

```
In [1]: ##--Regressão Espacial no R

#Bibliotecas Necessárias
library(sf) ## leitura da geometria da região
library(geobr) ## leitura da geometria dos municípios do Brasil
library(spdep) ## Estrutura de vizinhança espacial
library(spatialreg) ## análise espacial com o SAR, CAR e SMA
library(spData) ## banco de dados
library(ggplot2) ## elaboração de gráficos
library(tmap) ## elaboração de mapas
library(leaflet) ## mapas interativos
library(cartogram) # mapas distorcidos
library(viridis) ## paleta de cores
library(RColorBrewer) ## escala de cores
```

Linking to GEOS 3.10.6, GDAL 3.6.2, PROJ 9.3.1; sf_use_s2() is TRUE

Carregando pacotes exigidos: spData

To access larger datasets in this package, install the spDataLarge package with: ``install.packages('spDataLarge',
repos='https://nowosad.github.io/drat/', type='source')``

Carregando pacotes exigidos: Matrix

Attaching package: 'spatialreg'

The following objects are masked from 'package:spdep':

```
get.ClusterOption, get.coresOption, get.mcOption,  
get.VerboseOption, get.ZeroPolicyOption, set.ClusterOption,  
set.coresOption, set.mcOption, set.VerboseOption,  
set.ZeroPolicyOption
```

Breaking News: tmap 3.x is retiring. Please test v4, e.g. with `remotes::install_github('r-tmap/tmap')`

Carregando pacotes exigidos: viridisLite

```
In [2]: ##---PARTE I
##--Iniciando com a Regressão Linear simples para entender o processo

##--unidades vendidas
y = c(430, 335, 520, 490, 470, 210, 195, 270, 400, 480)

##--Total de anúncios
x = c(30, 21, 38, 42, 37, 20, 8, 17, 35, 25)

##--Iniciando as etapas da Regressão
##--diagrama de dispersão
plot(x, y, xlab= "Total de Anúncios", ylab = "Unidades Vendidas")
##- coeficiente de correlação
cor(x, y)
##--Resultados do modelo no R
```

```
# usar a função lm() do R base.
modelo_linear <- lm(y ~ x)
summary(modelo_linear)
```

0.872143063644016

Call:

```
lm(formula = y ~ x)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-99.775	-26.288	-1.325	21.921	122.126

Coefficients:

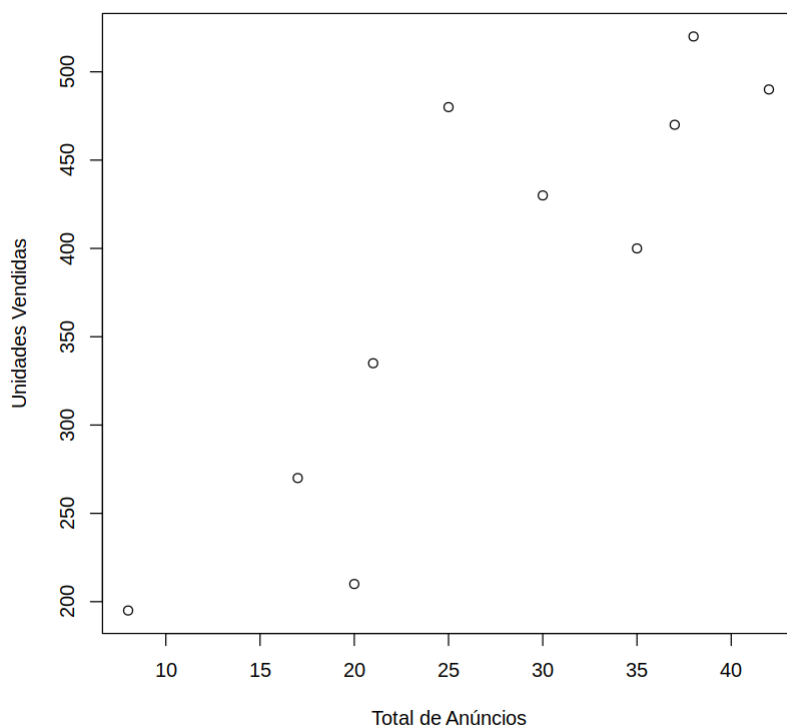
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	117.377	55.695	2.108	0.068137 .
x	9.620	1.908	5.042	0.000999 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 62.36 on 8 degrees of freedom

Multiple R-squared: 0.7606, Adjusted R-squared: 0.7307

F-statistic: 25.42 on 1 and 8 DF, p-value: 0.0009992



In [4]:

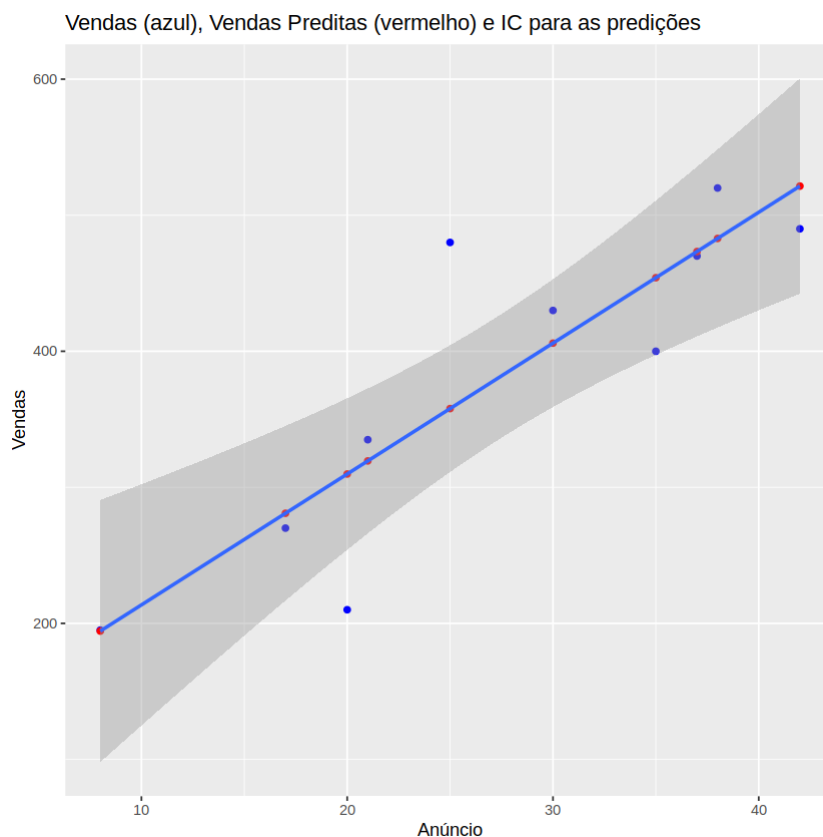
```
# Predição no Modelo
amostra_teste <- data.frame(x=x)
y_predito <- predict(modelo_linear, amostra_teste)
y_predito

##--construindo um data.frame dos dados
ml <- data.frame(x=x, y=y, y_predito = y_predito)

##--plotando o grafico de vendas e preditos
ggplot(data=ml, aes(x=x)) +
  geom_point(aes(y=y), colour = "blue") +
  geom_point(aes(y=y_predito), colour = "red") +
```

```
geom_smooth(aes(y=y), method = "lm", formula = y ~ x, se=TRUE) +
labs(title = "Vendas (azul), Vendas Preditas (vermelho) e IC para as pr
```

1: 405.97369160191 2: 319.394719595544 3: 482.932777829791 4: 521.412320943732 5:
473.312892051306 6: 309.774833817058 7: 194.336204475236 8: 280.915176481603 9:
454.073120494336 10: 357.874262709484



```
In [5]: ##---PARTE II ---
##--Regressão Espacial: Modelos SAR, CAR e GWR explicação nos slides!
# dados geospacial: municípios do AM via IBGE
am_muni <- read_municipality(code_muni = "AM", year= 2019, showProgress =
lviz <- poly2nb(am_muni)
Mviz <- nb2listw(lviz, style="W")

##---simulando um processo SAR no am ---##
set.seed(987654)
n <- dim(am_muni)[1]
uncorr_x <- rnorm(n)
#lambda=10
#uncorr_x <- rpois(n, lambda)
rho <- 0.90
autocorr_x <- invIrW(Mviz, rho) %*% uncorr_x
###--fim SAR SIMULADO--##
```

Using year 2019

```
In [6]: ## Moran teste
holmes<-moran.test(autocorr_x, Mviz)
holmes
holmes1<-moran.test(uncorr_x, Mviz)
holmes1
```

Moran I test under randomisation

data: autocorr_x
weights: Mvz

Moran I statistic standard deviate = 9.5332, p-value < 2.2e-16

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.74778154	-0.01639344	0.00642546

Moran I test under randomisation

data: uncorr_x
weights: Mvz

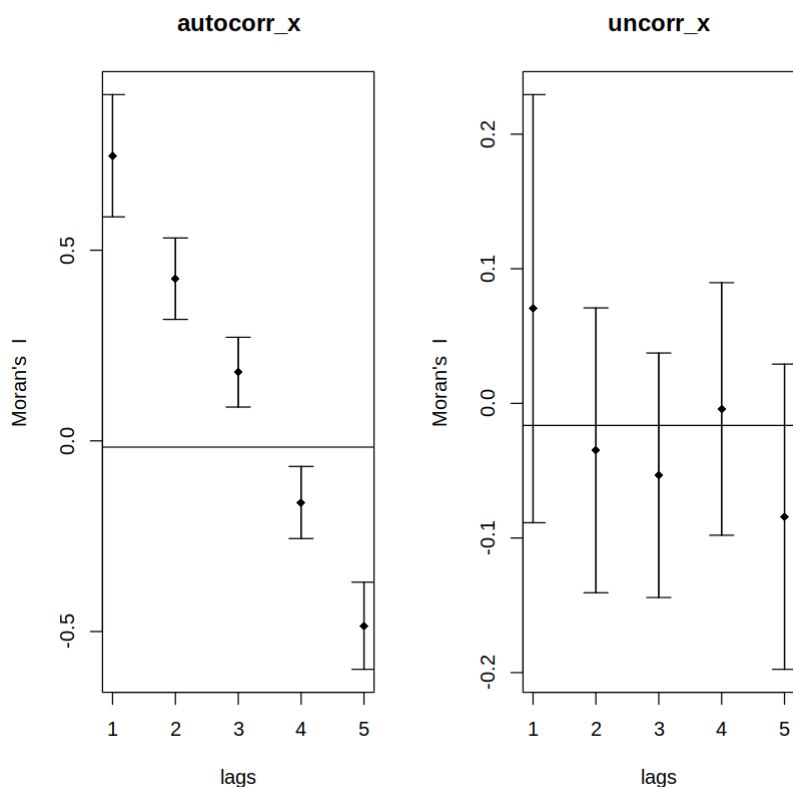
Moran I statistic standard deviate = 1.0907, p-value = 0.1377

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.070326806	-0.016393443	0.006322121

```
In [7]: ## Análise do correlograma
autocorr_x=as.vector(autocorr_x)
uncorr_x = as.vector(uncorr_x)
par(mfrow=c(1,2))
csp <- sp.correlogram(lviz, autocorr_x, order=5, method="I")
plot(csp)
csp1 <- sp.correlogram(lviz, uncorr_x, order=5, method="I")
plot(csp1)
```



```
In [8]: ##--PARTE III: APLICAÇÃO no R--##
## Dados
#contagens de casos de leucemia para 281 setores censitários
#de oito condados centrais do Estado de Nova York.
# Z = log(1000*(Y + 1)/n_i)
```

```
# n_i = população em risco na área i
# Z é apenas uma aproximação nos dados

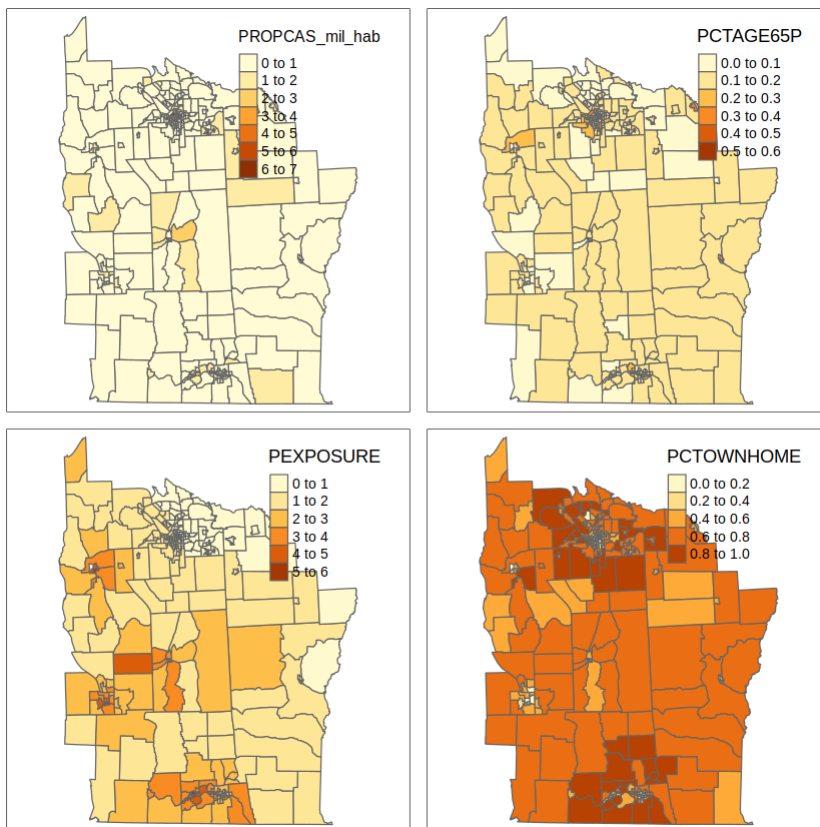
##Covariáveis:
#inverso da distância ao local de Tricloroeteno (TCE) mais próximo-Fator
#proporção de pessoas com 65 anos ou mais (PCTAGE65P);
#proporção de pessoas que possuem casa própria (PCTOWNHOME).

##Leitura dos Dados: os dados estão no pacote spData##
##NY8 <- as(sf::st_read(system.file("shapes/NY8_utm18.shp", package="spDa
NY8 <- sf::st_read(system.file("shapes/NY8_utm18.shp", package="spData"))
```

```
Reading layer `NY8_utm18' from data source
`/home/rocha/miniconda3/envs/regressao_especial_r/lib/R/library/spData/s
hapes/NY8_utm18.shp'
using driver `ESRI Shapefile'
Simple feature collection with 281 features and 17 fields
Geometry type: POLYGON
Dimension: XY
Bounding box: xmin: 358241.9 ymin: 4649755 xmax: 480393.1 ymax: 4808545
Projected CRS: WGS 84 / UTM zone 18N
```

In [9]: `##-Ajuste de Modelo de Regressão Espacial:--#`

```
# o que observamos ? Inteligência Espacial!
NY8$PROPCAS_mil_hab <- (NY8$PROPCAS)*1000
m1 <- qtm(NY8,"PROPCAS_mil_hab")
m2 <- qtm(NY8,"PCTAGE65P")
m3 <- qtm(NY8,"PEXPOSURE")
m4 <- qtm(NY8,"PCTOWNHOME")
tmap_arrange(m1, m2, m3, m4)
```



In [10]: `#Passo 1: Estimar o ml (modelo linear para dados independentes)`
`nylm <- lm(Z ~ PEXPOSURE + PCTOWNHOME, data = NY8)`
`summary(nylm)`

Call:

```
lm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.6726	-0.4413	-0.0400	0.4106	4.9133

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.08983	0.13765	0.653	0.5146
PEXPOSURE	0.06602	0.03749	1.761	0.0793 .
PCTOWNHOME	-0.72319	0.18062	-4.004	8e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7046 on 278 degrees of freedom

Multiple R-squared: 0.06915, Adjusted R-squared: 0.06245

F-statistic: 10.33 on 2 and 278 DF, p-value: 4.726e-05

```
In [11]: #Passo 2: Analisar os Resíduos - Moran Test
NY_nb = poly2nb(NY8)
Mviz <- nb2listw(NY_nb, style="W") ## define a matriz de vizinhança B, W,
lm.morantest(nylm, Mviz) ## moran test para os resíduos
```

Global Moran I for regression residuals

data:

model: lm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8)

weights: Mviz

Moran I statistic standard deviate = 3.6953, p-value = 0.0001098

alternative hypothesis: greater

sample estimates:

Observed Moran I	Expectation	Variance
0.124200902	-0.008853281	0.001296474

```
In [12]: #--parece haver dependência espacial?--#
NY8resid <- residuals(nylm)
csp <- sp.correlogram(NY_nb, NY8resid, order=4, method="I")
plot(csp)
#--Indicação de primeira ordem--#

#definindo a estrutura de vizinhança
NY_nb = poly2nb(NY8)
Mviz <- nb2listw(NY_nb, style="W") ## define a matriz de vizinhança W,...

#--Ajustando um Modelo SAR--#

NY8sar = spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME , data = NY8,
listw = Mviz, family = "SAR")
summary(NY8sar)

#--Ajustando um Modelo CAR--#
NY8car = spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME , data = NY8,
listw = Mviz, family = "CAR")
summary(NY8car)
```

```
Call: spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8, listw = M
viz,
  family = "SAR")
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.682146	-0.429646	-0.052537	0.370808	4.737739

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.057788	0.167590	-0.3448	0.73023
PEXPOSURE	0.082409	0.049102	1.6783	0.09328
PCTOWNHOME	-0.517641	0.208581	-2.4817	0.01307

Lambda: 0.30206 LR test value: 10.96 p-value: 0.00093093

Numerical Hessian standard error of lambda: 0.087457

Log likelihood: -293.337

ML residual variance (sigma squared): 0.46428, (sigma: 0.68138)

Number of observations: 281

Number of parameters estimated: 5

AIC: NA (not available for weighted model)

Warning message in spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8, listw = Mviz, :

"Non-symmetric spatial weights in CAR model"

```
Call: spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8, listw = M
viz,
  family = "CAR")
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.637778	-0.417563	-0.023747	0.339230	4.532105

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.093867	0.175274	-0.5355	0.59227
PEXPOSURE	0.102957	0.052418	1.9642	0.04951
PCTOWNHOME	-0.488608	0.214408	-2.2789	0.02267

Lambda: 0.60779 LR test value: 12.242 p-value: 0.00046728

Numerical Hessian standard error of lambda: 0.1465

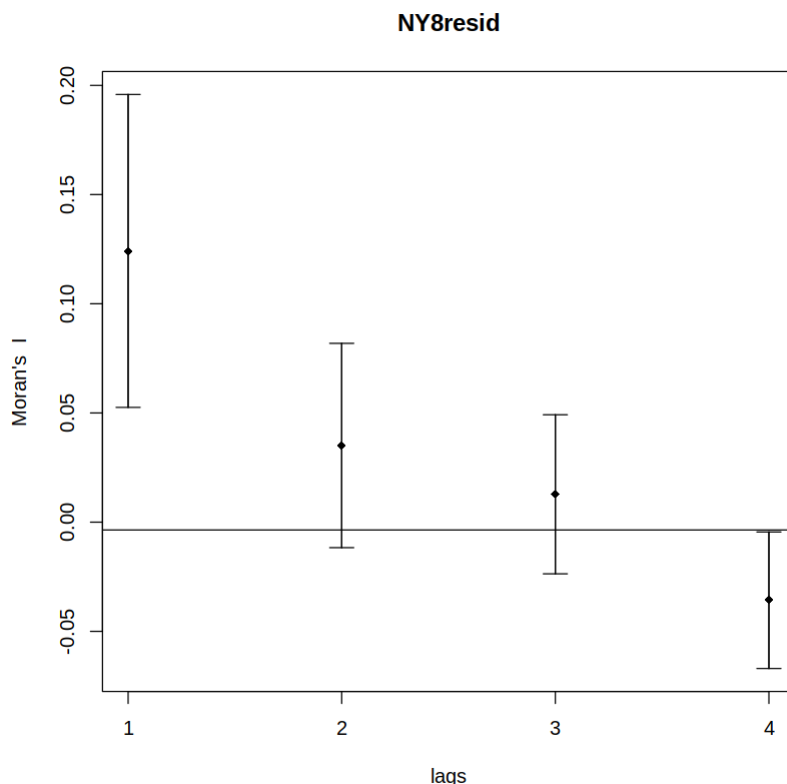
Log likelihood: -292.6962

ML residual variance (sigma squared): 0.45161, (sigma: 0.67202)

Number of observations: 281

Number of parameters estimated: 5

AIC: NA (not available for weighted model)



```
In [13]: ##- Comparando os Modelos com o AIC#
## 0 que nos diz o valor de Lambda e o AIC dos modelos
c(AIC(nylm),AIC(NY8sar),AIC(NY8car))
## quem tem o menor AIC?
##AICcar=-2*(summary(NY8car)$LL) + 2*(summary(NY8car)$parameters)

##Mapas para valores ajustados, observado, resíduo e risco
NY8$observado <-NY8$Z
NY8$ajustado_car = NY8car$fit$fitted.values
resíduo_car = NY8car$fit$residuals

# o que observamos ? Inteligência Espacial!
observado = qtm(NY8,"observado")
ajustado = qtm(NY8, "ajustado_car")
tmap_arrange(observado, ajustado)
```

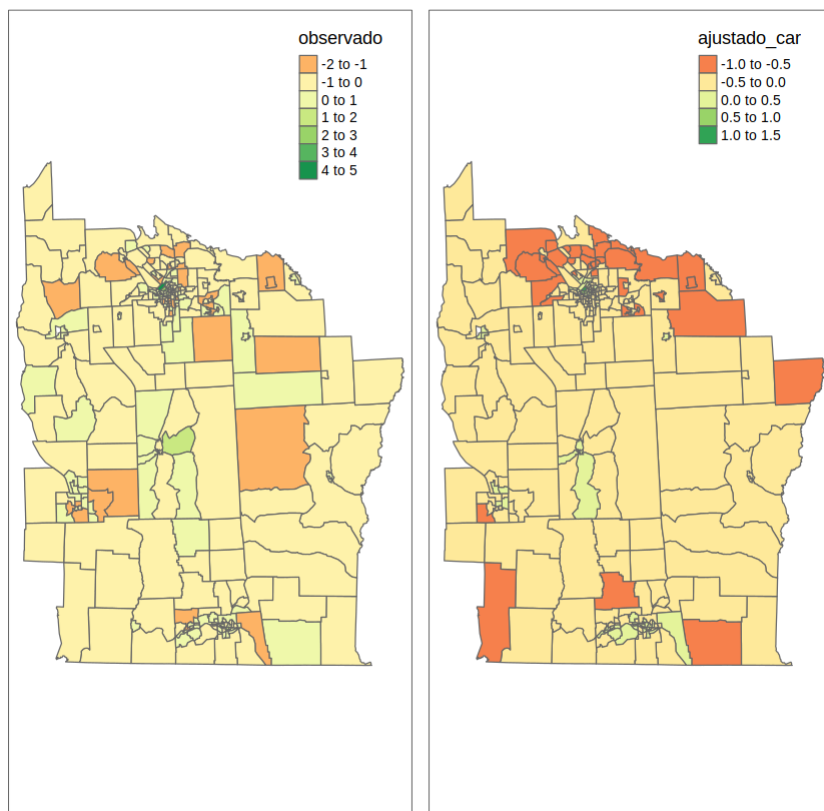
605.634227541287 · 596.674082413497 · 595.392336830635

Variable(s) "observado" contains positive and negative values, so midpoint is set to 0. Set midpoint = NA to show the full spectrum of the color palette.

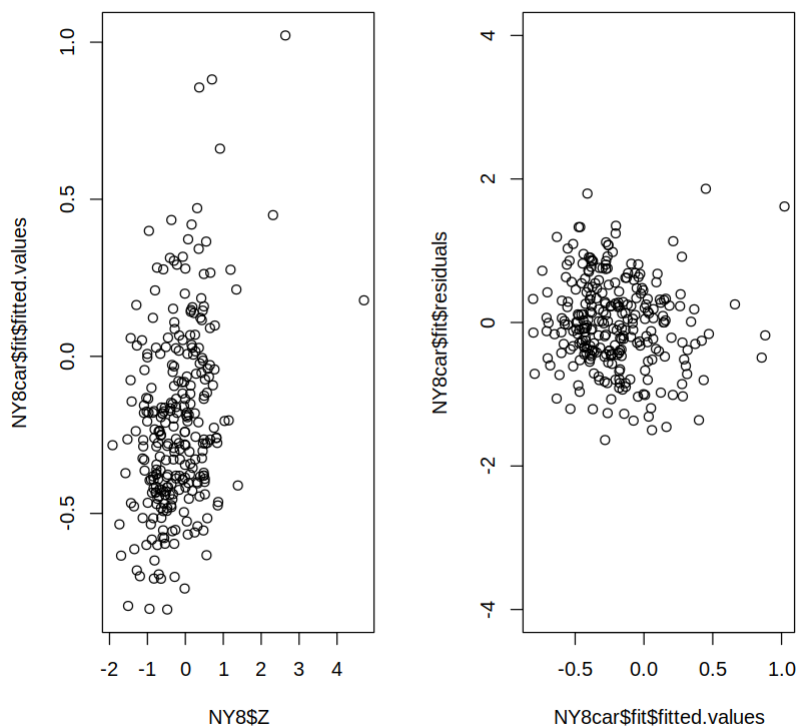
Variable(s) "ajustado_car" contains positive and negative values, so midpoint is set to 0. Set midpoint = NA to show the full spectrum of the color palette.

Variable(s) "observado" contains positive and negative values, so midpoint is set to 0. Set midpoint = NA to show the full spectrum of the color palette.

Variable(s) "ajustado_car" contains positive and negative values, so midpoint is set to 0. Set midpoint = NA to show the full spectrum of the color palette.



```
In [14]: #checagem de ajuste do modelo: gráfico do observado pelo ajustado e resíd
par(mfrow=c(1,2))
plot(NY8$Z, NY8car$fit$fitted.values)
plot(NY8car$fit$fitted.values, NY8car$fit$residuals, ylim=c(-4,4))
```



```
In [23]: ##Ainda possível avaliar outras possibilidades com vários
## tipos de matrizes de vizinhança.

# Certifique-se que NY_nb, NY8, e nylm estão definidos em células anterior
```

```

MvizB <- nb2listw(NY_nb, style="B") ## define a matriz de vizinhança tipo

# Ajustando um Modelo SAR com Matriz de Vizinhança do tipo B
NY8sarB <- spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8,
                    listw = MvizB, family = "SAR")

summary(NY8sarB)

# Ajustando um Modelo CAR com Matriz de Vizinhança do tipo B
NY8carB <- spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8,
                    listw = MvizB, family = "CAR")

summary(NY8carB)

#-- Comparando e Selecionando o Modelo mais "plausível" com o AIC#
## O que nos diz o valor de Lambda e o AIC dos modelos
aic_vals <- c(AIC(nylm), AIC(NY8sarB), AIC(NY8carB))
names(aic_vals) <- c("Linear", "SAR_B", "CAR_B")
aic_vals

```

Call:

```
spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8, listw = MvizB,
         family = "SAR")
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.65554	-0.43580	-0.06946	0.37643	4.66134

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.123412	0.170588	-0.7235	0.46940
PEXPOSURE	0.098484	0.048956	2.0117	0.04425
PCTOWNHOME	-0.472749	0.213432	-2.2150	0.02676

Lambda: 0.054719 LR test value: 11.809 p-value: 0.00058948

Numerical Hessian standard error of lambda: 0.01516

Log likelihood: -292.9127

ML residual variance (sigma squared): 0.46196, (sigma: 0.67968)

Number of observations: 281

Number of parameters estimated: 5

AIC: NA (not available for weighted model)

Call:

```
spautolm(formula = Z ~ PEXPOSURE + PCTOWNHOME, data = NY8, listw = MvizB,
  family = "CAR")
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.683686	-0.410274	-0.025502	0.368597	4.357031

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.206348	0.181321	-1.1380	0.25511
PEXPOSURE	0.113423	0.052712	2.1518	0.03142
PCTOWNHOME	-0.387549	0.221426	-1.7502	0.08008

Lambda: 0.10972 LR test value: 13.474 p-value: 0.00024187

Numerical Hessian standard error of lambda: 0.022765

Log likelihood: -292.08

ML residual variance (sigma squared): 0.44683, (sigma: 0.66846)

Number of observations: 281

Number of parameters estimated: 5

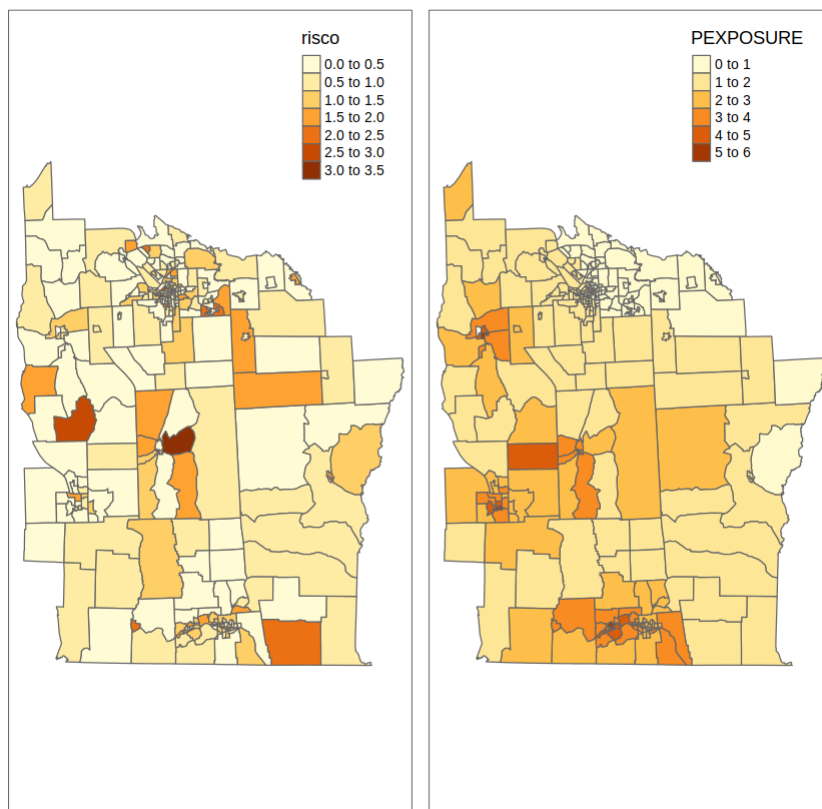
AIC: NA (not available for weighted model)

Linear: 605.634227541287 **SAR_B:** 595.825318012671 **CAR_B:** 594.160045934059

```
In [17]: ##--inteligência espacial e insights práticos!
## veja, no modelo CAR (modelo selecionado), que a variável preditora PCT
## Isso nos dá evidências que PCTOWNHOME é possivelmente um fator
## de confusão (confundimento - é uma variável que influencia tanto
# a variável dependente, quanto a variável independente,
# causando uma associação espúria.)

##--Identificando áreas com maior risco: mapa de predição do risco relati
##-- risco > 1, significa número de casos acima do esperado
##-- risco = 1, significa número de casos dentro do esperado
##-- risco < 1, número de casos abaixo do esperado (processo sob controle
ni = NY8$POP8 ##população em risco no site i
zi_predito = NY8car$fit$fitted.values ##valor predito para
casos_predito <- (ni/1000)*exp(zi_predito) ## casos preditos para o site
risco = NY8$Cases/casos_predito
NY8$risco = risco

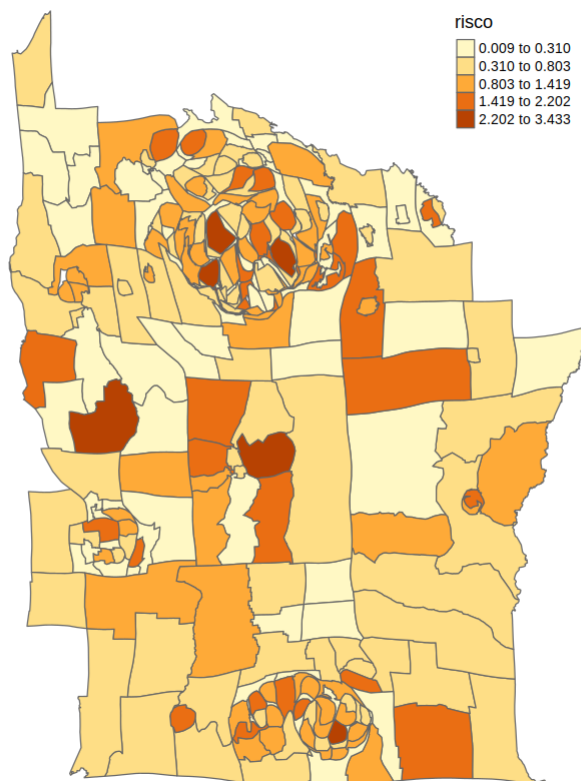
map1 <- qtm(NY8,"risco")
map2 <- qtm(NY8,"PEXPOSURE")
tmap_arrange(map1, map2)
## Através regressão espacial foi possível estimar adequadamente
## as áreas com maiores riscos de casos
```



```
In [18]: ##--PARTE III - Visualização de Dados: novo tipo de mapa--##
##--Cartograma
##--pacotes necessários--##
library(cartogram)
library(sf)
library(tmap)
```

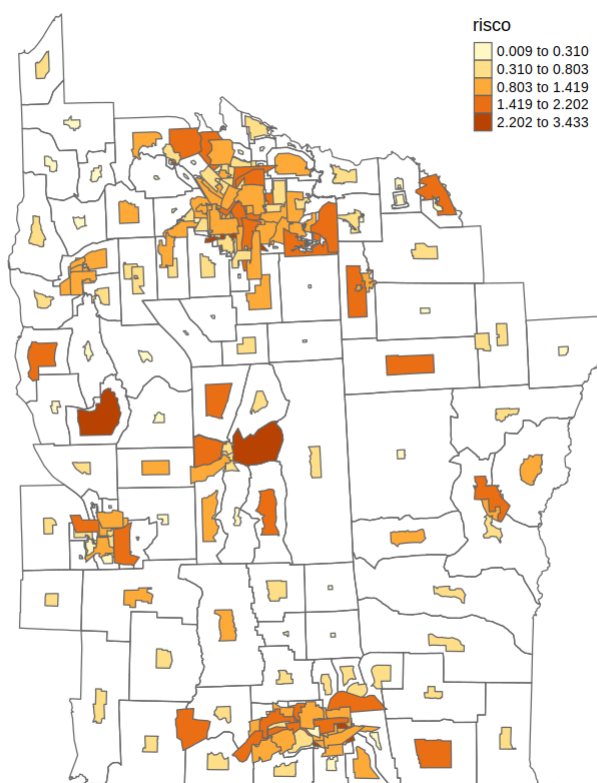
```
In [19]: ###--Cartograma de Área Contínua--##
##--preparando os dados--#
NY8_sf = st_as_sf(NY8, crs = 3395)

##--construindo o cartograma
ny8_cont <- cartogram_cont(NY8_sf, "risco", itermax = 5)
# plot
tm_shape(ny8_cont) + tm_polygons("risco", style = "jenks") +
tm_layout(frame = FALSE, legend.position = c("right", "top"))
```



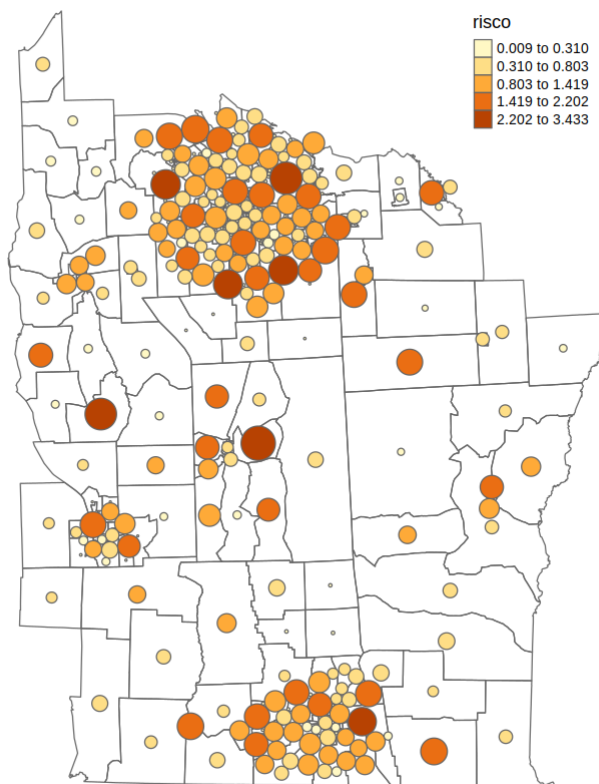
```
In [20]: ##-Cartograma de Área Não Contígua--##
# construindo cartograma
ny8_cont <- cartogram_ncont(NY8_sf, "risco")

# plot
tm_shape(NY8_sf) + tm_borders() +
tm_shape(ny8_cont) + tm_polygons("risco", style = "jenks") +
tm_layout(frame = FALSE, legend.position = c("right", "top"))
```



```
In [21]: ##--Cartograma de círculos não sobrepostos
# construindo o cartograma
ny8_dorling <- cartogram_dorling(NY8_sf, "risco", k=0.25)

# plot
tm_shape(NY8_sf) + tm_borders() +
tm_shape(ny8_dorling) + tm_polygons("risco", style = "jenks") +
tm_layout(frame = FALSE, legend.position = c("right", "top"))
```



```
In [22]: ##--PARTE IV - Detectando Cluster com Regressão Espacial de taxas--##

##--ler os dados de mortalidade infantil no amazonas
mortalidade_infantil <- read.delim("C:\\Users\\maxso\\Desktop\\max\\spati
#mortalidade_infantil <- read.delim(file=" endereço no seu computador\\da
am_muni <- read_municipality(code_muni = "AM", year= 2019, showProgress =
lviz <- poly2nb(am_muni)
Mviz <- nb2listw(lviz, style="W")
tosca <- data.frame(cod=am_muni$code_muni,mortalidade_infantil)
am_mortalidade_infantil = dplyr::left_join(am_muni, toasca, by = c("code_m
mor_inf_lm <- lm(InfantMor ~ IDHM + IAMM, data = am_mortalidade_infantil)
summary(mor_inf_lm)
lm.morantest(mor_inf_lm, Mviz) ## moram test para os resíduos

resid_ml <- residuals(mor_inf_lm)
csp <- sp.correlogram(lviz, resid_ml, order=5, method="I")
plot(csp)
```

Warning message in file(file, "rt"):

"não foi possível abrir o arquivo 'C:\\Users\\maxso\\Desktop\\max\\spatialdata\\Plano_de_Ensino\\data_infant_mortality.txt': Arquivo ou diretório inexistente"

Error in file(file, "rt"): não é possível abrir a conexão
Traceback:

```
1. read.delim("C:\\Users\\maxso\\Desktop\\max\\spatialdata\\Plano_de_Ensin  
o\\data_infant_mortality.txt")  
2. read.table(file = file, header = header, sep = sep, quote = quote,  
  .      dec = dec, fill = fill, comment.char = comment.char, ...)  
3. file(file, "rt")
```

```
In [ ]: ##--Ajustando um Modelo CAR--#  
mor_inf_car = spautolm(InfantMor ~ IDHM + IAMM, data = am_mortalidade_inf  
summary(mor_inf_car)  
  
##--visualizando valores preditos para detecção de clusters espaciais  
am_mortalidade_infantil$predita = mor_inf_car$fit$fitted.values ##valor p  
tmap_mode("view")  
tm_shape(am_mortalidade_infantil) +  
tm_fill(col = "predita", title = "", id="name_muni") +  
tm_layout(frame = FALSE, title = "Taxa de Mortalidade Infantil no Amazona  
tm_layout(legend.outside = TRUE)
```