Assignment Block B

Bianca Neubert and Franziska Wehrmann

Theory

Markov random fields

Let Y denote a GMRF with mean vector μ and precision matrix Q of dimension $n \in \mathbb{N}$.

i) To show:

$$Y_i|Y_{-i} \sim \mathcal{N}\left(\mu_i - \frac{1}{q_{ii}} \sum_{j\in \mathrm{ne}(i)} q_{ij}(y_j - \mu_j), \frac{1}{q_{ii}}\right)$$

Proof: Since Y is a GMRF with expectation μ and precision matrix Q, we have for the conditional density of $Y_i|Y_{-i}$

$$\pi(y_i|y_{-i}) = \pi(y_i|y_{\mathrm{ne}(i)})$$

where $y_{\mathrm{ne}(i)}$ denotes the vector with indices in $\mathrm{ne}(i)$. Because the MRF is Gaussian, we have further for every $A \subset \{1, ..., n\}$

$$Y_A \sim \mathcal{N}(\mu_A, Q_A^{-1})$$

where $Y_A = (Y_i)_{i\in A}$, $\mu_A = (\mu_i)_{i\in A}$ and $Q_A = (q_{ij})_{i,j\in A}$ for the entries q_{ij} of Q.

Combining this, we get for the density of $Y_i|Y_{-i}$, since $Y_{\mathrm{ne}(i)}$ is also normally distributed and hence has strictly positive density, that

$$\begin{aligned}
\pi(y_i|y_{-i}) &= \frac{\pi(y_{\{i\}\cup \mathrm{ne}(i)})}{\pi(y_{\mathrm{ne}(i)})} \\
&\propto \exp\left\{-\frac{1}{2}(y_{\{i\}\cup\mathrm{ne}(i)} - \mu_{\{i\}\cup\mathrm{ne}(i)})^T Q_{\{i\}\cup\mathrm{ne}(i)}(y_{\{i\}\cup\mathrm{ne}(i)} - \mu_{\{i\}\cup\mathrm{ne}(i)}) \right. \\
&\quad \left. +\frac{1}{2}(y_{\mathrm{ne}(i)} - \mu_{\mathrm{ne}(i)})^T Q_{\mathrm{ne}(i)}(y_{\mathrm{ne}(i)} - \mu_{\mathrm{ne}(i)})\right\} \\
&= \exp\left\{-\frac{1}{2}\left[q_{ii}y_{ii}^2 + 2\sum_{j\in\mathrm{ne}(i)} y_i q_{ij} y_j + q_{ii}\mu_i^2 + 2\sum_{j\in\mathrm{ne}(i)} \mu_i q_{ij}\mu_j \right.\right. \\
&\quad \left.\left. -2q_{ii}\mu_i y_i - 2\sum_{j\in\mathrm{ne}(i)} \mu_i q_{ij} y_j - 2\sum_{j\in\mathrm{ne}(i)} y_i q_{ij}\mu_j\right]\right\} \\
&= \exp\left\{-\frac{q_{ii}}{2}\left[(y_i-\mu_i)^2 + \frac{2}{q_{ii}}(y_i - \mu_i)\sum_{j\in\mathrm{ne}(i)} q_{ij}(y_j - \mu_j)\right]\right\} \\
&\propto \exp\left\{-\frac{q_{ii}}{2}\left[y_i - \mu_i + \frac{1}{q_{ii}}\sum_{j\in\mathrm{ne}(i)} q_{ij}(y_j-\mu_j)\right]^2\right\}
\end{aligned}$$

which is the kernel of a normal distribution with the given expecation and variance.

iii) To show: $\mathbb{C}\text{or}[Y_i, Y_j | Y_{-ij}] = -\frac{q_{ij}}{\sqrt{q_{ii}q_{jj}}}$

Proof:

Local charcteristics

Suppose S is a finite set equipped with a symmetry relation $\sim$. For count outcomes, a specification commonly used in disease mapping is the *auto-Poisson* model where

$$\pi(y_i | y_{S\setminus i}) = \exp(-\mu_i)\frac{\mu_i^{y_i}}{y_i}$$
$$\log(\mu_i) = -\sum_{j \in \text{ne}(i)} y_j \;\; i \neq j$$

for $y \in \mathbb{N}_0$, $i \in S$. Determine the canonical potential with respect to $a = 0$.

Computation

Guerry's data on social morals

This task is based on Guerry's data on moral statistics including aggregated numbers of suicide (`suicids`). Let us first import and inspect the data.

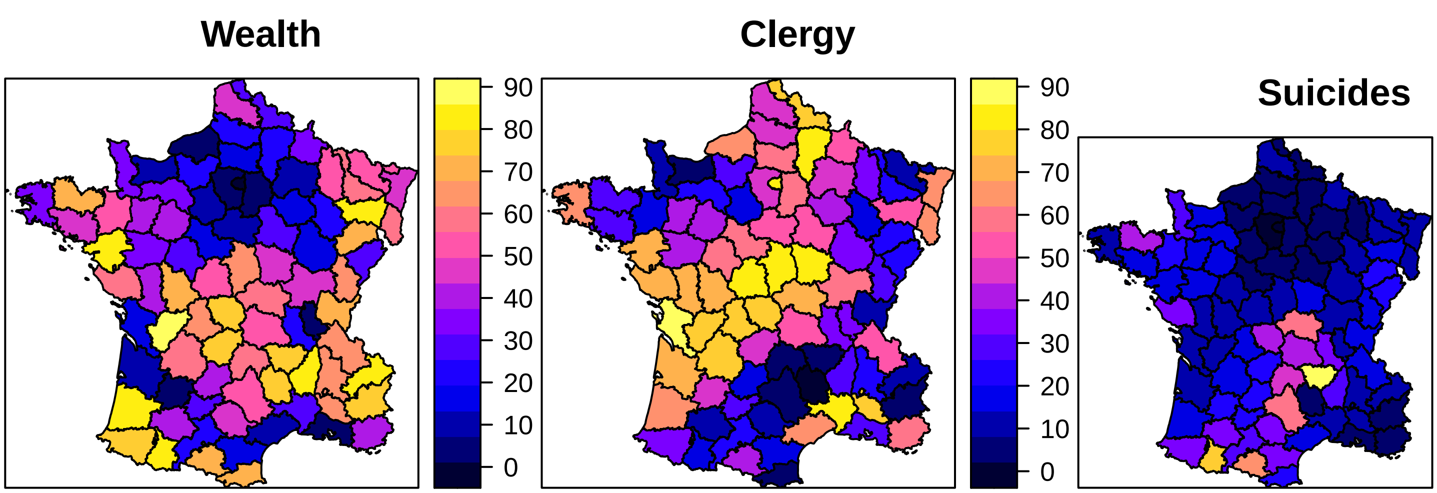
```
guerry <- readOGR("guerry/Guerry.shp")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/franziska/SpatStat/guerry/Guerry.shp", layer: "Guerry"
## with 85 features
## It has 26 fields
## Integer64 fields read as strings:  dept Crm_prs Crm_prp Litercy Donatns Infants Suicids MainCty Weal
```

We see that the variables are stored in form of strings, so we first change the class to integer of the three variables of interest in this assignment: `Suicids`, `Wealth` and `Clergy`

```
guerry$Suicids <- as.integer(guerry$Suicids)
guerry$Wealth <- as.integer(guerry$Wealth)
guerry$Clergy <- as.integer(guerry$Clergy)
```

Using `spplot` we can get a first impression on the spatial distribution of these variables.



Specify a (non-spatial) regression model of `Wealth` and `Clergy` on `Suicids`.

```r
reg <- lm(Suicids ~ Wealth + Clergy, data=guerry)
summary(reg)
```

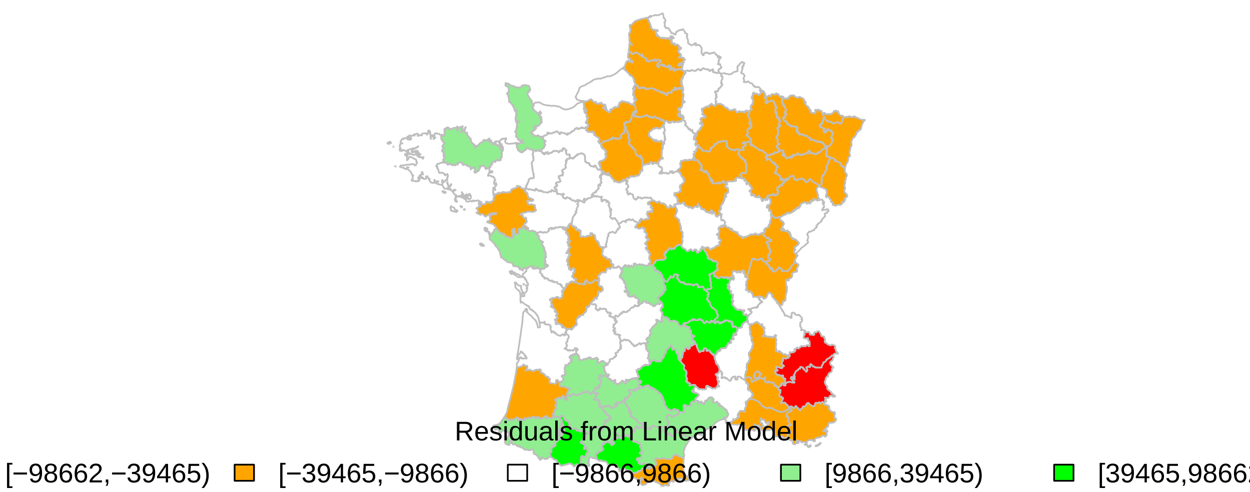
```
##
## Call:
## lm(formula = Suicids ~ Wealth + Clergy, data = guerry)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -56593 -14816  -5650   7680  98662
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30919.3     8476.1   3.648 0.000464 ***
## Wealth         486.7      120.5   4.038 0.000121 ***
## Clergy        -355.4      122.5  -2.902 0.004760 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27560 on 82 degrees of freedom
## Multiple R-squared:  0.2527, Adjusted R-squared:  0.2345
## F-statistic: 13.86 on 2 and 82 DF,  p-value: 6.508e-06
```

We can see, that in this the coefficients of Clergy and Wealth are not significant which at first does not indicate a relationship between these and Suicides. This model does not explain the variation in Suicides very well (e.g. Multiple R-squared = 0.011).

We can plot the residuals: First, we fix color palette once for all plots of this type. This way it is easier to compare the results later.

```r
m = 98661.66  # maximum residual of (lm, car, sar, sdm)
breaks = round(c(-m, -0.4*m, -0.1*m, 0.1*m, 0.4*m, m), digits=0)
res.palette <- colorRampPalette(c("red","orange","white", "lightgreen","green"), space = "rgb")
pal <- res.palette(5)
```

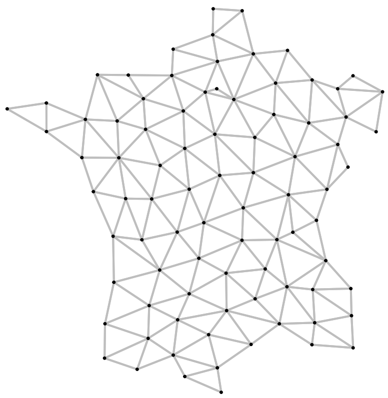
Now we can plot the residuals of the linear model.



TODO: Moron manual ?? Meinst du Moran's I :)

To test for autocorrelation in the data, we can use Moran's I and Geary's C. For this, we need to specify a neighborhood structure in a matrix.

```r
xy <- coordinates(guerry)
W_cont_el <- poly2nb(guerry, queen=FALSE)
W_cont_el_mat <- nb2listw(W_cont_el, style="B", zero.policy=TRUE)
plot(W_cont_el_mat, coords=xy, cex=0.1, col="gray")
```



For the Moran's I ad Geary's C we get the following results

```
##
##  Moran I test under randomisation
##
## data:  guerry$Suicids
## weights: W_cont_el_mat
##
## Moran I statistic standard deviate = 6.4289, p-value = 6.428e-11
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##       0.404857670      -0.011904762       0.004202492
```

```
##
##  Geary C test under randomisation
##
## data:  guerry$Suicids
## weights: W_cont_el_mat
##
## Geary C statistic standard deviate = 4.2472, p-value = 1.082e-05
## alternative hypothesis: Expectation greater than statistic
## sample estimates:
## Geary C statistic       Expectation          Variance
##        0.55293883        1.00000000        0.01107975
```

The Moran's I test statistic and its expectation are both close to zero (0.074 and -0.011). However, the Moran's I test statistic is larger than the expectation which is an indication for a positive autocorrelation. If we look the given p-value we can reject the null of no autocorrelation for a 0.1-niveau, but not for a 0.05-niveau. On the other hand, Geary's C is close to 1 and the p-value with 0.197 suggests that the null of no autocorrelation can not be rejeceted for a 0.1-niveau. To sum this up, there seems to be very little global autocorrelation in the data, if there is any at all.

TODO: whats the difference? to moran below? , Interpretation Warum moran with residuals? mc nicht?

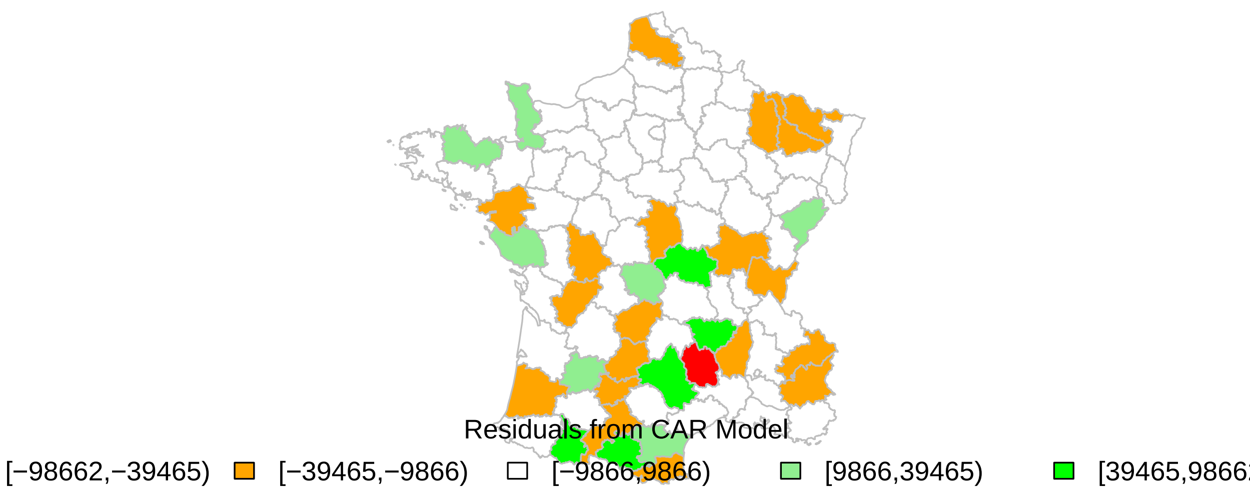
CAR

Let us compute a CAR model.

```r
car.out <- spdep::spautolm(Suicids ~ Wealth + Clergy, data=guerry,
                           listw=W_cont_el_mat, family="CAR")
mod.car <- fitted(car.out)
summary(car.out)
```

```
##
## Call: spdep::spautolm(formula = Suicids ~ Wealth + Clergy, data = guerry,
##     listw = W_cont_el_mat, family = "CAR")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -81951.2  -8586.6  -2583.9   3722.5  94835.8
##
## Coefficients:
##             Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 27409.16    9856.88  2.7807 0.0054240
## Wealth        421.99     123.21  3.4250 0.0006147
## Clergy       -275.75     128.56 -2.1449 0.0319596
##
## Lambda: 0.15913 LR test value: 15.459 p-value: 8.4312e-05
## Numerical Hessian standard error of lambda: 0.018933
##
## Log likelihood: -980.4015
## ML residual variance (sigma squared): 551060000, (sigma: 23475)
## Number of observations: 85
## Number of parameters estimated: 5
## AIC: 1970.8
```

Again, the coefficients for Wealth and Clergy are not significant. The plot of the residuals:



SAR

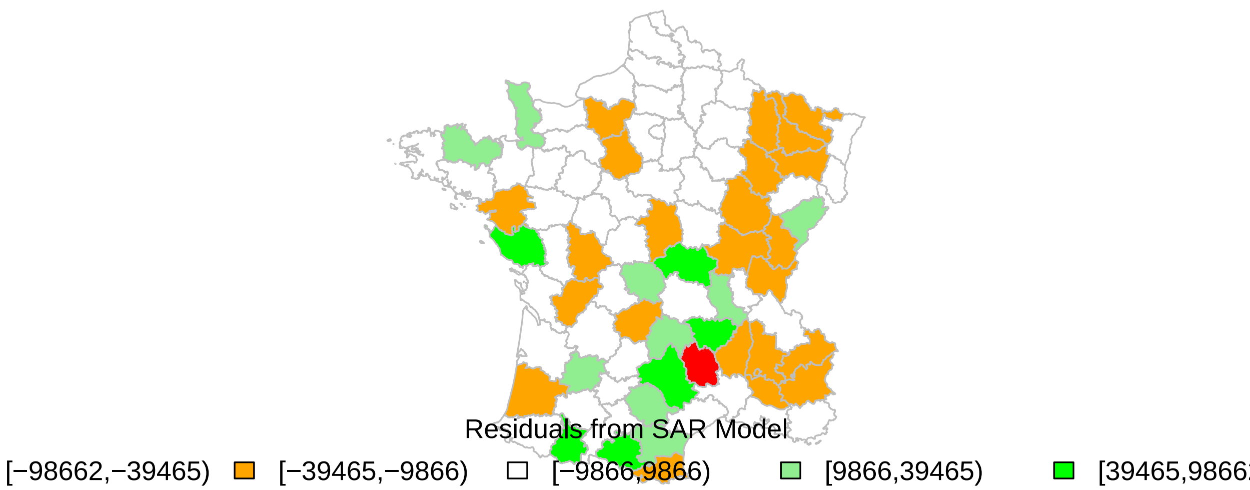
Now, the same for SAR

```r
mod.sar <- lagsarlm(Suicids ~ Wealth + Clergy, data=guerry,
                    listw=W_cont_el_mat, zero.policy=T, tol.solve=1e-12)
summary(mod.sar)
```

```
##
```

```
## Call:lagsarlm(formula = Suicids ~ Wealth + Clergy, data = guerry,
##     listw = W_cont_el_mat, zero.policy = T, tol.solve = 1e-12)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -71691.5 -10439.1   -4823.3    6113.9   94298.9
##
## Type: lag
## Coefficients: (numerical Hessian approximate standard errors)
##             Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 16317.90    8188.54  1.9928 0.0462863
## Wealth        400.52     107.90  3.7121 0.0002056
## Clergy       -279.20     109.10 -2.5591 0.0104940
##
## Rho: 0.082595, LR test value: 15.313, p-value: 9.1111e-05
## Approximate (numerical Hessian) standard error: 0.019307
##     z-value: 4.2781, p-value: 1.8852e-05
## Wald statistic: 18.302, p-value: 1.8852e-05
##
## Log likelihood: -980.4748 for lag model
## ML residual variance (sigma squared): 588440000, (sigma: 24258)
## Number of observations: 85
## Number of parameters estimated: 5
## AIC: 1970.9, (AIC for lm: 1984.3)
```

Again, the coefficients for Wealth and Clergy are not significant.The plot of the residuals:



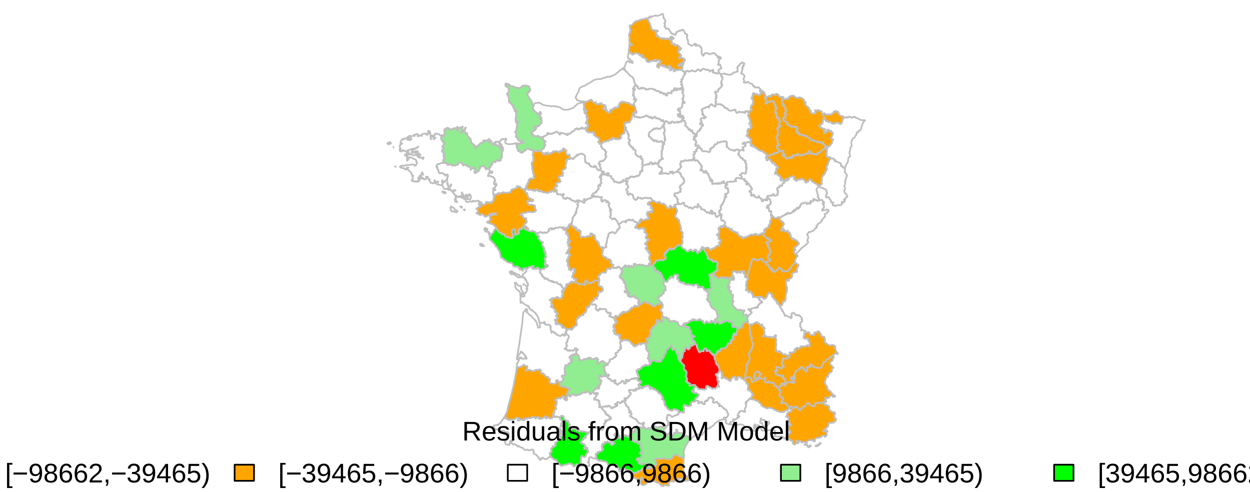
SDM

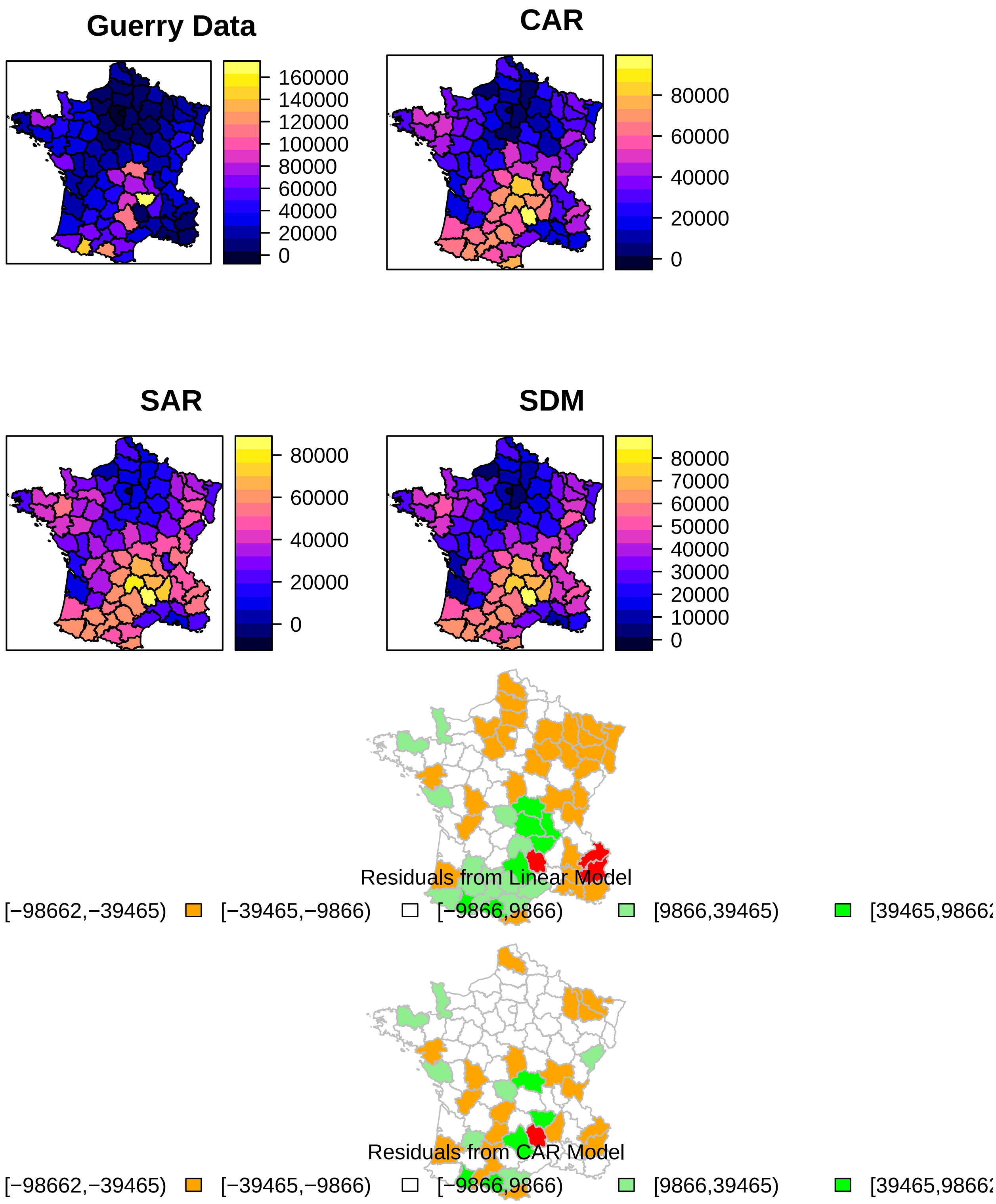
And SDM

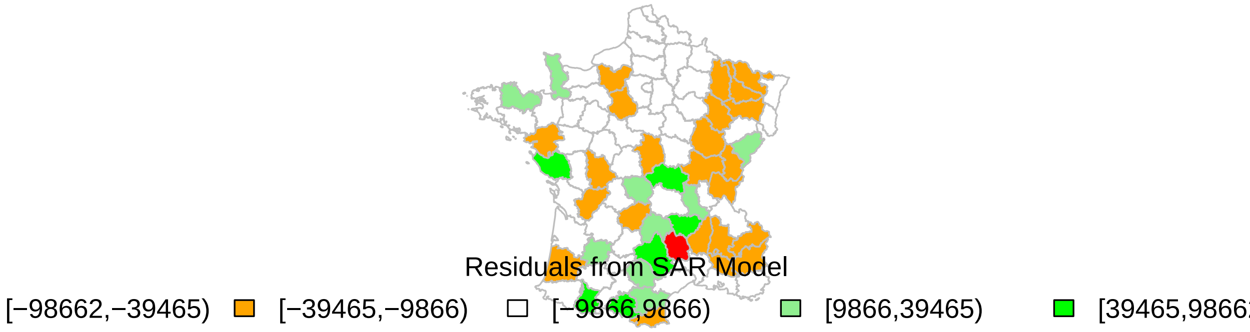
```
mod.sdm <- lagsarlm(Suicids ~ Wealth + Clergy, data=guerry, listw=W_cont_el_mat,
                    zero.policy=T, type="mixed", tol.solve=1e-12)
summary(mod.sdm)
```

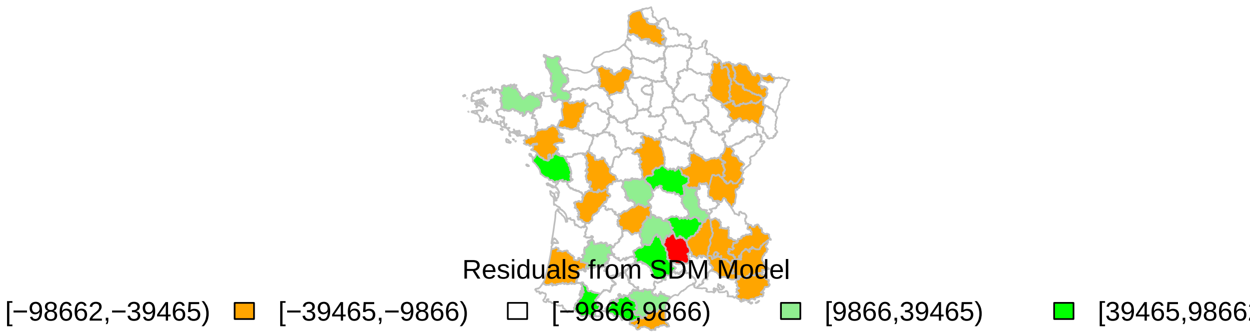
```
##
## Call:lagsarlm(formula = Suicids ~ Wealth + Clergy, data = guerry,
##     listw = W_cont_el_mat, type = "mixed", zero.policy = T, tol.solve = 1e-12)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -72844.2 -10442.8  -5159.5   4116.9  94379.8
##
## Type: mixed
## Coefficients: (numerical Hessian approximate standard errors)
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     20408.816  12688.840  1.6084 0.107746
## Wealth            417.431    132.596  3.1481 0.001643
## Clergy           -211.493    140.755 -1.5026 0.132954
## lag.(Intercept)  -158.197   3430.786 -0.0461 0.963222
## lag.Wealth        -21.019     47.436 -0.4431 0.657698
## lag.Clergy        -21.646     45.160 -0.4793 0.631706
##
## Rho: 0.094743, LR test value: 12.159, p-value: 0.00048859
## Approximate (numerical Hessian) standard error: 0.023939
##     z-value: 3.9577, p-value: 7.5666e-05
## Wald statistic: 15.664, p-value: 7.5666e-05
##
## Log likelihood: -979.7269 for mixed model
## ML residual variance (sigma squared): 570060000, (sigma: 23876)
## Number of observations: 85
## Number of parameters estimated: 8
## AIC: 1975.5, (AIC for lm: 1985.6)
```




Guerry Data
CAR
SAR
SDM
Residuals from Linear Model
[−98662,−39465)
[−39465,−9866)
[−9866,9866)
[9866,39465)
[39465,9866
Residuals from CAR Model
[−98662,−39465)
[−39465,−9866)
[−9866,9866)
[9866,39465)
[39465,9866


Residuals from SAR Model
[−98662,−39465)
[−39465,−9866)
[−9866,9866)
[9866,39465)
[39465,9866


Residuals from SDM Model
[−98662,−39465)
[−39465,−9866)
[−9866,9866)
[9866,39465)
[39465,9866