##--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

##---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 correlaçaõ

cor(x,y)

##--Resultados do modelo no R

# usar a função lm() do R base.

modelo\_linear <- lm(y ~ x)

summary(modelo\_linear)

# 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)

##--plotanto o grafico de vendas e preditos

ggplot(data=ml, aes(x=x, y=y, y\_predito=y\_predito)) +

geom\_point(aes(x=x, y=y), colour = "blue") +

geom\_point(aes(x=x, y=y\_predito), colour = "red") +

geom\_smooth(method = "lm", formula = y ~ x) +

labs(title = "Vendas (azul), Vendas Preditas (vermelho) e IC para as predições", y = "Vendas", x = "Anúncio")

##---PARTE II ---

##--Regressão Espacial: Modelos SAR, CAR e GWR explicação nos slides!

# dados geospatial: municípios do AM via IBGE

am\_muni <- read\_municipality(code\_muni = "AM", year= 2019, showProgress = FALSE)

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--##

## Moran teste

holmes<-moran.test(autocorr\_x, Mviz)

holmes

holmes1<-moran.test(uncorr\_x, Mviz)

holmes1

## 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)

##--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 de exposição (PEXPOSIÇÃO);

#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="spData")), "Spatial")

NY8 <- sf::st\_read(system.file("shapes/NY8\_utm18.shp", package="spData"))

#--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)

#Passo 1: Estimar o ml (modelo linear para dados independentes)

nylm <- lm(Z ~ PEXPOSURE + PCTOWNHOME, data = NY8)

summary(nylm)

#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) ## moram test para os resíduos

#--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)

#-- Comparando os Modelos com o AIC#

## O 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)

#checagem de ajuste do modelo: gráfico do observado pelo ajustado e resíduos para o modelo escolhido

par(mfrow=c(1,2))

plot(NY8$Z,NY8car$fit$fitted.values)

plot(NY8car$fit$fitted.values, NY8car$fit$residuals, ylim=c(-4,4))

##Ainda possível avaliar outras possibilidades com vários

## tipos de matrizes de vizinhança.

MvizB <- nb2listw(NY\_nb, style="B") ## define a matriz de vizinhança B, W,...

# 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

c(AIC(nylm),AIC(NY8sarB),AIC(NY8carB))

##--inteligência espacial e insights práticos!

## veja, no modelo CAR (modelo selecionado), que a variável preditora PCTOWNHOME agora não é mais significativa.

## 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 relativo de casos--#

##-- 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 i

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

##--PARTE III - Visualização de Dados: novo tipo de mapa--##

##--Cartograma

##--pacotes necessários--##

library(cartogram)

library(sf)

library(tmap)

##--fim pacotes necessários--##

###--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"))

#--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"))

#--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"))

##--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\\spatialdata\\Plano\_de\_Ensino\\data\_infant\_mortality.txt")

#mortalidade\_infantil <- read.delim(file=" endereço no seu computador\\data\_infant\_mortality.txt")

am\_muni <- read\_municipality(code\_muni = "AM", year= 2019, showProgress = FALSE)

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, tosca, by = c("code\_muni" = "cod"))

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)

#--Ajustando um Modelo CAR--#

mor\_inf\_car = spautolm(InfantMor ~ IDHM + IAMM, data = am\_mortalidade\_infantil, listw = Mviz, family ="CAR")

summary(mor\_inf\_car)

##--visualizando valores preditos para deteção de clusters espaciais

am\_mortalidade\_infantil$predita = mor\_inf\_car$fit$fitted.values ##valor predito para mor\_inf\_car

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 Amazonas") +

tm\_layout(legend.outside = TRUE)