Fitting CAR and SAR Models

Lecture 20

Dr. Colin Rundel

Fitting areal models

Revised SAR Model

- Formula Model

$$y(s_i) = X_{i\cdot}\,\beta + \phi \sum_{j=1}^{n} D_{jj}^{-1}\; A_{ij}\,\Big(y(s_j) - X_{j\cdot}\,\beta\Big) + \epsilon_i$$

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

- Joint Model

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

Revised CAR Model

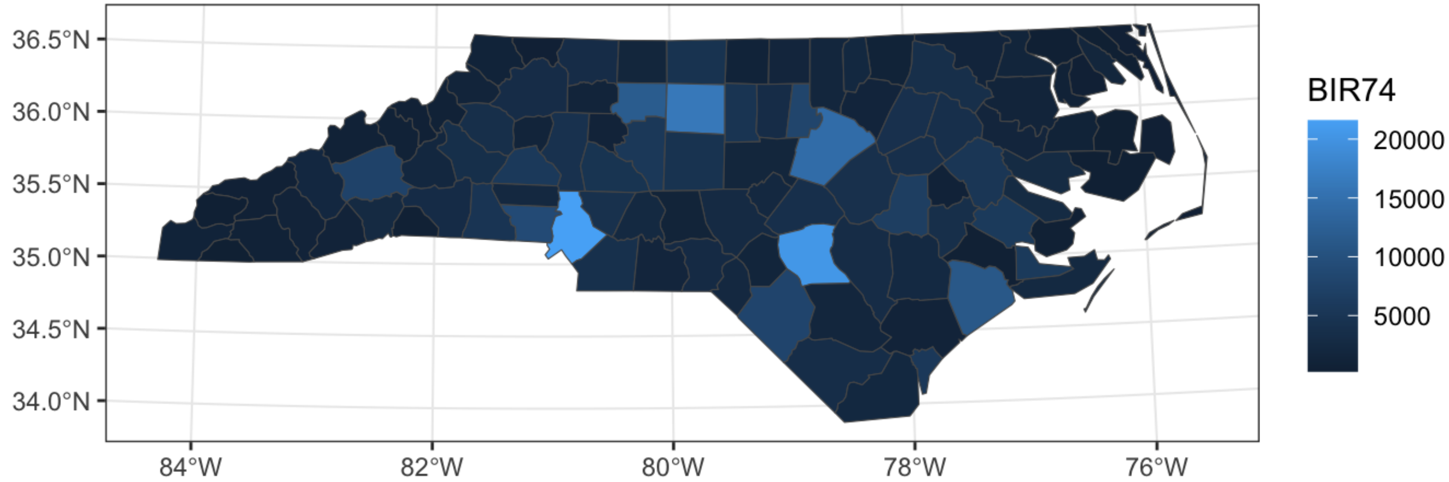
- Conditional Model

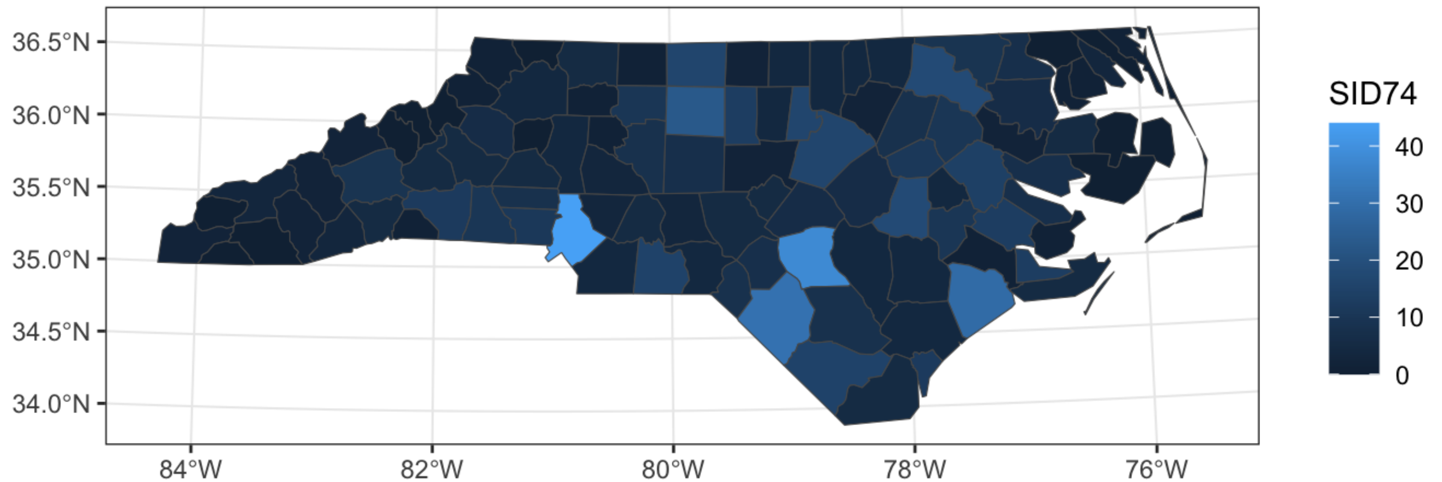
$$y(s_i)|\boldsymbol{y}_{-s_i} \sim N\left(X_{i\cdot}\beta + \phi \sum_{j=1}^{n} \frac{A_{ij}}{D_{ii}} \big(y(s_j) - X_{j\cdot}\beta\big),\ \sigma^2 D_{ii}^{-1}\right)$$

- Joint Model

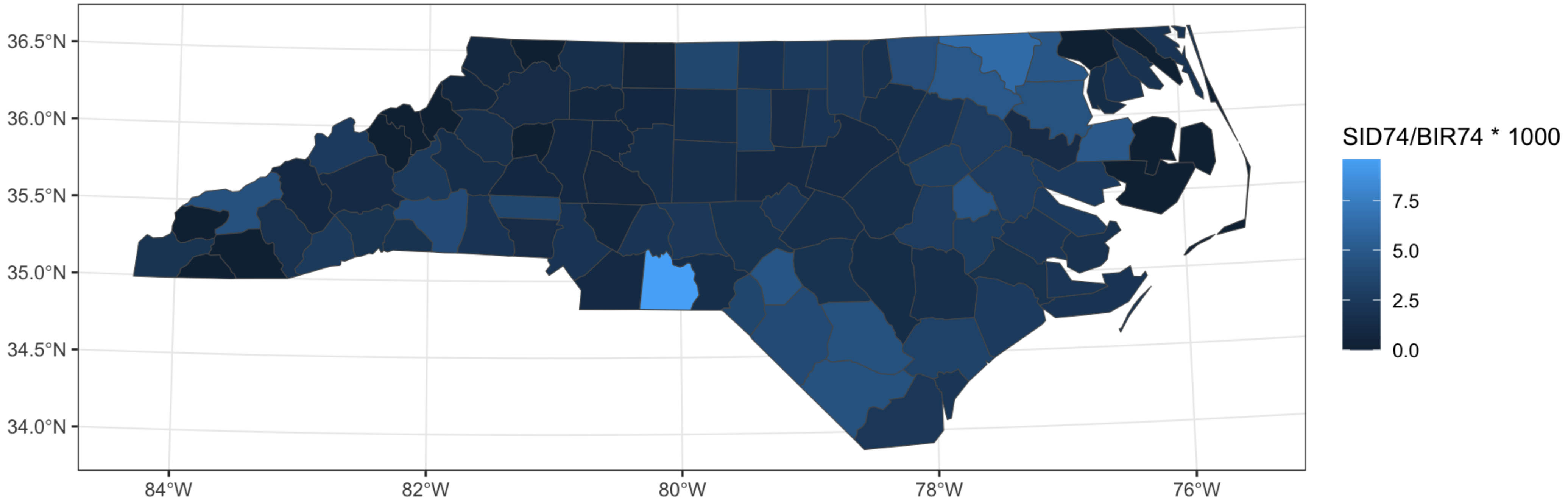
$$\boldsymbol{y} \sim N(\boldsymbol{X\beta},\ \sigma^2(\boldsymbol{D} - \phi\boldsymbol{A})^{-1})$$

Example - NC SIDS





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



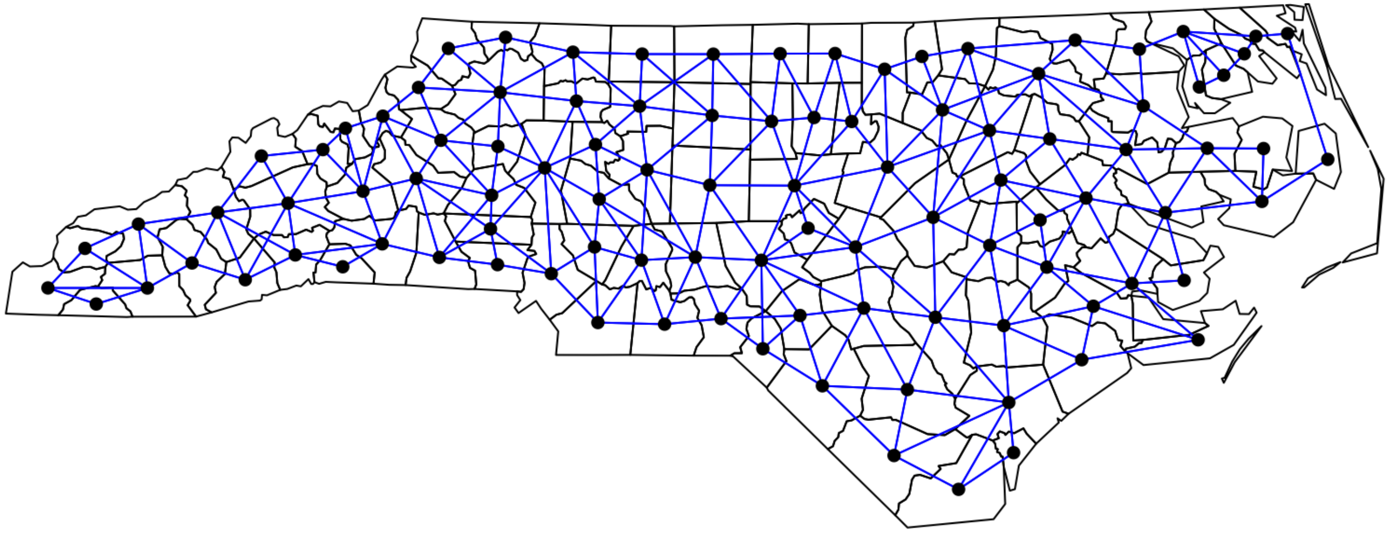
Using spdep + spatialreg

```r
library(spdep)

A = st_touches(nc, sparse=FALSE)
(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

```r
nc_coords = nc %>% st_centroid() %>% st_coordinates()

plot(st_geometry(nc))
plot(listW, nc_coords, add=TRUE, col="blue", pch=16)
```



Moran's I

```
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
```

```
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

```r
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
```

```r
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

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

summary(nc_car)
```

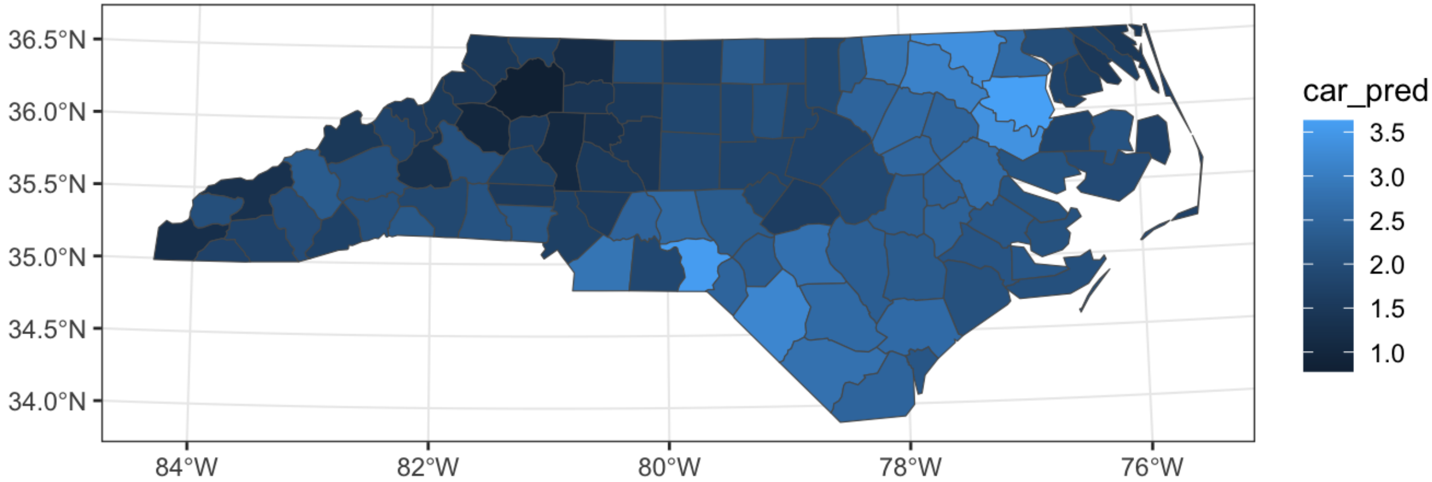
```
Call:
spatialreg::spautolm(formula = 1000 * SID74/BIR74 ~ 1, data = nc,
    listw = listW, family = "CAR")

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
```

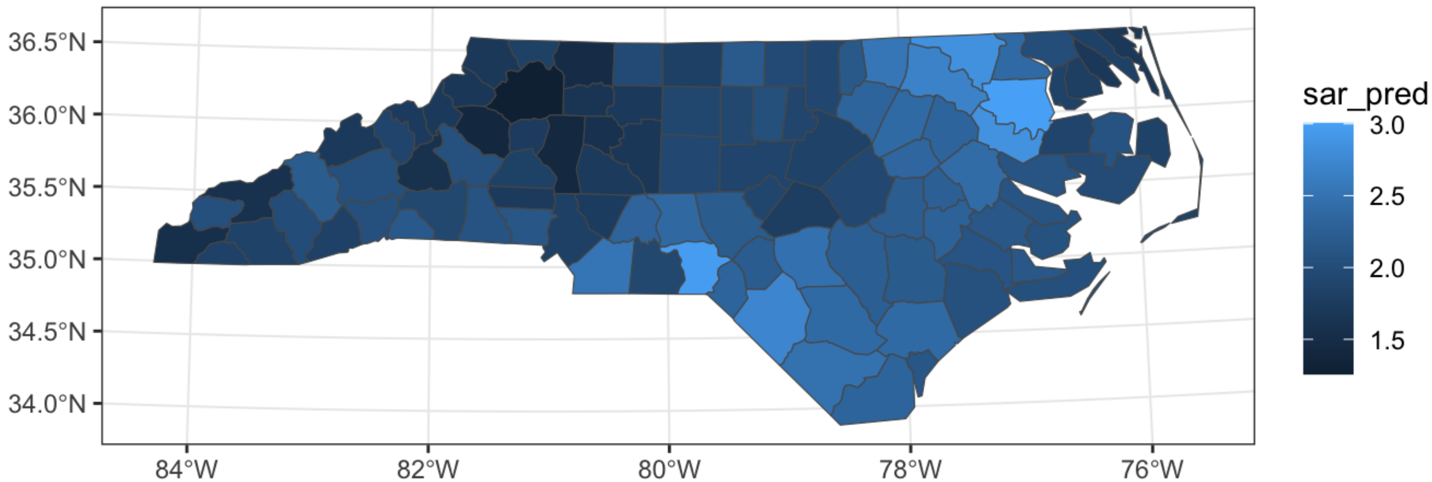
SAR Model

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

summary(nc_sar)
```

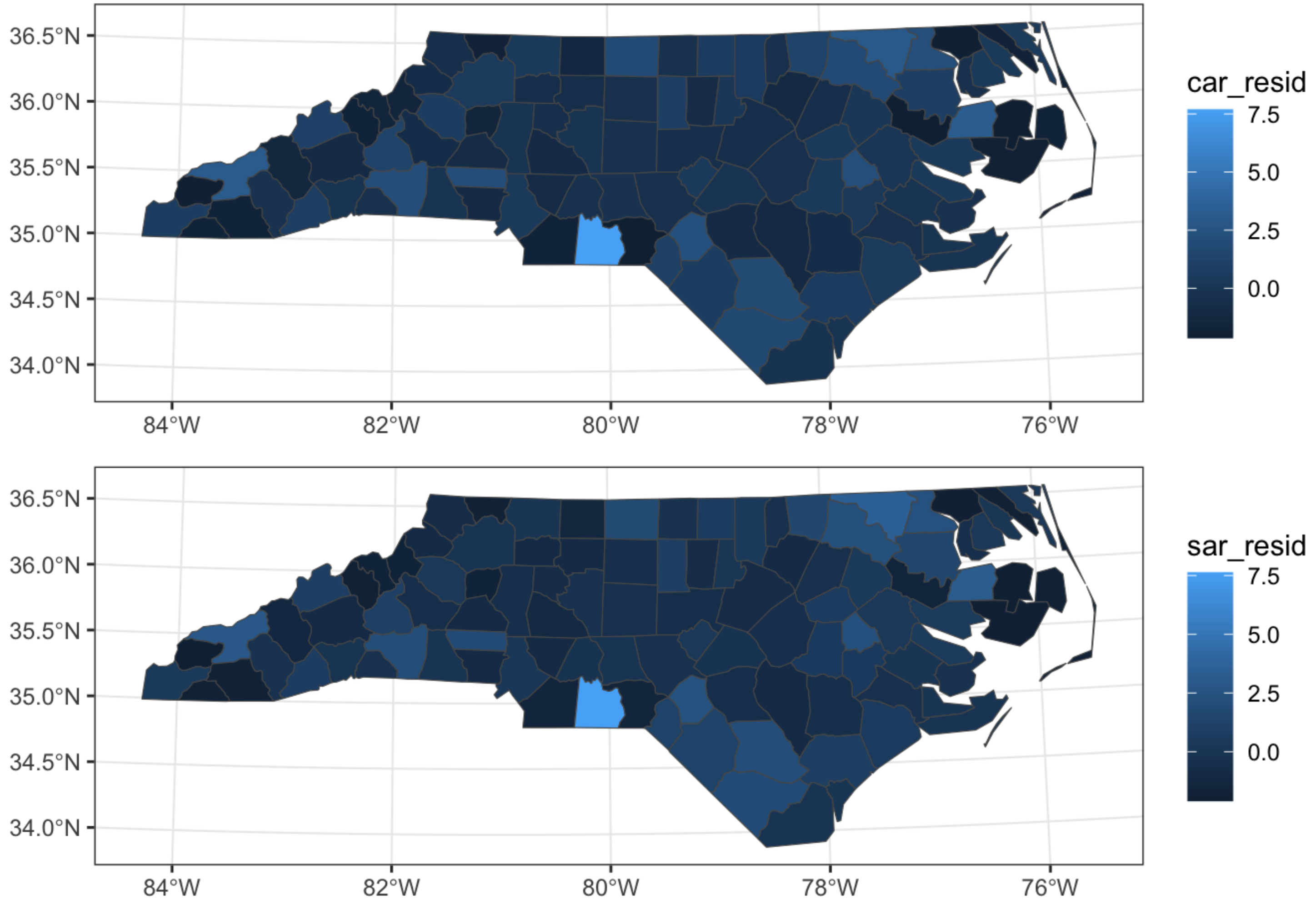
```
Call:
spatialreg::spautolm(formula = 1000 * SID74/BIR74 ~ 1, data = nc,
    listw = listW, family = "SAR")

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



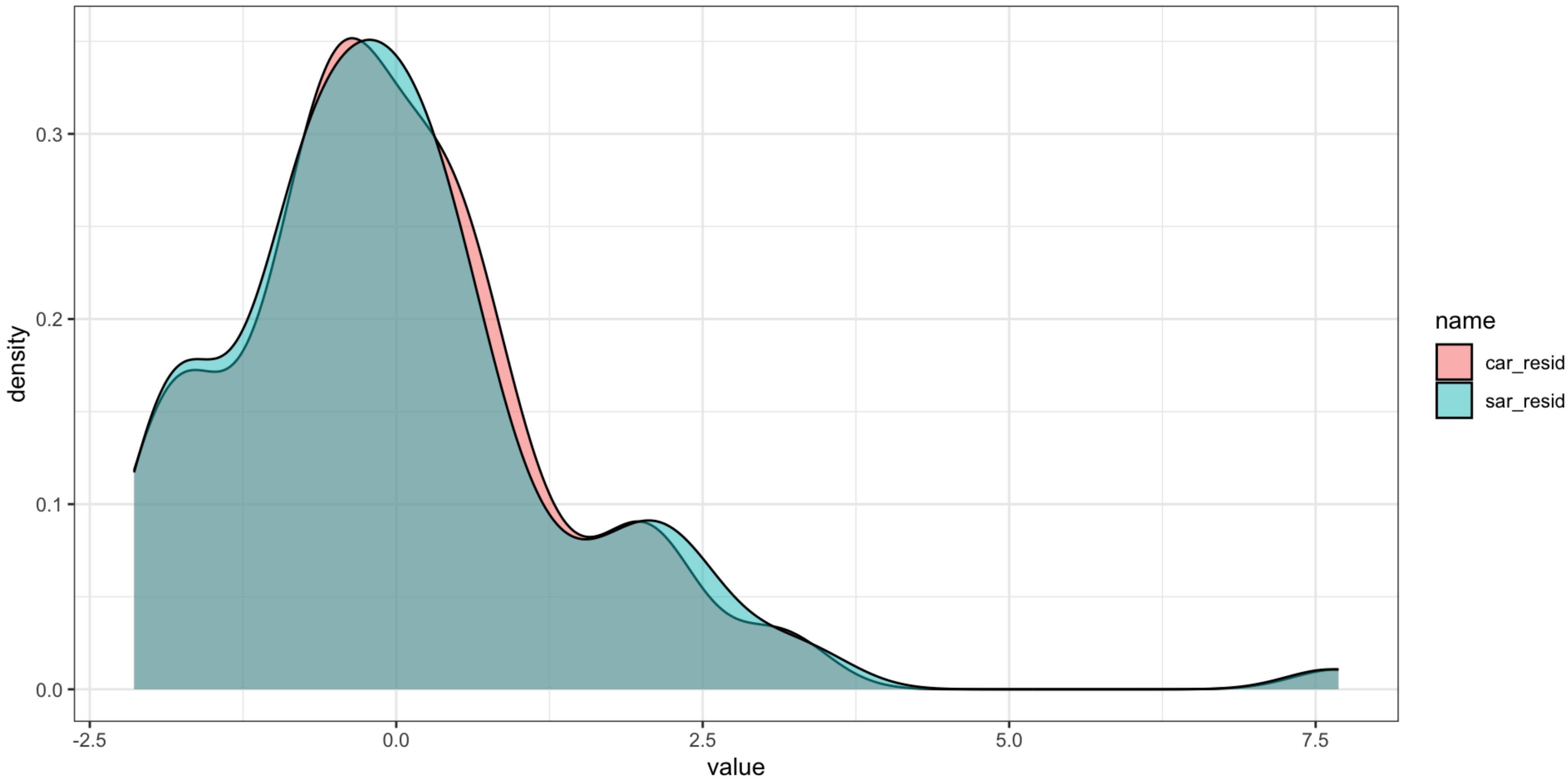


Residuals



36.5°N
36.0°N
35.5°N
35.0°N
34.5°N
34.0°N
84°W
82°W
80°W
78°W
76°W
car_resid
7.5
5.0
2.5
0.0
36.5°N
36.0°N
35.5°N
35.0°N
34.5°N
34.0°N
84°W
82°W
80°W
78°W
76°W
sar_resid
7.5
5.0
2.5
0.0

Residual distributions



Residual autocorrelation

```r
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 
```

```r
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     -0.0101010101 
         Variance 
     0.0036020941 
```

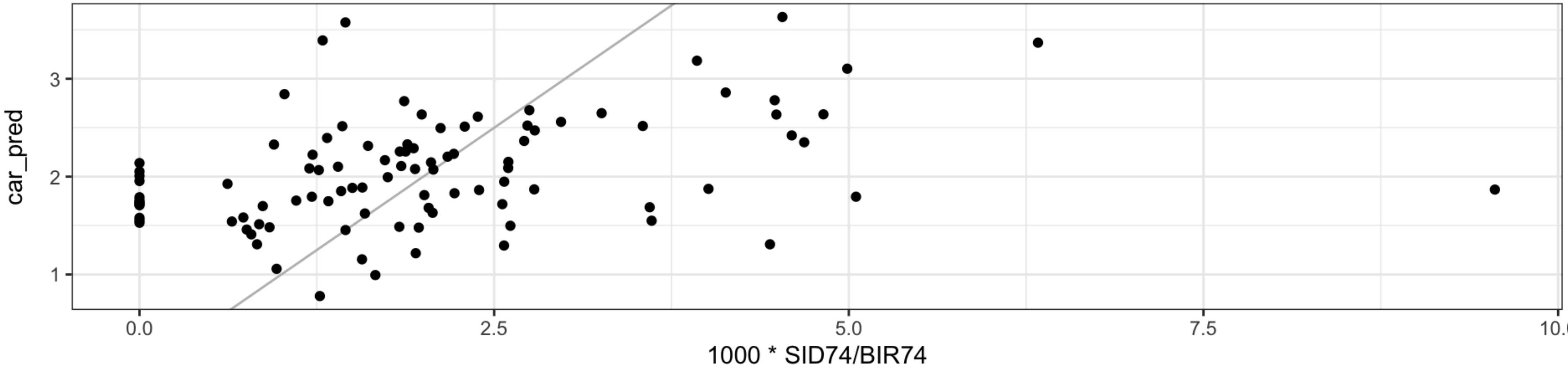
```r
spdep::moran.test(nc$car_resid, listW, alterr
```

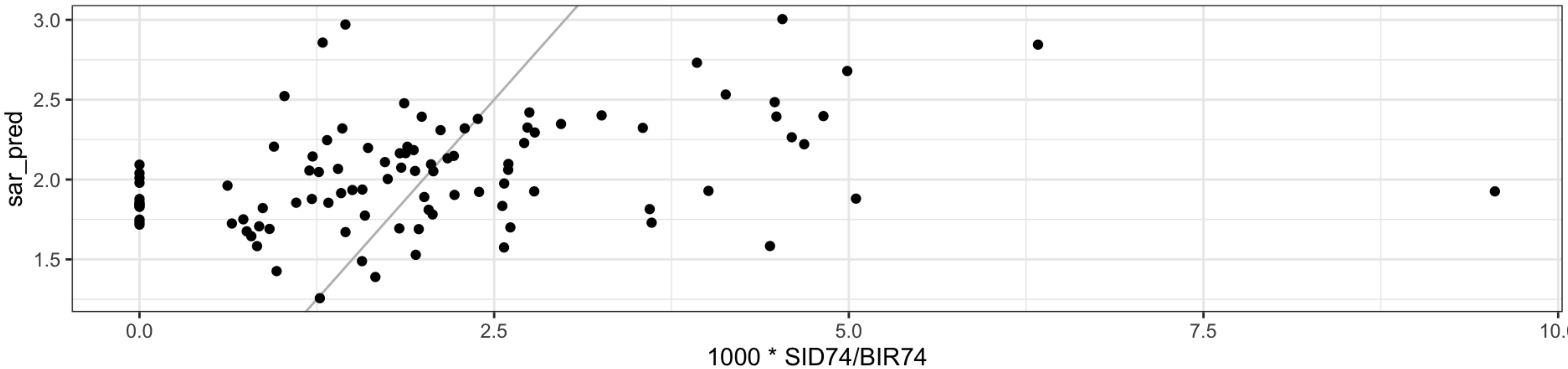
```
    Moran I test under randomisation

data:  nc$car_resid
weights: listW

Moran I statistic standard deviate =
-1.7952, p-value = 0.07261
alternative hypothesis: two.sided
sample estimates:
Moran I statistic       Expectation
     -0.117449316      -0.010101010
         Variance
      0.003575538
```

```r
spdep::moran.test(nc$sar_resid, listW, alterr
```

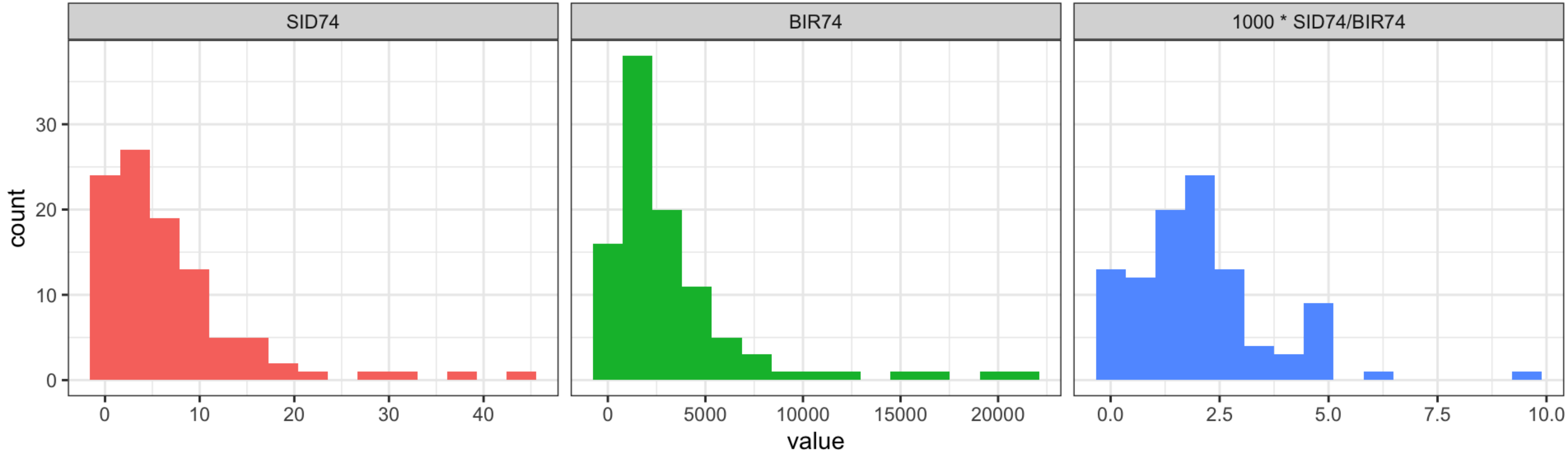
```
    Moran I test under randomisation

data:  nc$sar_resid
weights: listW

Moran I statistic standard deviate =
0.17958, p-value = 0.8575
alternative hypothesis: two.sided
sample estimates:
Moran I statistic       Expectation
     0.0006769074     -0.0101010101
         Variance
     0.0036020941
```

Predicted vs Observed



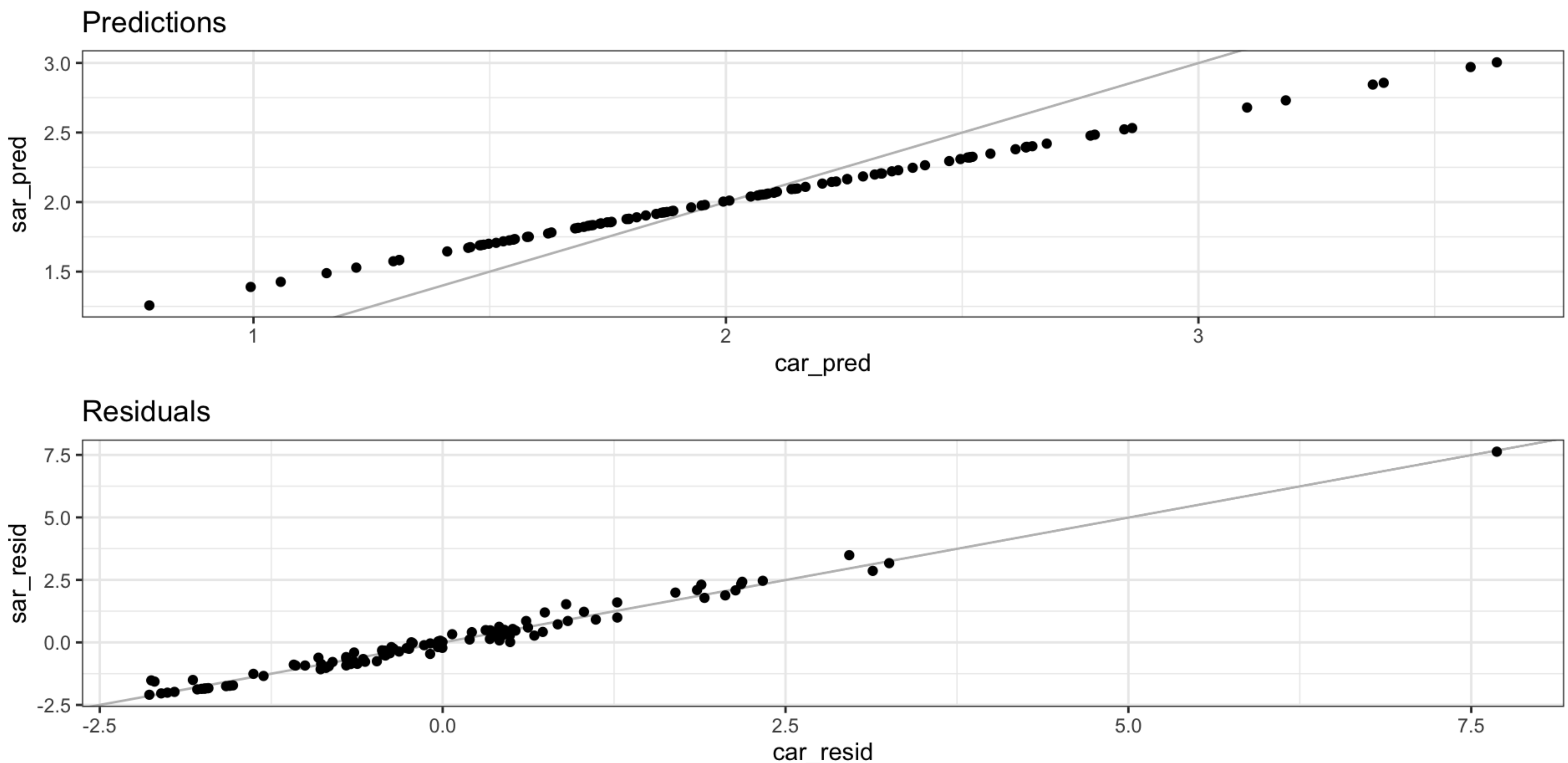


What's wrong?



SID74
BIR74
1000 * SID74/BIR74
count
value
30
20
10
0
0
10
20
30
40
0
5000
10000
15000
20000
0.0
2.5
5.0
7.5
10.0

Comparing CAR vs SAR.



Transforming the data

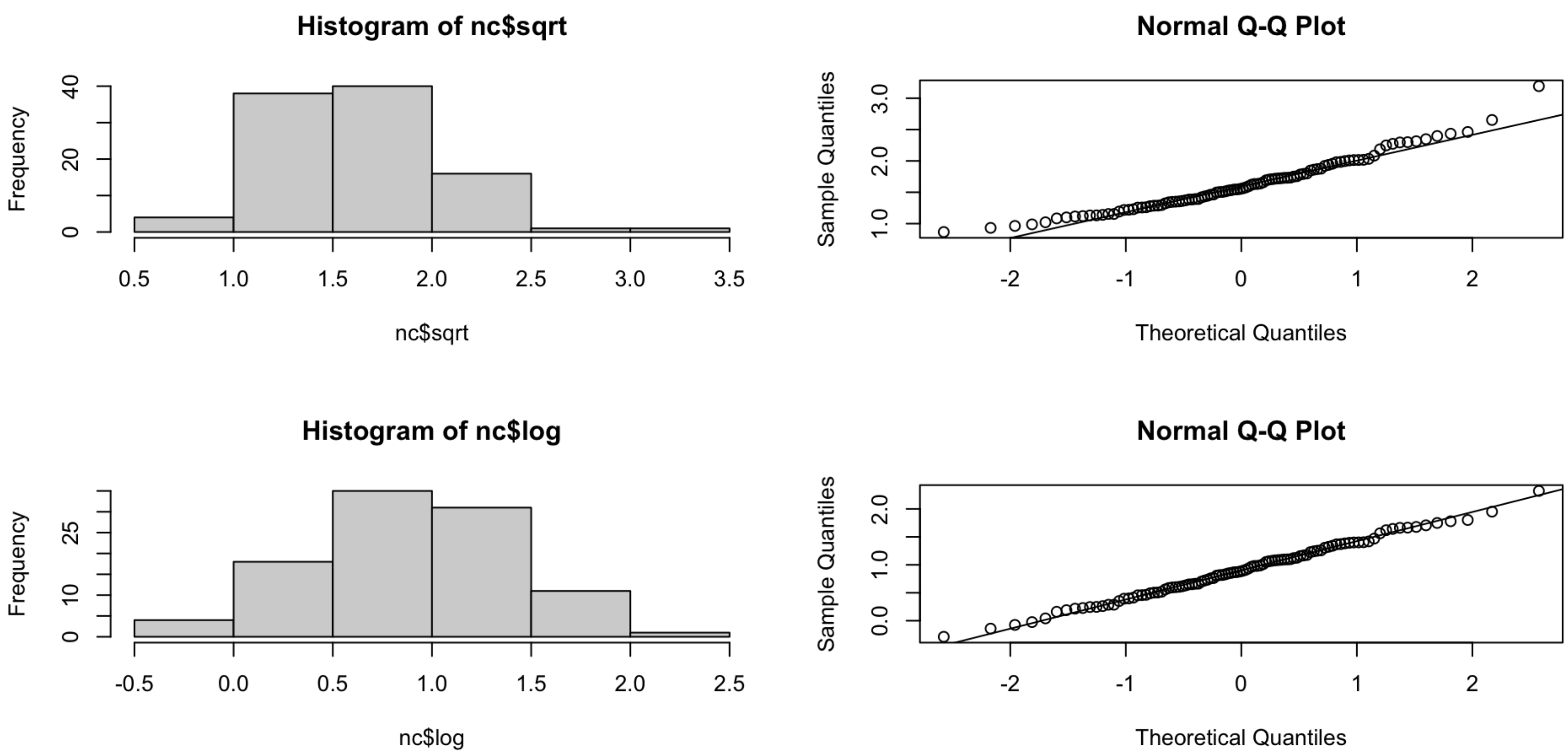
Freeman-Tukey's transformation

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

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

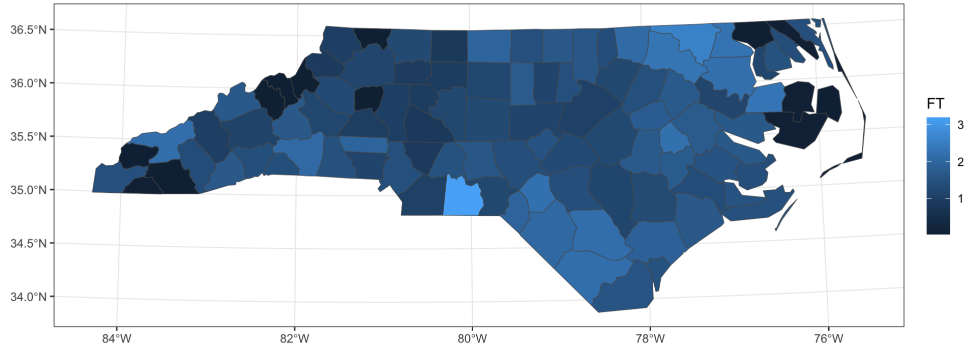
Other possibilities

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



FT transformation

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

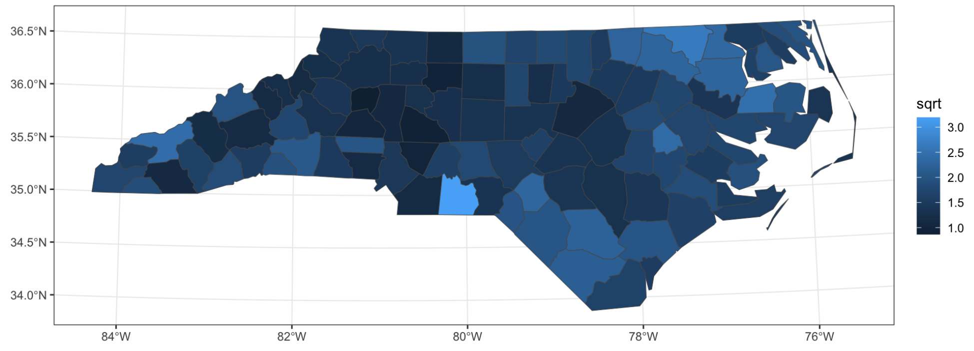



```r
spdep::moran.test(nc$FT, listW)
```

```
	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

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

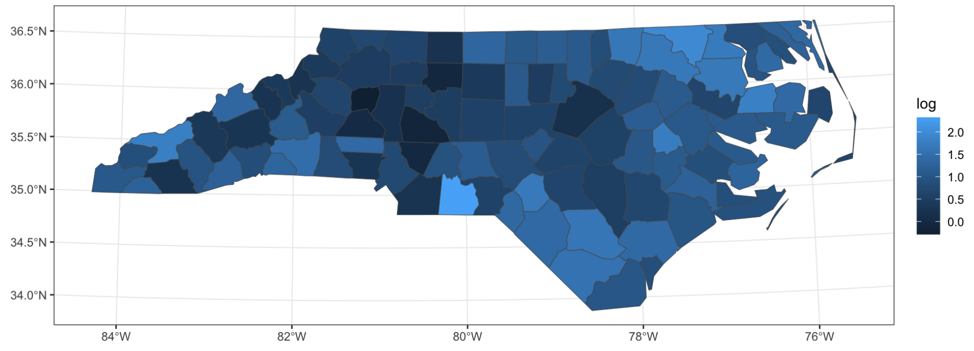



```
spdep::moran.test(nc$sqrt, listW)
```

```
	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

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




```r
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

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

AIC(nc_car_ft)
```

```
[1] 192.1781
```

```
AIC(nc_car_sqrt)
```

```
[1] 100.8898
```

```
AIC(nc_car_log)
```

```
[1] 134.644
```

SAR Model

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

AIC(nc_sar_ft)
```

```
[1] 191.9918
```

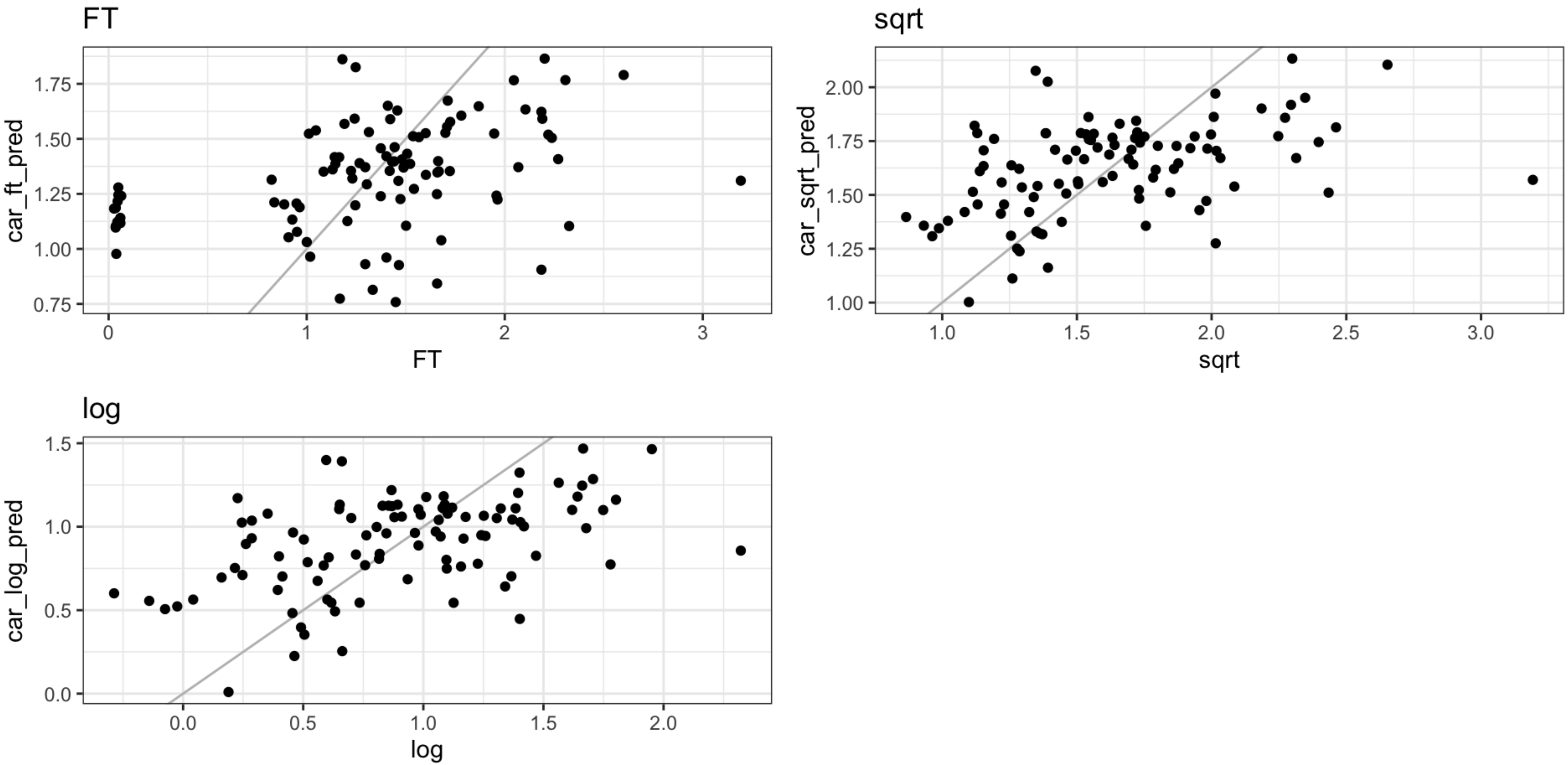
```r
AIC(nc_sar_sqrt)
```

```
[1] 102.717
```

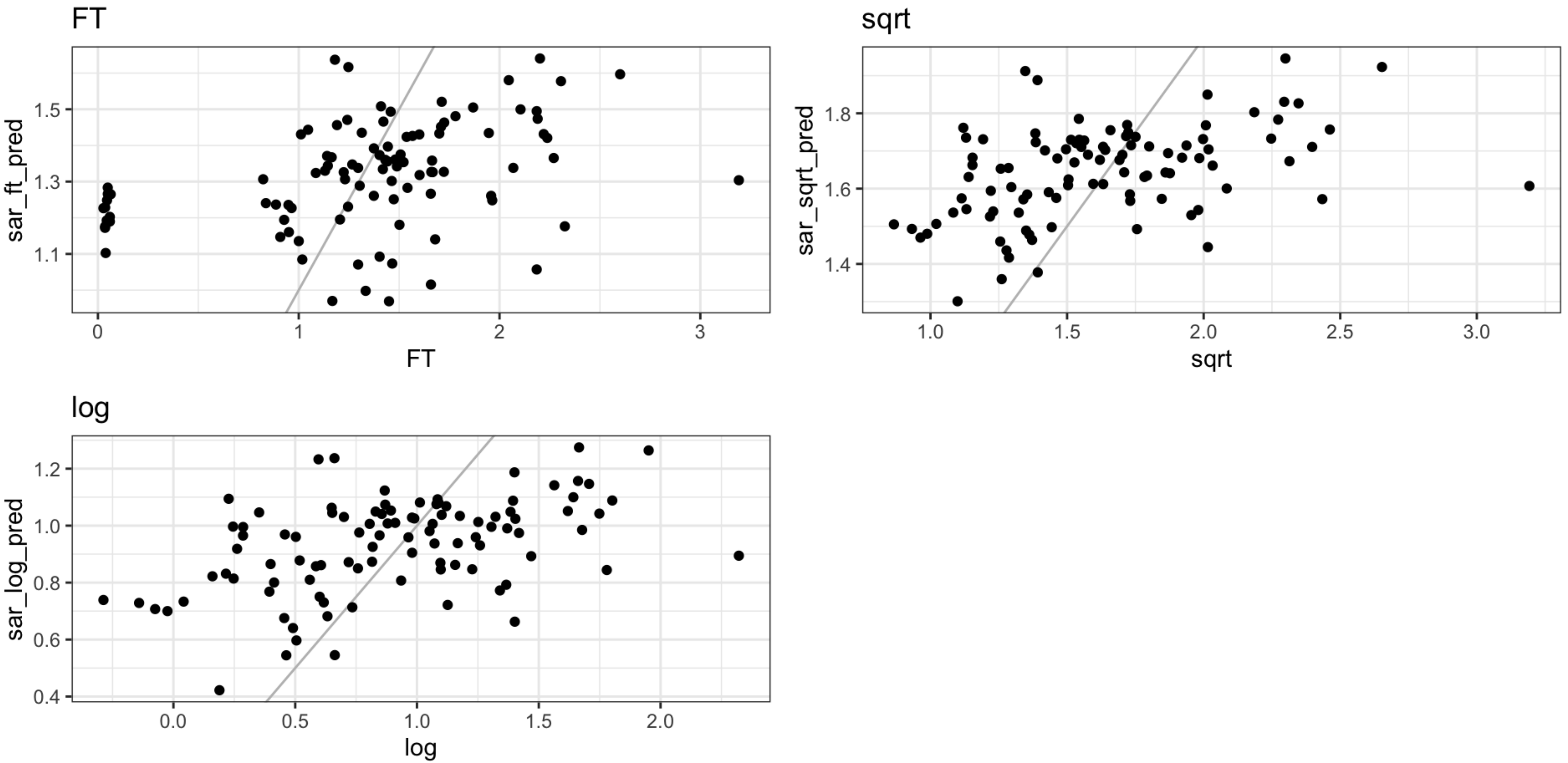
```r
AIC(nc_sar_log)
```

```
[1] 137.4095
```

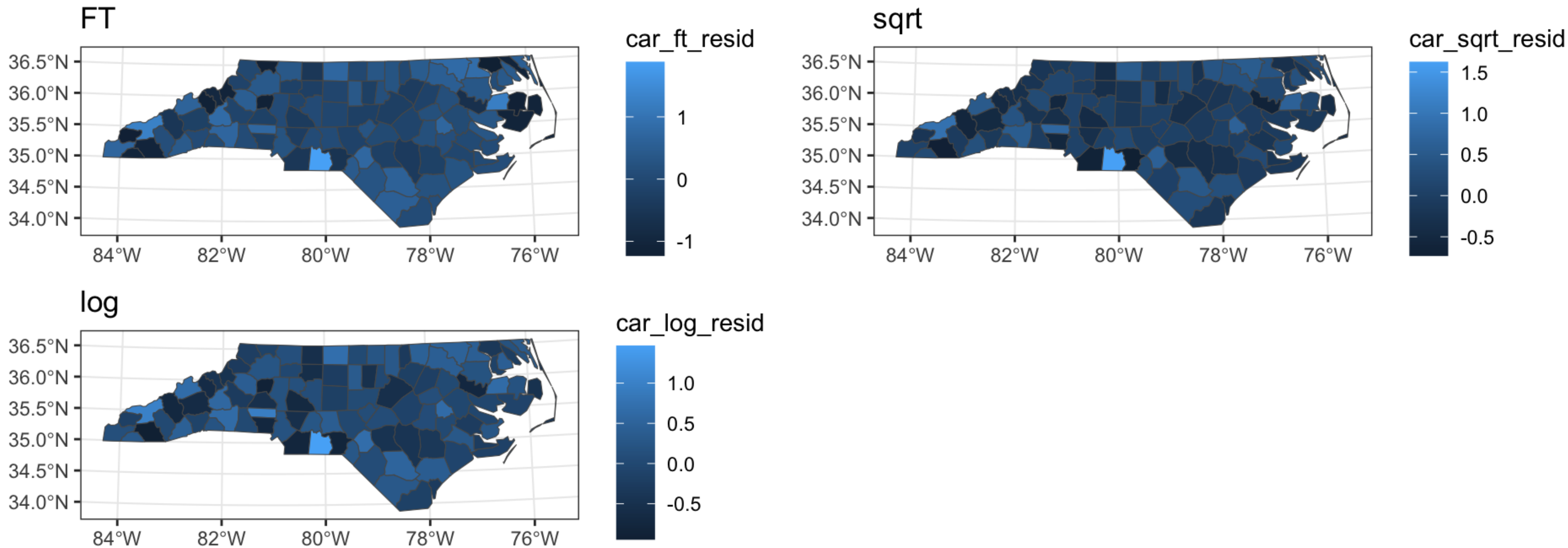
CAR predictions



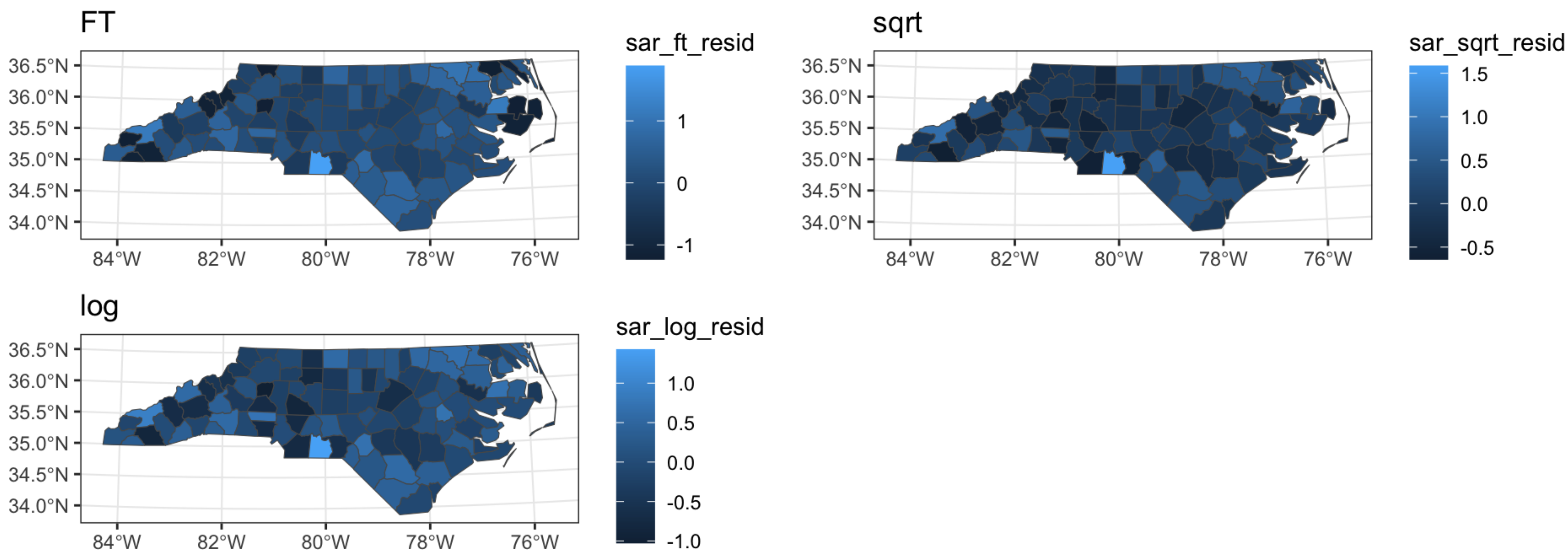
SAR predictions



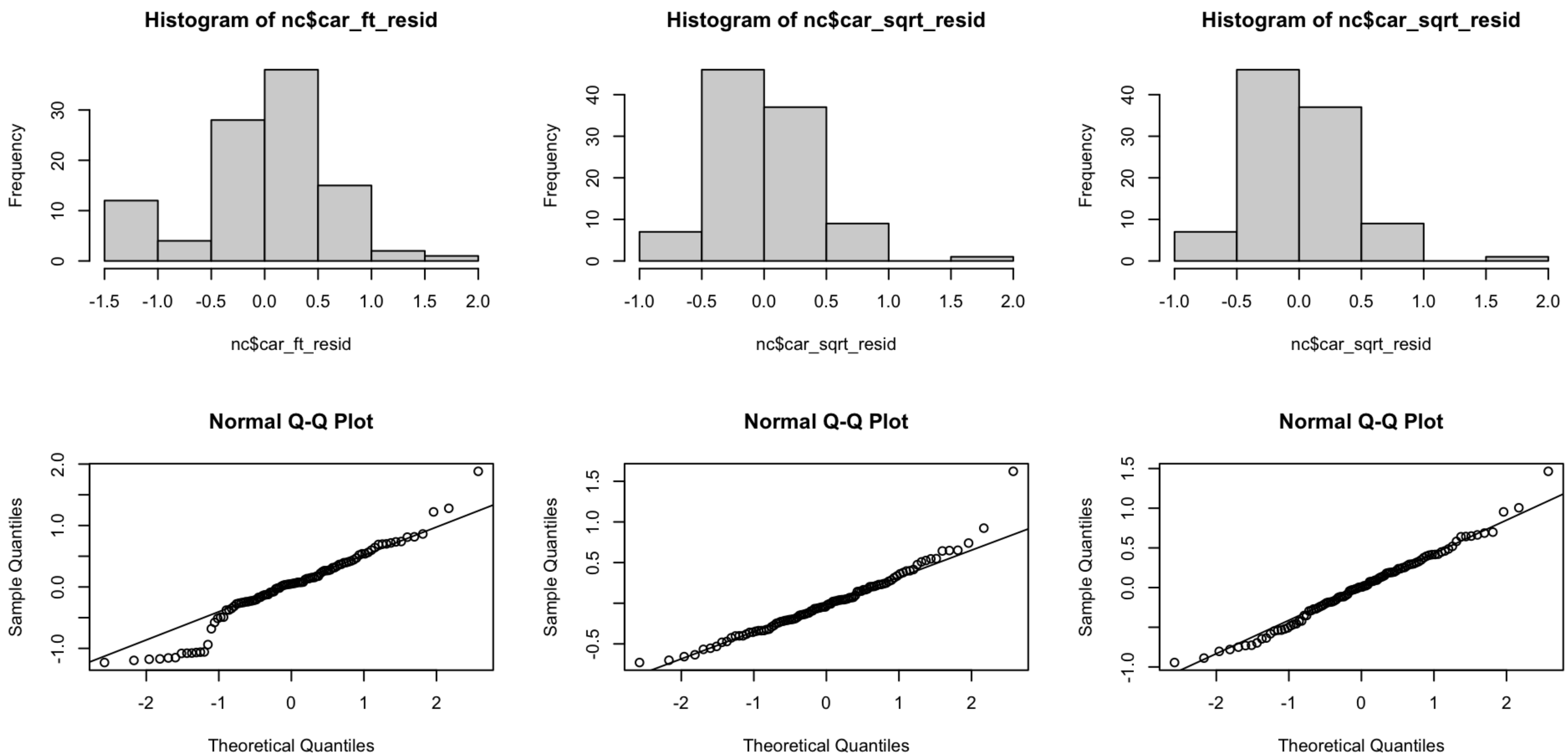
CAR residuals



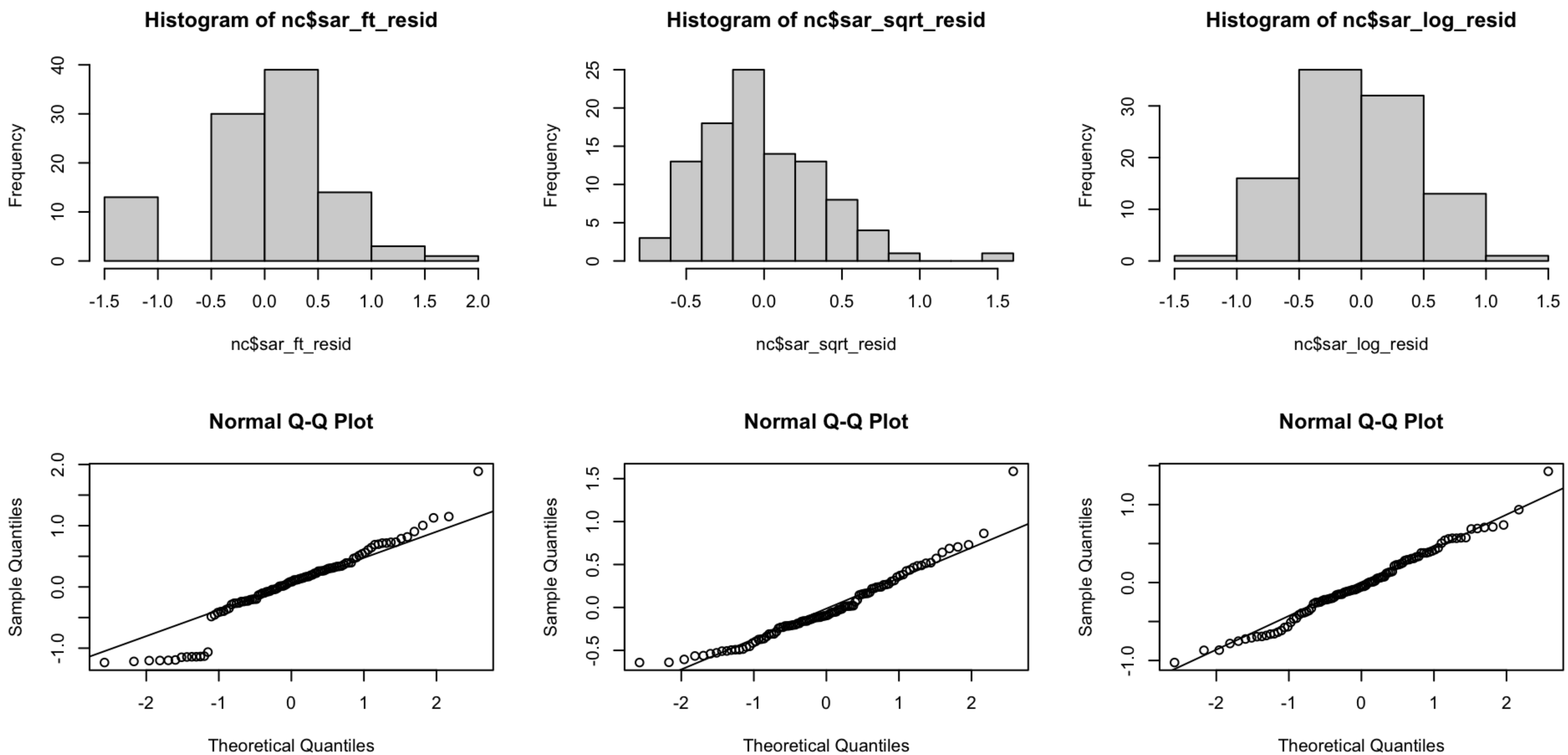
SAR predictions



CAR residual distributions



SAR residual distributions



Residual spatial autocorrelation

```r
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
```

```r
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

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

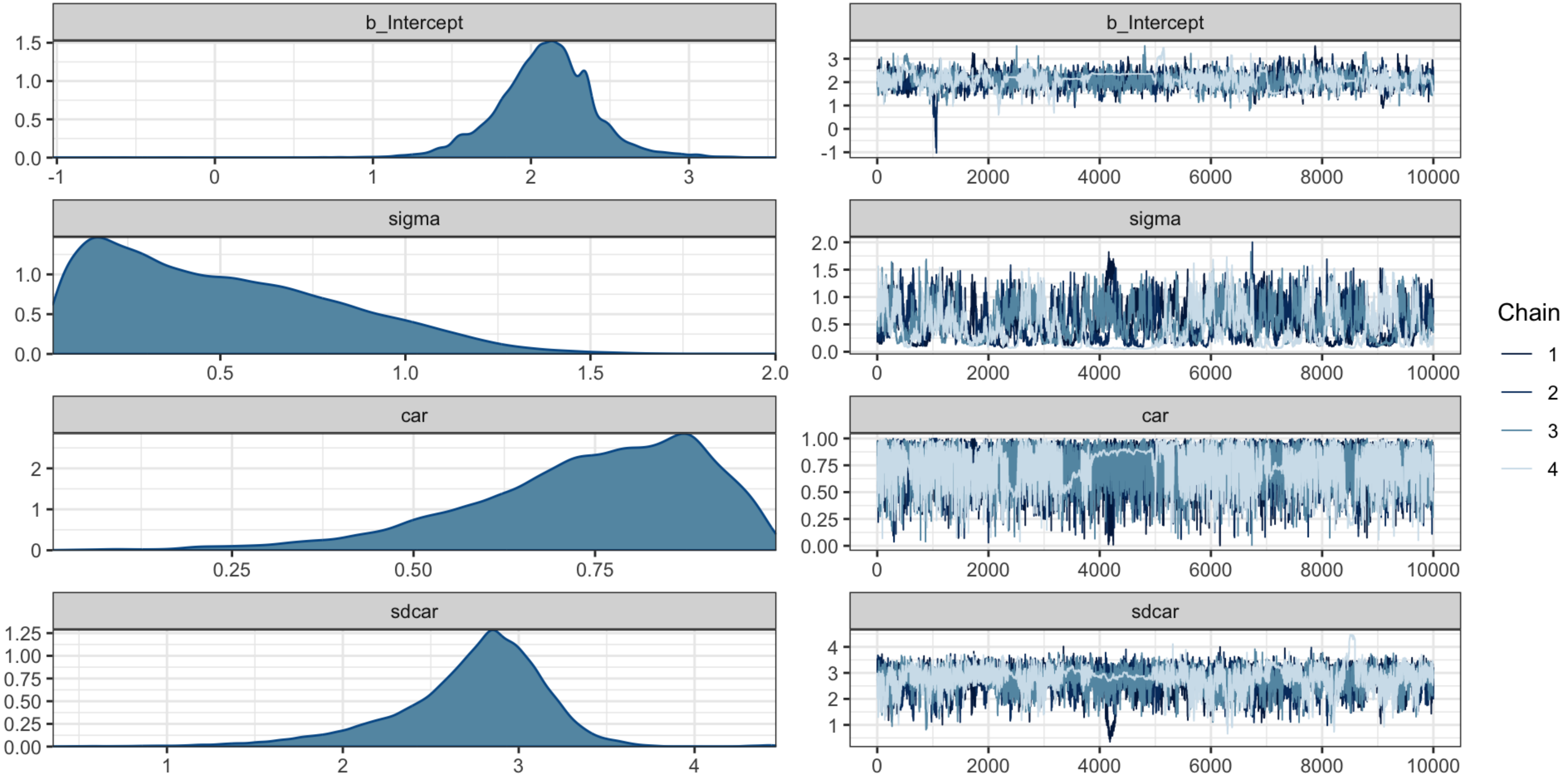
```
Running MCMC with 4 parallel chains...

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: 12.0 seconds.
Total execution time: 12.8 seconds.
```

```r
b_car
```

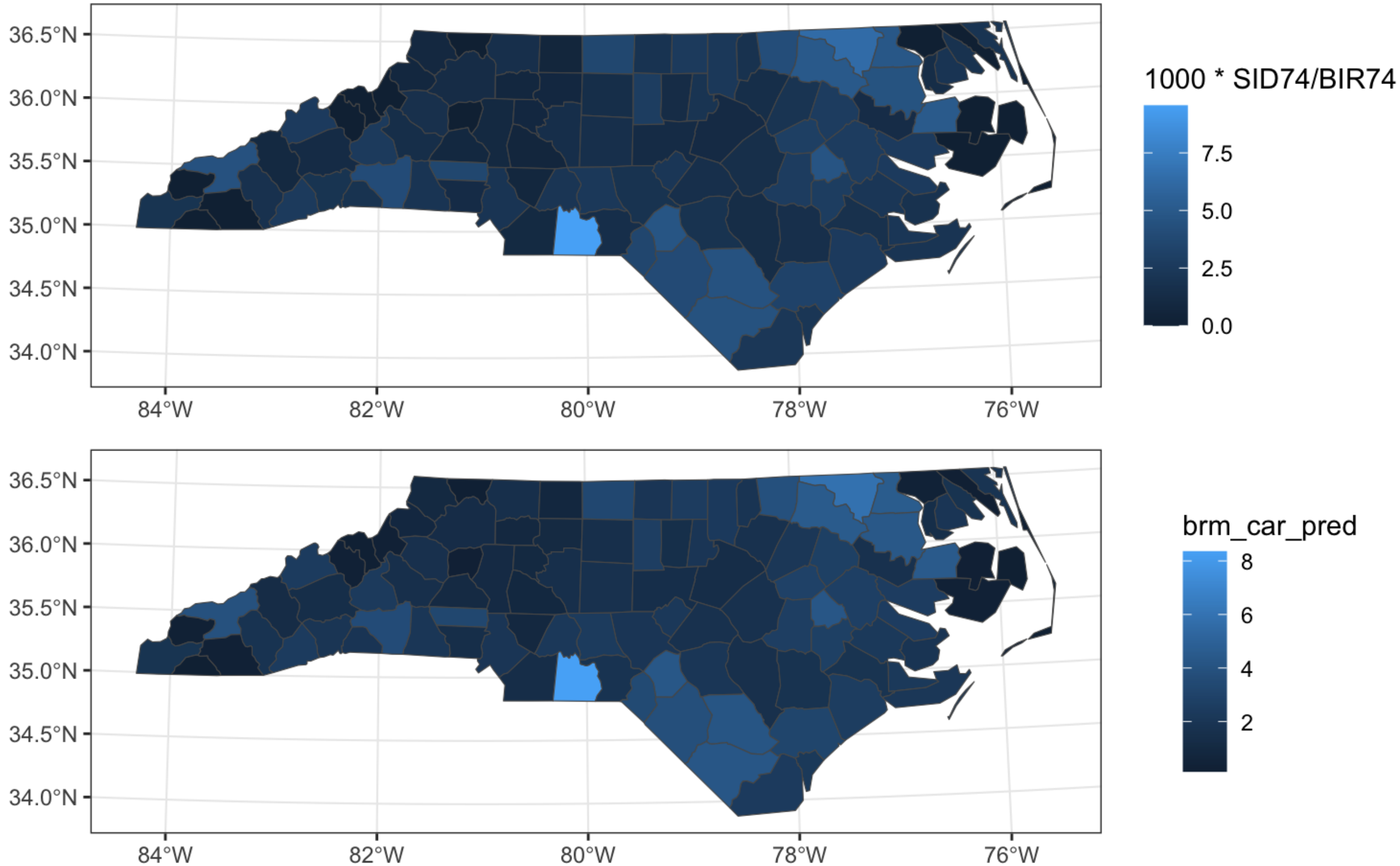
```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: 1000 * SID74/BIR74 ~ 1 + car(A)
   Data: nc (Number of observations: 100)
  Draws: 4 chains, each with iter = 20000; warmup = 10000; thin = 1;
         total post-warmup draws = 40000

Correlation Structures:
      Estimate Est.Error l-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

```r
plot(b_car)
```

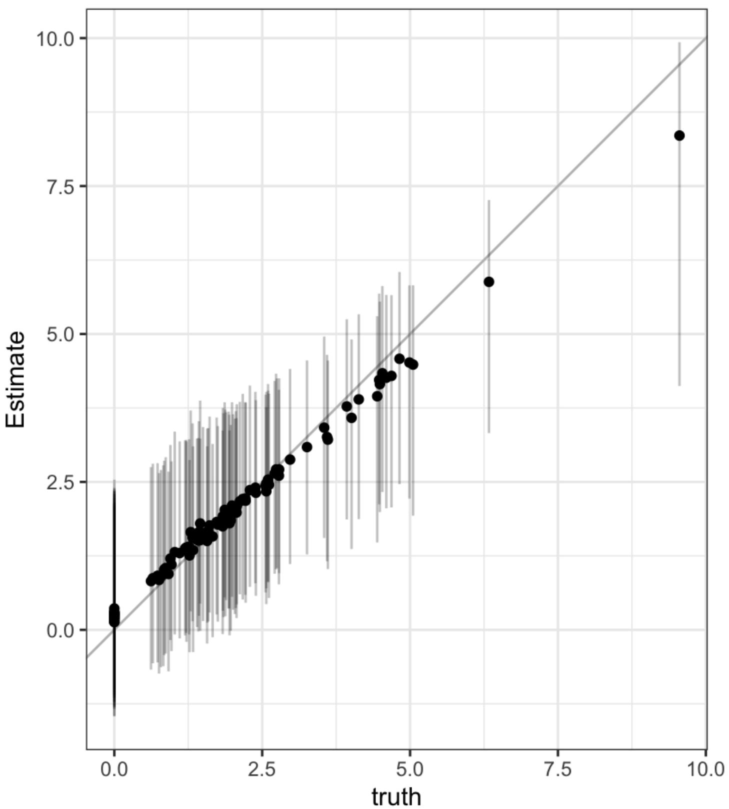


Predictions



Observed vs predicted

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



brms SAR

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

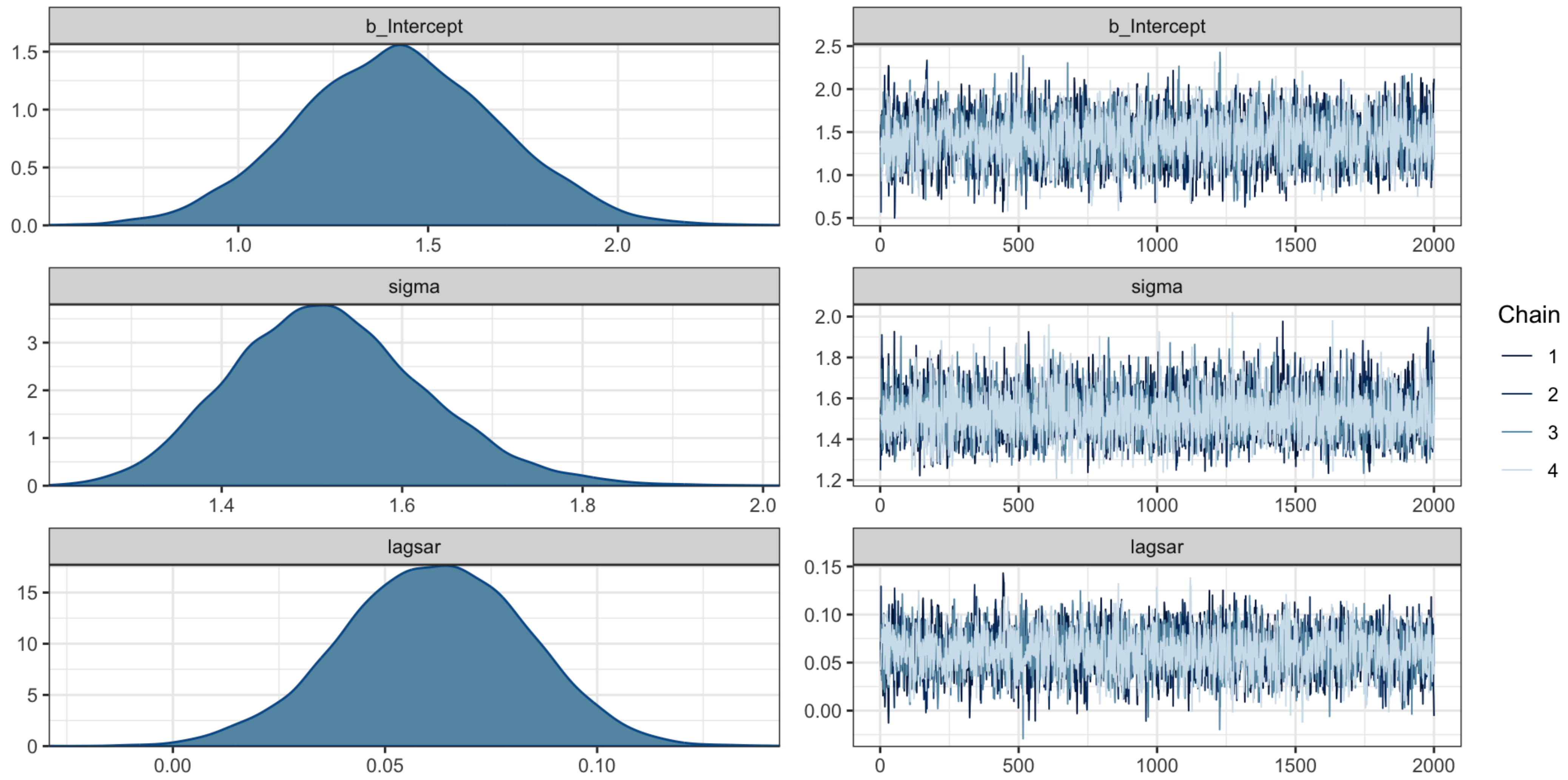
```
Running MCMC with 4 parallel chains...

Chain 3 finished in 2.9 seconds.
Chain 4 finished in 2.8 seconds.
Chain 1 finished in 2.9 seconds.
Chain 2 finished in 2.8 seconds.

All 4 chains finished successfully.
Mean chain execution time: 2.9 seconds.
Total execution time: 3.2 seconds.
```

```r
b_sar
```

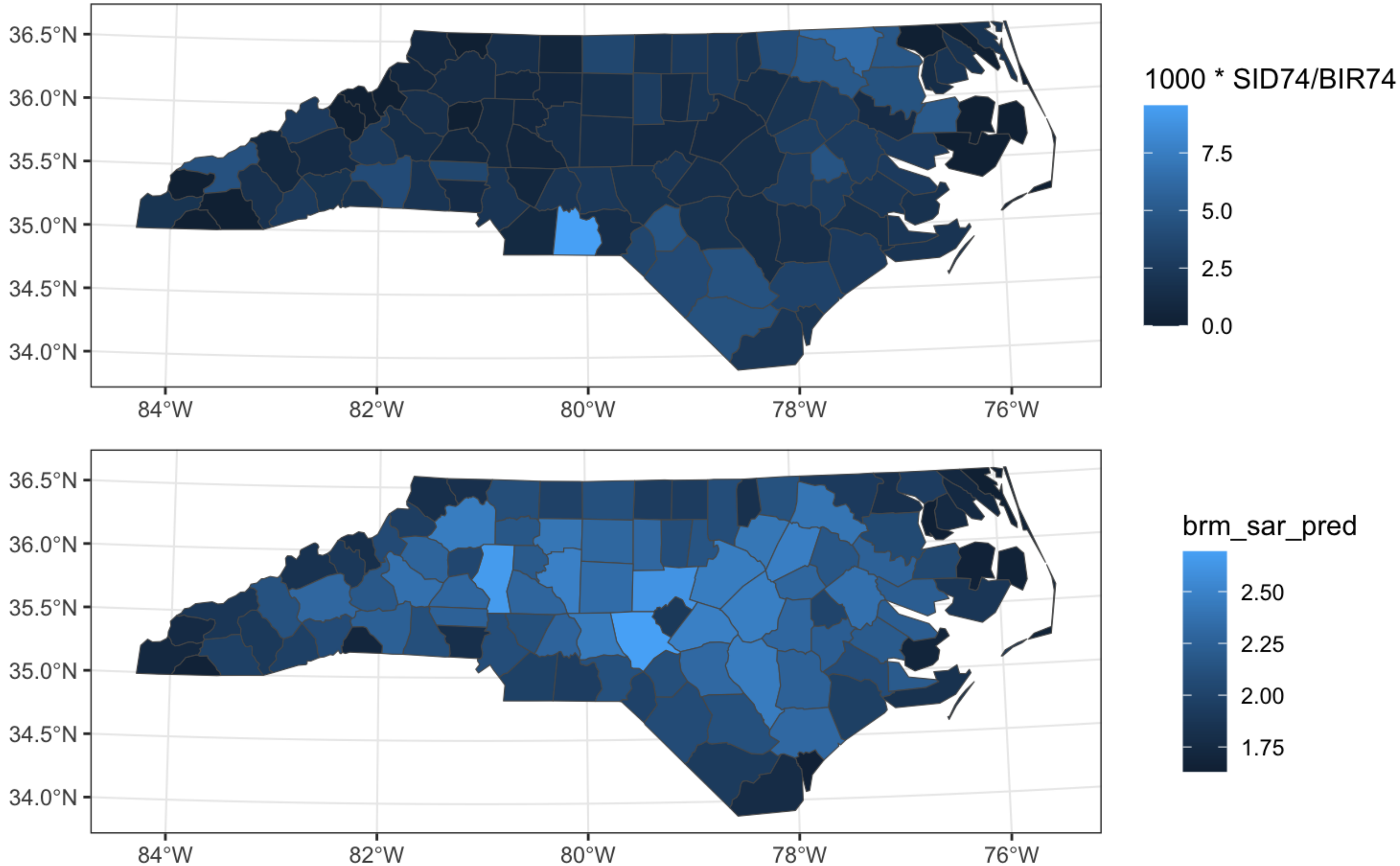
```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: 1000 * SID74/BIR74 ~ 1 + sar(listW)
   Data: nc (Number of observations: 100)
  Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
         total post-warmup draws = 8000

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
lagsar     3128     3877
```

Diagnostics

```r
plot(b_sar)
```

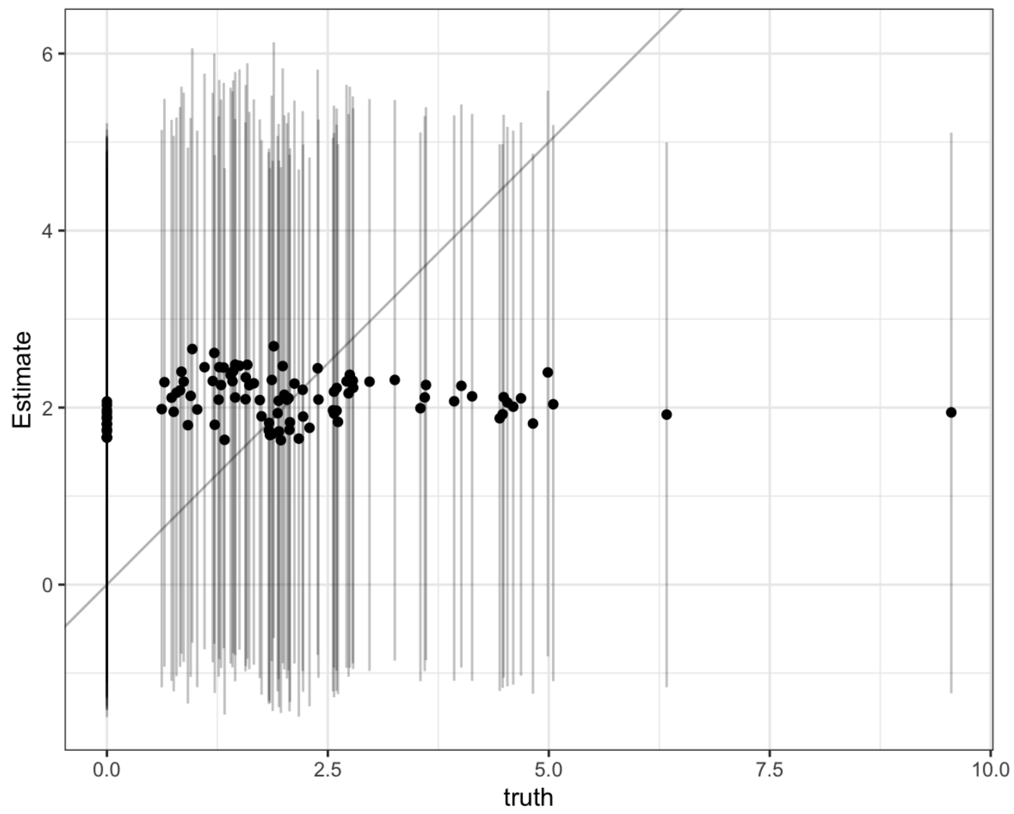


Predictions



Observed vs predicted

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



Brief Aside - CAR & SAR precision matrices

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

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

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