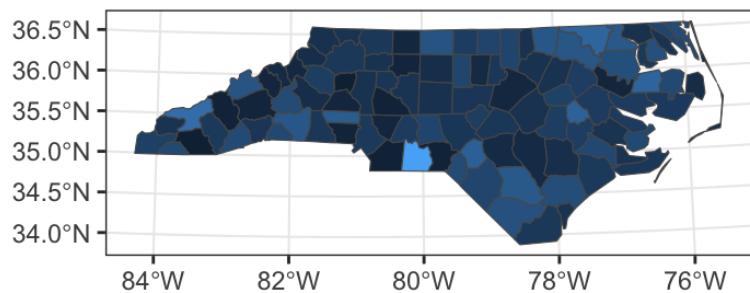
FT



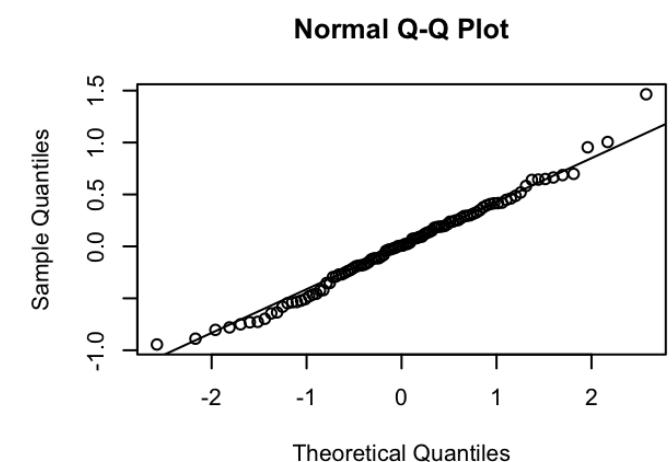
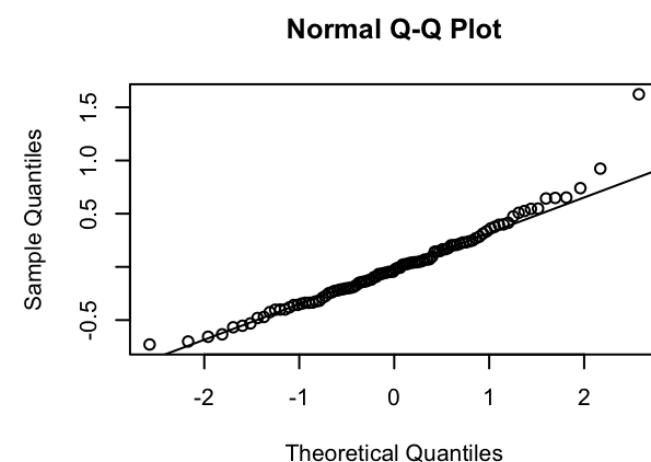
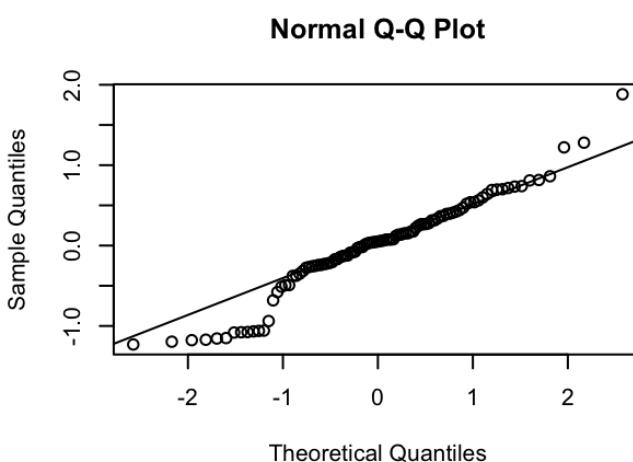
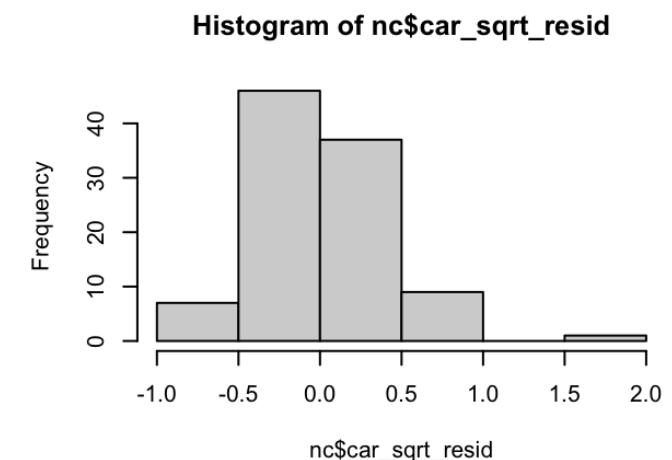
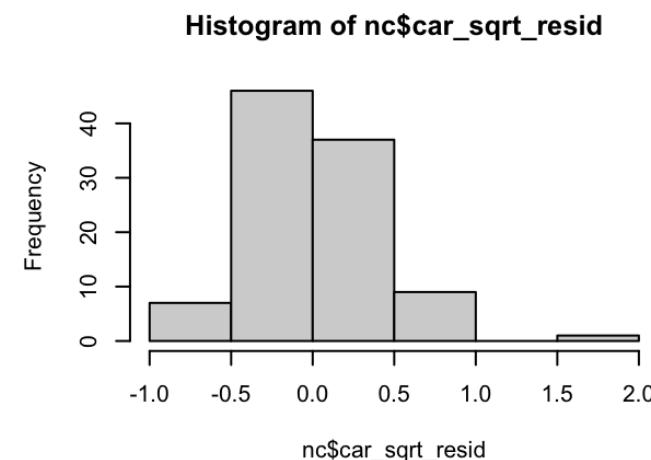
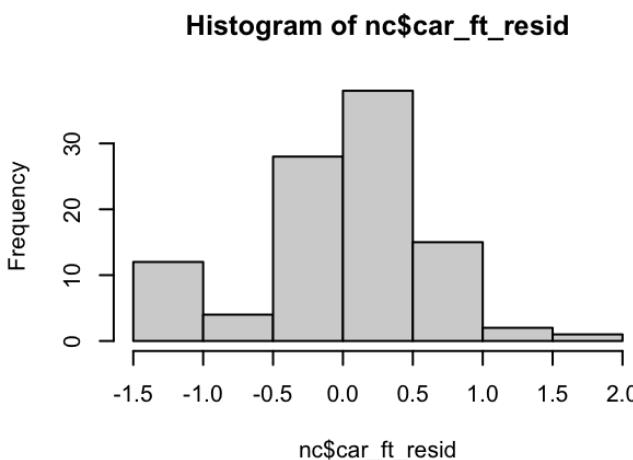
log



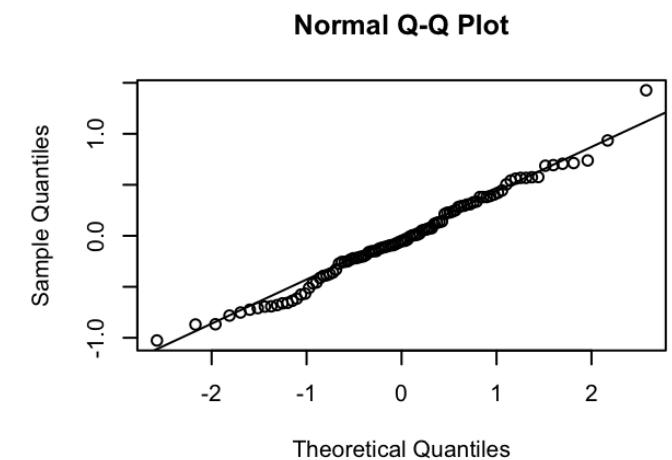
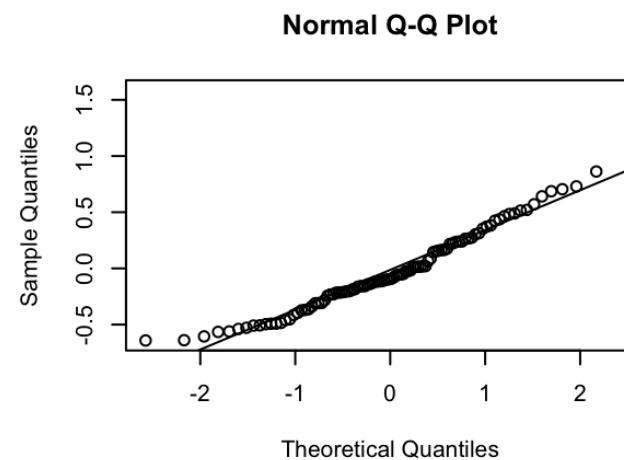
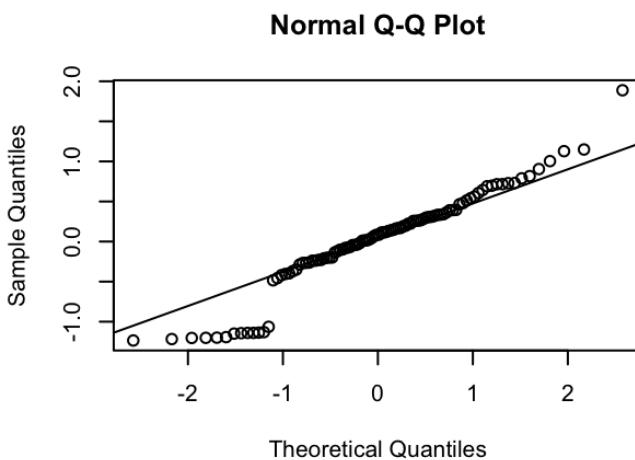
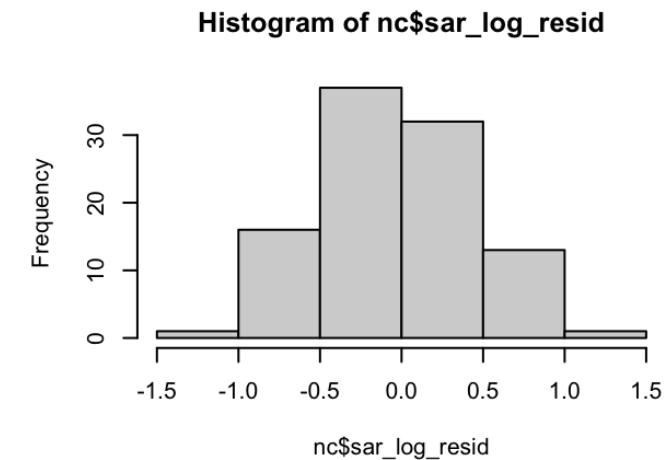
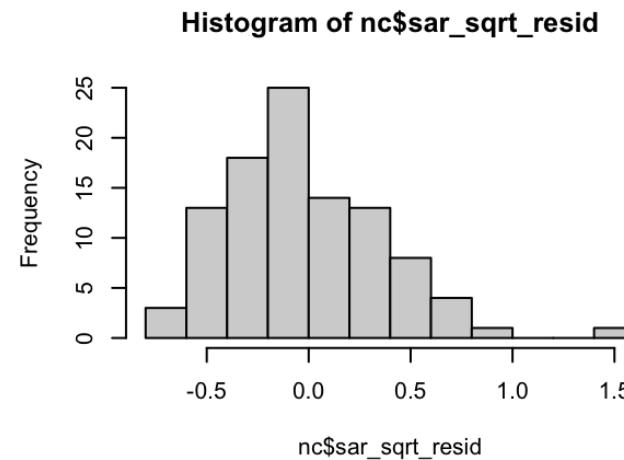
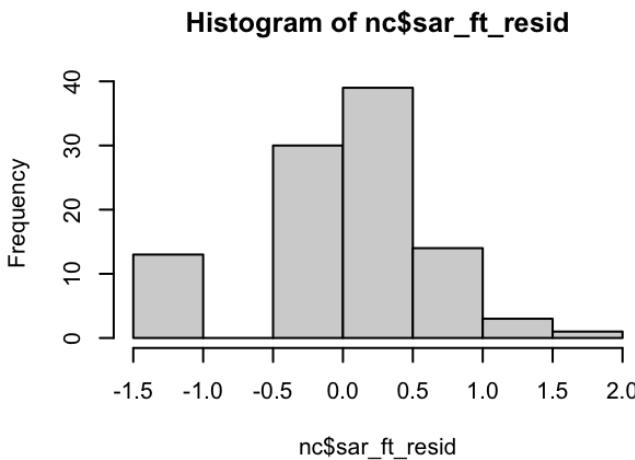
sqrt



CAR residual distributions



SAR residual distributions



Residual spatial autocorrelation

```
1 spdep::moran.test(nc$car_sqrt_resid, listW)
```

Moran I test under randomisation

```
data: nc$car_sqrt_resid  
weights: listW  
  
Moran I statistic standard deviate =  
-3.1196, p-value = 0.9991  
alternative hypothesis: greater  
sample estimates:  
Moran I statistic      Expectation  
-0.200890550      -0.010101010  
Variance  
0.003740354
```

```
1 spdep::moran.test(nc$sar_sqrt_resid, listW)
```

Moran I test under randomisation

```
data: nc$sar_sqrt_resid  
weights: listW  
  
Moran I statistic standard deviate = -0.422,  
p-value = 0.6635  
alternative hypothesis: greater  
sample estimates:  
Moran I statistic      Expectation  
-0.035976080      -0.010101010  
Variance  
0.003759585
```

CAR & SAR with brms

brms CAR

```
1 b_car = brms::brm(  
2     1000*SID74/BIR74 ~ 1 + car(A), data=nc, data2=list(A=A),  
3     adapt_delta = 0.95,  
4     silent=2, refresh=0, iter=20000,  
5     cores = 4, backend = "cmdstanr"  
6 )
```

Running MCMC with 4 parallel chains...

Chain 3 finished in 11.0 seconds.

Chain 2 finished in 11.7 seconds.

Chain 4 finished in 12.4 seconds.

Chain 1 finished in 12.7 seconds.

All 4 chains finished successfully.

Mean chain execution time: 11.9 seconds.

Total execution time: 12.7 seconds.

```
1 b_car
```

Family: gaussian

Links: mu = identity; sigma = identity

Formula: 1000 * SID74/BIR74 ~ 1 + car(A)

Data: nc (Number of observations: 100)

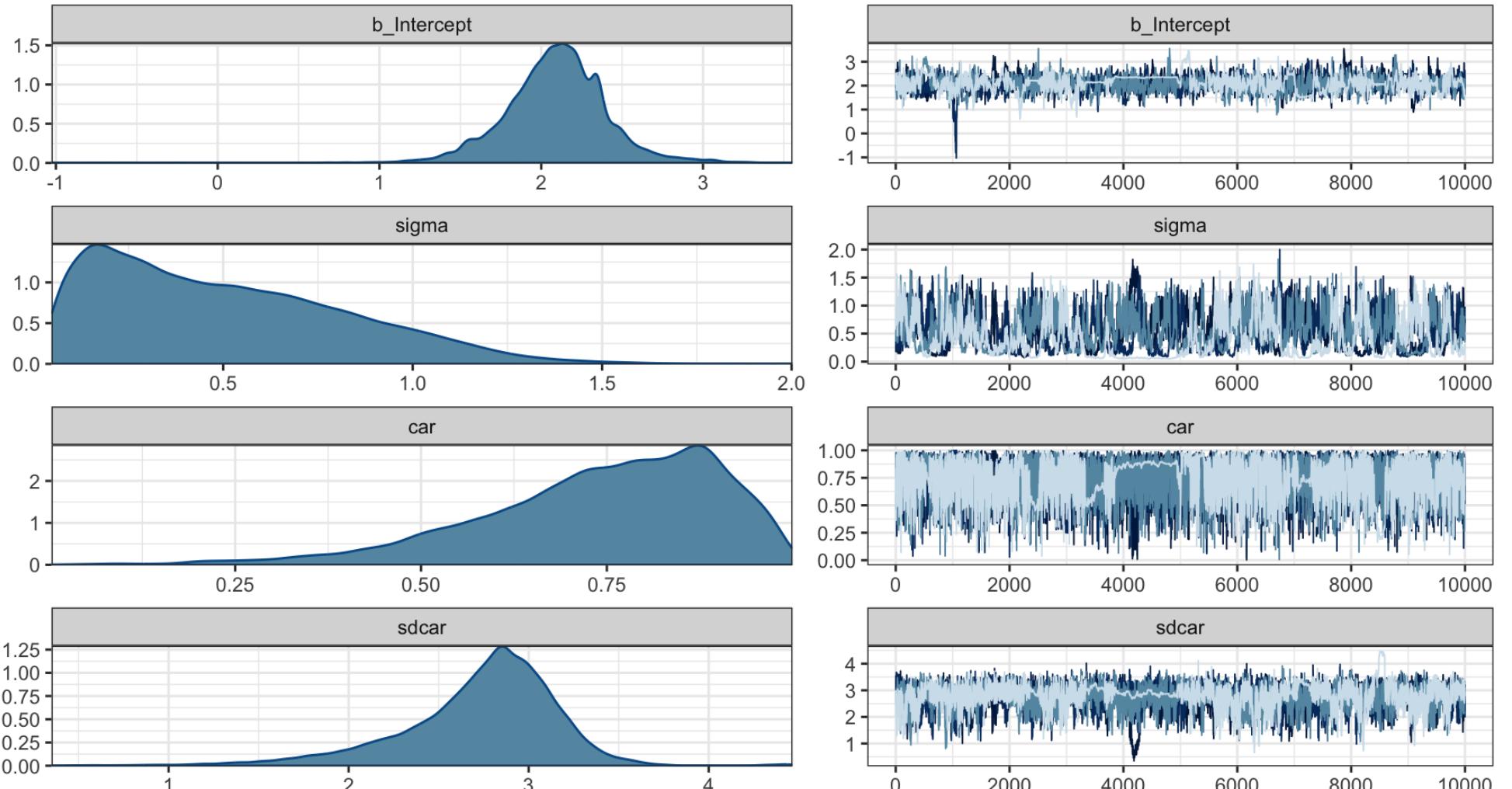
Draws: 4 chains, each with iter = 20000; warmup = 10000; thin = 1;
total post-warmup draws = 40000

Correlation Structures:

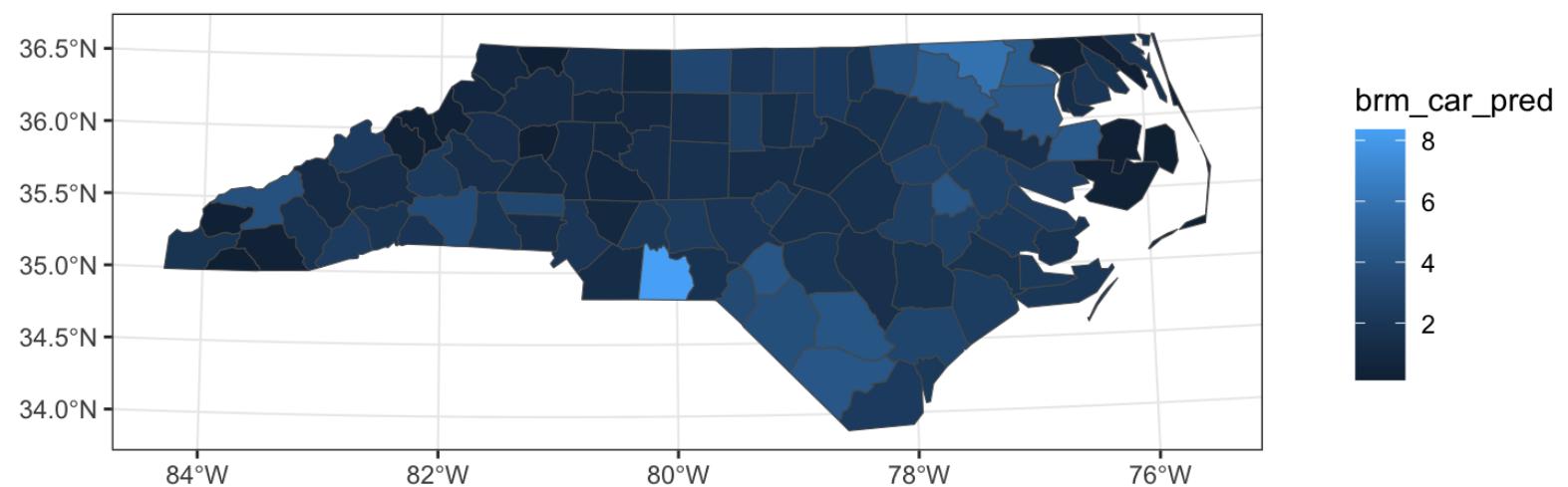
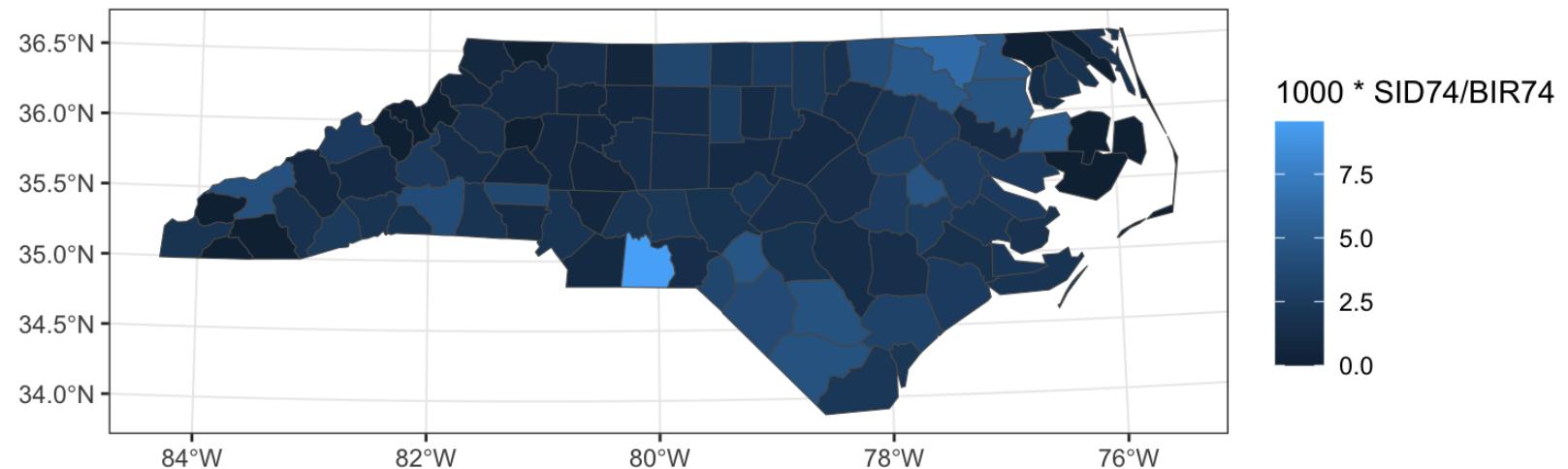
	Estimate	Est.Error	1-95% CI	u-95% CI	Rhat
car	0.74	0.16	0.35	0.97	1.00
sdcar	2.73	0.44	1.67	3.40	1.01
Bulk ESS Tail ESS					

Diagnostics

```
1 plot(b_car)
```

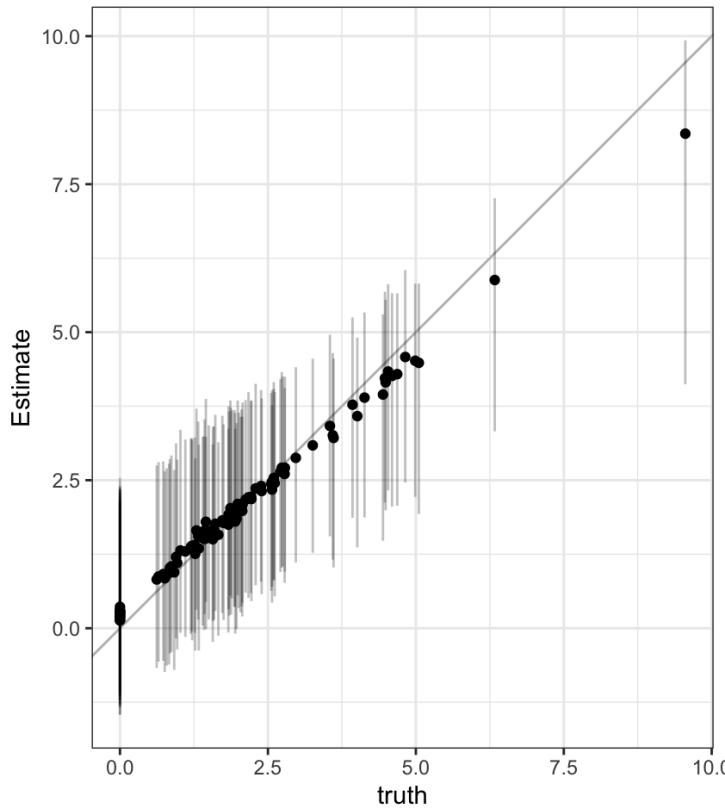


Predictions



Observed vs predicted

```
1 ggplot(p, aes(x=truth, y=Estimate)) +  
2   geom_abline(intercept=0, slope=1, color="grey") +  
3   geom_point() +  
4   geom_errorbar(aes(ymin=Q2.5, ymax=Q97.5), alpha=0.25) +  
5   coord_fixed()
```



brms SAR

```
1 b_sar = brms::brm(  
2     1000*SID74/BIR74 ~ 1 + sar(listW), data=nc, data2=list(listW=listW),  
3     silent=2, refresh=0, iter=4000,  
4     cores = 4, backend = "cmdstanr"  
5 )
```

Running MCMC with 4 parallel chains...

Chain 4 finished in 2.8 seconds.

Chain 3 finished in 2.9 seconds.

Chain 1 finished in 2.9 seconds.

Chain 2 finished in 2.9 seconds.

All 4 chains finished successfully.

Mean chain execution time: 2.9 seconds.

Total execution time: 3.3 seconds.

```
1 b_sar
```

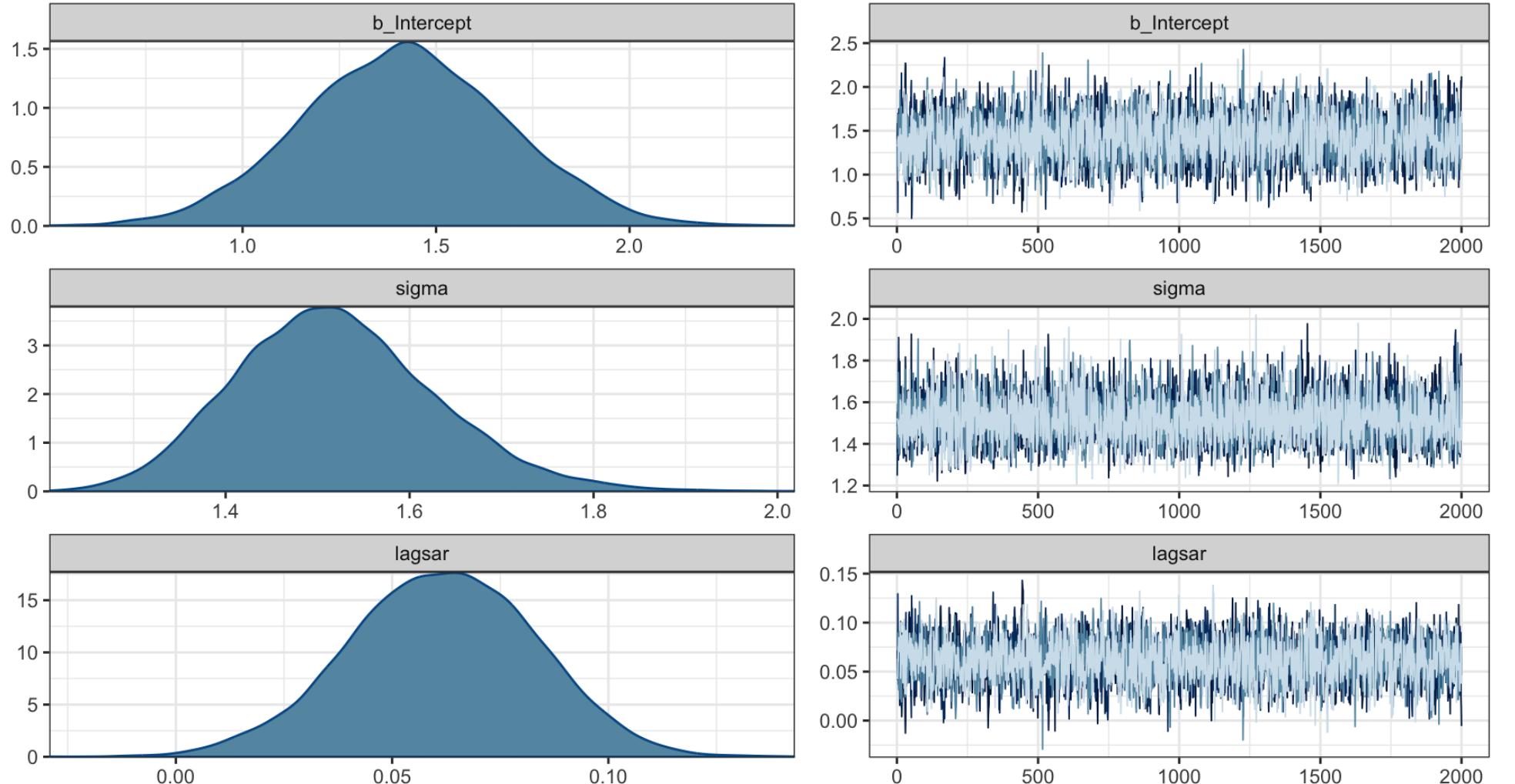
Family: gaussian
Links: mu = identity; sigma = identity
Formula: 1000 * SID74/BIR74 ~ 1 + sar(listW)
Data: nc (Number of observations: 100)
Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
total post-warmup draws = 8000

Correlation Structures:

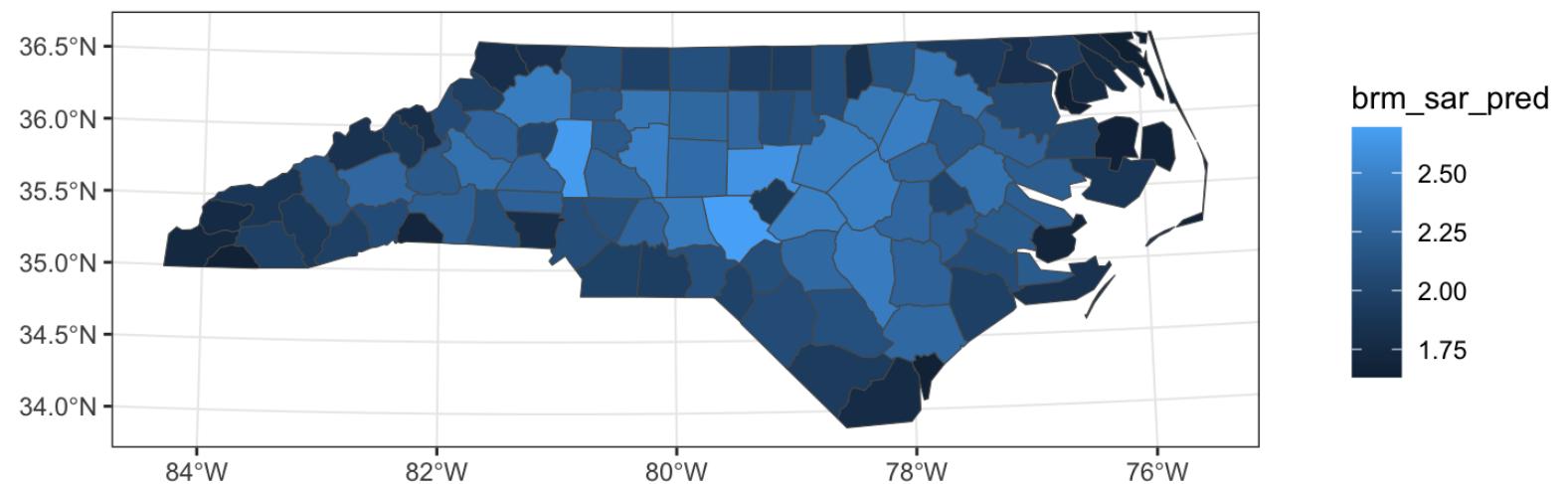
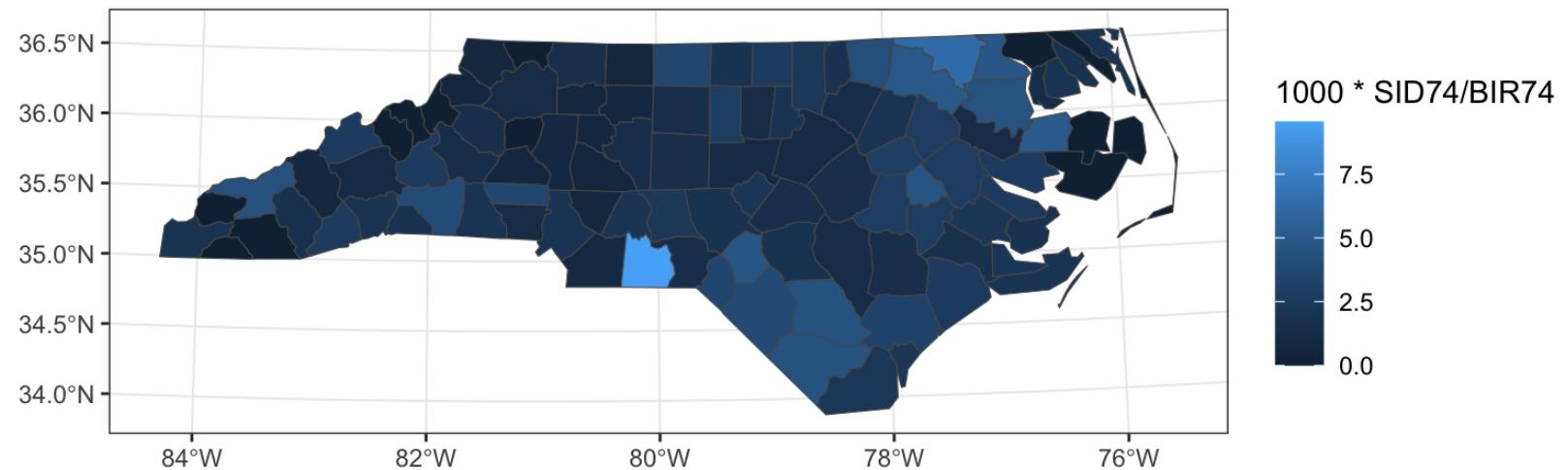
	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat
lagsar	0.06	0.02	0.02		0.10	1.00	
	Bulk_ESS	Tail_ESS					
laqsar	3128	3877					

Diagnostics

```
1 plot(b_sar)
```

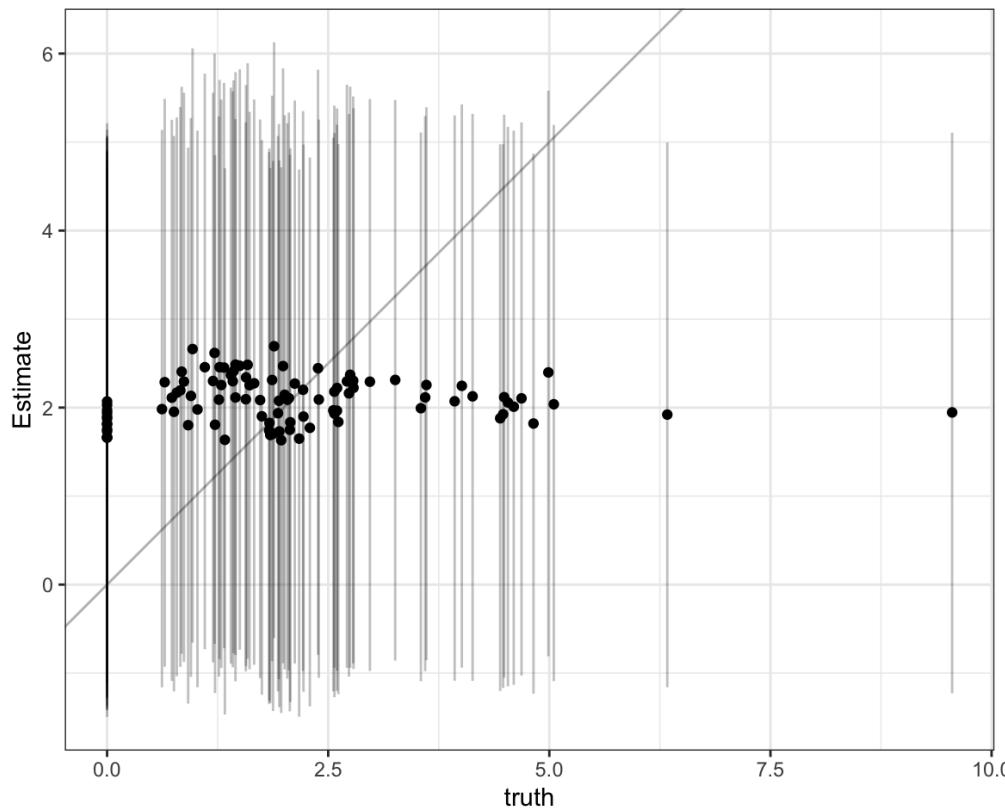


Predictions



Observed vs predicted

```
1 ggplot(p, aes(x=truth, y=Estimate)) +  
2   geom_abline(intercept=0, slope=1, color="grey") +  
3   geom_point() +  
4   geom_errorbar(aes(ymin=Q2.5, ymax=Q97.5), alpha=0.25) +  
5   coord_fixed()
```



Brief Aside - CAR & SAR precision matrices

$$\Sigma_{\text{SAR}} = (\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A})^{-1} \sigma^2 \mathbf{D}^{-1} ((\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A})^{-1})^t$$

$$\begin{aligned}\Sigma_{\text{SAR}}^{-1} &= \left((\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A})^{-1} \sigma^2 \mathbf{D}^{-1} ((\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A})^{-1})^t \right)^{-1} \\ &= \left(((\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A})^{-1})^t \right)^{-1} \frac{1}{\sigma^2} \mathbf{D} (\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A}) \\ &= \frac{1}{\sigma^2} (\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A})^t \mathbf{D} (\mathbf{I} - \phi \mathbf{D}^{-1} \mathbf{A})\end{aligned}$$

$$\Sigma_{\text{CAR}}^{-1} = \frac{1}{\sigma^2} (\mathbf{D} - \phi \mathbf{A})$$