Geo-épidémiologie – Statistiques spatiales et analyses écologiques

Jean Gaudart, Jordi Landier, Cédric Bationo, Mady Cissoko, Eva Legendre

SESSTIM, *Institute of Public Health Sciences*, Aix Marseille Univ, France
http://sesstim.univ-amu.fr/

1. Définitions

- Statistiques Spatiales

1ere loi de la géographie (Waldo Tobler) :

« Tout interagit avec tout, mais deux objets proches ont plus de chances de le faire que deux objets éloignés »

  - Informations sur un phénomène géographique

  - Cartographie != SIGS != Statistiques spatiales

1. Définitions

- Statistiques Spatiales
  - Analyse
    - « Structure » spatial du phénomène (auto-correlation, clusters spatiaux, tendances…)
    - Explication de la structure (facteurs explicatifs)

$\Rightarrow$ Auto-correlation spatiale $\Rightarrow$ sous-estimation de la variance $\Rightarrow$ Biais

  - Données individuelles / agrégées (études écologiques)

1. Définitions

- Etudes Ecologiques
  - Etude épidémiologique
  - Etude de la co-variation de facteurs mesurés au niveau d'un groupe d'individu
  - Unité statistique = groupe
    - Par ex. ménage, ville, département, région, pays
  - Aucune information individuelle != Cohorte
  - ⇒ Biais écologiques
    - biais de pure spécification : pas d'information sur l'exposition individuelle ⇒ estimation biaisée de l'effet individuel
    - Variabilité intra-unité (mesure et facteurs)

Sylvia Richardson. « Problèmes méthodologiques dans les études écologiques santé-environnement » Comptes rendus de l'Académie des Sciences – series III – Sciences de la vie. 2000;323(7):611-6.

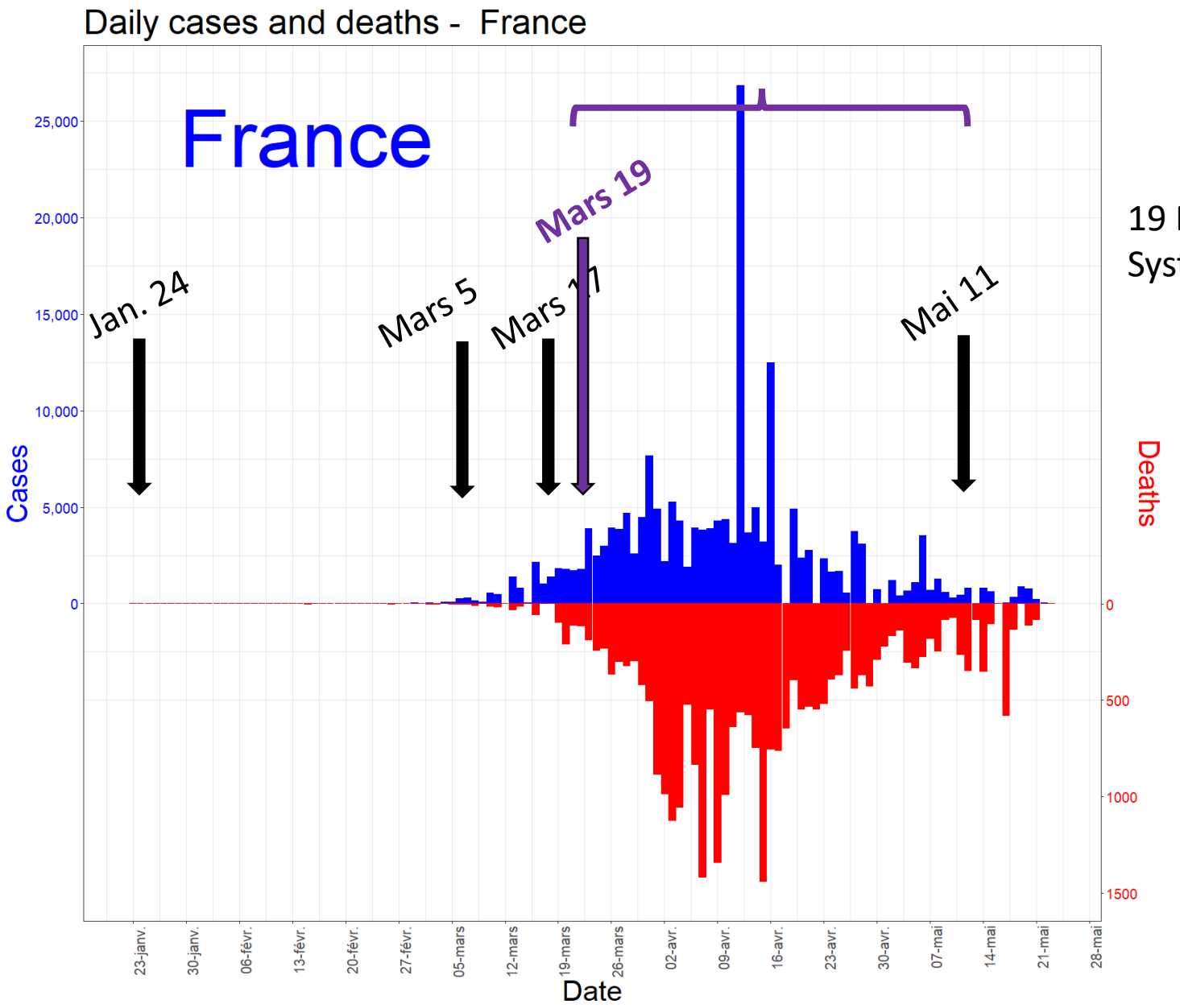
2. Quelle question de Recherche

- Hétérogénéité spatiale de la 1ere vague de COVID-19 en France

- Unité statistique : Départements métropolitains
- Outcome : Taux d'incidence hospitalier
  - Nombre cumulé de cas COVID-19 hospitalisés / 100 000 habitants

Gaudart J, Landier J, Huiart L, Legendre E, Lehot L, Bendiane MK, Chiche L, Petitjean A, Mosnier E, Kirakoya-Samadoulougou F, Demongeot J, Piarroux R, Rebaudet S. Lancet Public Health 2021, Feb 5

https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(21)00006-2/fulltext



19 Mars 2020 :
Système d'information Hosp. SIVIC

2. Quelle question de Recherche

- Co-facteurs :
  - Distribution Age et Sexe (40)
  - Délai entre 1er décès et Confinement (17 Mars) (1)
  - Accès aux soins (14)
  - Urbanisation (5)
  - Climat (9)
  - HCQ/CQ (1)

DataLPH.csv

```
pathToData <- "C:/.../DataLPH.csv"
DataSet <- read.csv2(file = pathToData,  header = TRUE)
DataSet$HIncid <- 100000*DataSet$HCases / DataSet$Pop2020
colnames(DataSet)
```

AK1 | HISTO_2.Pct

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	ID_GEOFLA	CODE_DEPT	NOM_DEPT	X_CENTROID	Y_CENTROID	CODE_REG	NOM_REG	Pop2020	MortMaAv1819	ICUBed	HCases	ICUCases	HDeath	HCQCQn	MedD	MedGD	InfD	PharmD	LitsD	VieilInd2019	Urbpct	ComAutrPct	ComMu
2	DEPARTEM0(	39	JURA	886172	6641548	27	BOURGOGNE	257849	931	8	349	59	55	81	242	145	1065	103	336,243305	105,1	55,6	20,7	
3	DEPARTEM0(	42	LOIRE	795655	6521581	84	AUVERGNE-F	764737	3060	78	1290	179	208	166	339	161	1160	99	437,405278	92,2	88,5	3,6	
4	DEPARTEM0(	76	SEINE-MARIT	550942	6952842	28	NORMANDIE	1243788	4825	103	1017	243	164	287	328	149	1064	108	431,182806	81,2	86,8	6,4	
5	DEPARTEM0(	89	YONNE	748211	6750855	27	BOURGOGNE	332096	1555	18	389	46	75	108	222	116	917	98	396,270958	108,4	59,5	16,1	
6	DEPARTEM0(	68	HAUT-RHIN	1016174	6763894	44	ALSACE-CHA	763204	2560	70	3707	550	774	297	303	139	1083	97	393,734834	86,3	88	8	
7	DEPARTEM0(	28	EURE-ET-LOI	579506	6810114	24	CENTRE-VAL	429425	1437	20	701	71	105	111	214	106	774	104	336,729347	84,2	78,4	11,3	
8	DEPARTEM0(	10	AUBE	787771	6802945	44	ALSACE-CHA	309907	1197	12	897	60	124	98	234	120	860	94	365,916227	89,8	65,8	13,5	
9	DEPARTEM0(	55	MEUSE	873653	6885669	44	ALSACE-CHA	181641	797	8	595	60	99	40	213	135	1065	92	416,756129	102,1	53,2	13,4	
10	DEPARTEM0(	61	ORNE	502675	6833540	28	NORMANDIE	276903	1208	16	280	25	35	45	242	135	1152	95	426,864281	120	33,8	28,7	
11	DEPARTEM0(	67	BAS-RHIN	1035762	6839265	44	ALSACE-CHA	1132607	3670	134	3265	587	579	391	405	174	1147	109	400,050503	80,8	87,2	3,5	
12	DEPARTEM0(	47	LOT-ET-GARO	502233	6367330	75	AQUITAINE-L	330336	1435	11	87	13	9	97	252	130	1048	111	385,667926	123,4	64,3	15,5	
13	DEPARTEM0(	40	LANDES	408685	6332024	75	AQUITAINE-L	411979	1525	14	90	16	13	108	276	160	925	99	283,509596	122,8	48,6	24,5	
14	DEPARTEM0(	14	CALVADOS	448366	6868155	28	NORMANDIE	691453	2917	99	434	100	70	201	367	168	1177	106	457,297893	93,3	88,7	6,3	
15	DEPARTEM0(	93	SEINE-SAINT	659903	6867852	11	ILE-DE-FRAN	1670149	2818	100	5153	697	908	640	258	115	659	70	248,600574	41,3	100	0	
16	DEPARTEM0(	62	PAS-DE-CALA	648085	7046488	32	NORD-PAS-D	1452778	5358	85	1756	293	265	259	255	136	854	103	378,516195	73,9	93,6	1,7	
17	DEPARTEM0(	15	CANTAL	676537	6438974	84	AUVERGNE-F	142811	799	8	66	17	10	50	290	168	1308	134	553,178677	152,1	45	13,5	
18	DEPARTEM0(	45	LOIRET	648620	6757240	24	CENTRE-VAL	682890	2276	51	813	190	83	173	244	110	841	102	380,149072	81,1	82,6	8,4	
19	DEPARTEM0(	63	PUY-DE-DOM	706634	6519140	84	AUVERGNE-F	660240	2570	81	232	72	39	207	363	158	1186	134	443,77802	98	81,6	7,6	
20	DEPARTEM0(	2	AISNE	743736	6938772	32	NORD-PAS-D	526050	2110	34	1324	119	233	137	230	110	1062	94	364,414029	83,1	69,8	13,4	
21	DEPARTEM0(	35	ILLE-ET-VILAI	326237	6845888	53	BRETAGNE	1082073	3450	66	530	93	80	168	344	158	902	101	383,615523	70,4	91,4	4,3	
22	DEPARTEM0(	71	SAONE-ET-LO	819998	6618053	27	BOURGOGNE	547824	2410	26	1049	112	177	112	262	129	1037	103	417,104764	120,6	62,5	17,7	
23	DEPARTEM0(	9	ARIEGE	570122	6206008	76	LANGUEDOC	152398	641	8	42	8	2	39	284	174	1140	104	288,7177	128,2	35,1	32,2	
24	DEPARTEM0(	78	YVELINES	614838	6851681	11	ILE-DE-FRAN	1448625	3088	84	2877	353	456	455	287	122	704	99	247,337993	63,3	100	0	
25	DEPARTEM0(	58	NIEVRE	740760	6666907	27	BOURGOGNE	199596	1098	12	125	26	25	41	233	122	1061	119	448,40578	155,4	46	16,5	
26	DEPARTEM0(	85	VENDEE	291784	6637100	52	PAYS DE LA L	683187	2513	18	261	39	34	123	219	124	755	90	265,227529	106,3	49,7	19,2	
27	DEPARTEM0(	64	PYRENEES-A	395765	6238257	75	AQUITAINE-L	683169	2828	37	258	44	29	228	386	189	1213	120	409,415533	113,6	77,3	11,3	
28	DEPARTEM0(	24	DORDOGNE	519054	6456311	75	AQUITAINE-L	408393	1882	19	169	14	11	93	238	135	985	101	300,690756	153,1	47	16,9	
29	DEPARTEM0(	69	RHONE	832095	6530600	84	AUVERGNE-F	1876051	5424	222	3994	753	579	546	448	175	1260	144	475,946549	67,1	97,7	1,9	
30	DEPARTEM0(	37	INDRE-ET-LO	524350	6682655	24	CENTRE-VAL	605380	2244	86	467	71	71	161	382	162	1022	122	443,523076	92,8	84,4	6,8	
31	DEPARTEM0(	92	HAUTS-DE-SE	644365	6856485	11	ILE-DE-FRAN	1613762	3715	210	6005	1076	959	779	399	157	868	143	439,160174	62,2	100	0	
32	DEPARTEM0(	88	VOSGES	952788	6792216	44	ALSACE-CHA	359520	1574	8	946	147	243	84	254	143	961	98	398,030708	114,3	55,9	26,2	
33	DEPARTEM0(	48	LOZERE	738959	6374438	76	LANGUEDOC	76286	334	3	20	2	1	18	237	141	1466	137	365,728967	123,5	0,1	49,3	
34	DEPARTEM0(	4	ALPES-DE-HA	953735	6346713	93	PROVENCE-A	165197	710	6	151	11	9	52	304	185	1103	115	309,327651	126,4	41,6	19,8	
35	DEPARTEM0(	8	ARDENNES	823553	6954788	44	ALSACE-CHA	265531	992	12	295	41	55	50	239	135	1089	97	354,384234	96,9	61,6	18,6	
36	DEPARTEM0(	81	TARN	629978	6299541	76	LANGUEDOC	387898	1662	24	147	44	21	98	299	151	1212	111	338,491047	114,9	56,9	23,4	
37	DEPARTEM0(	75	PARIS	650100	6862280	11	ILE-DE-FRAN	2148271	6371	471	8542	1918	1610	1622	858	248	1771	188	670,259944	89,9	100	0	
38	DEPARTEM0(	82	TARN-ET-GA	556115	6333159	76	LANGUEDOC	262618	948	18	51	19	3	91	264	135	1027	88	297,009344	91,7	77,2	7,9	
39	DEPARTEM0(	17	CHARENTE-M	360881	6576233	75	AQUITAINE-L	647080	2786	30	306	49	44	185	311	170	866	107	338,443469	135,9	68,8	10,8	
40	DEPARTEM0(	19	CORREZE	607137	6472059	75	AQUITAINE-L	240336	1230	15	198	33	30	53	279	149	1407	124	487,650623	135	60,7	12,7	
41	DEPARTEM0(	87	HAUTE-VIEN	565400	6537770	75	AQUITAINE-L	370774	1739	26	167	26	25	73	413	194	1501	149	528,084494	115,8	77,1	6,7	
42	DEPARTEM0(	83	VAR	983126	6221091	93	PROVENCE-A	1073836	3978	60	1010	159	125	508	337	166	1104	117	362,066461	125,1	87,9	8,2	
43	DEPARTEM0(	60	OISE	703330	6903185	32	NORD-PAS-D	825077	2257	36	1554	173	371	216	217	113	670	90	272,095817	65,7	91,5	3,5	
44	DEPARTEM0(	5	HAUTES-ALP	964813	6400089	93	PROVENCE-A	141756	542	8	141	25	12	51	441	246	1626	132	398,572195	113,7	45,7	21,8	
45	DEPARTEM0(	7	ARDECHE	808517	6399690	84	AUVERGNE-F	326875	1304	8	597	48	90	129	242	128	1137	103	324,282983	114,4	63	15,7	

DataLPH

Prêt | Moyenne : 2,061480882 | Nb (non vides) : 873 | Somme : 1781,119482

3. Plan d'analyse

- Cartographie :
  - datafile + shapefile ⇒ attention identifiants
  - Package « ggplot2 » ⇒ attention formatage préalage
- Auto-corrélation spatiale
- Co-facteurs : Réduction de dimension des co-facteurs
- Approche regressive / tenir compte du spatial
- Analyse de sensibilité (étude écologique)

4. Cartographie

- Cartographie : datafile + shapefile

EPICLIN21 › Data › DPTShp

Nom	Modifié le	Type	Taille
ShapeIdF	07/06/2021 17:14	Dossier de fichiers	
DEPARTEMENT.cpg	06/07/2016 14:28	Fichier CPG	1 Ko
DEPARTEMENT	06/07/2016 14:28	Classeur OpenOffi...	17 Ko
DEPARTEMENT.prj	06/07/2016 14:28	Fichier PRJ	1 Ko
DEPARTEMENT.shp	06/07/2016 14:28	Fichier SHP	3 052 Ko
DEPARTEMENT.shx	06/07/2016 14:28	Fichier SHX	1 Ko
LIMITE_DEPARTEMENT.cpg	06/07/2016 14:28	Fichier CPG	1 Ko
LIMITE_DEPARTEMENT	06/07/2016 14:28	Classeur OpenOffi...	1 992 Ko
LIMITE_DEPARTEMENT.prj	06/07/2016 14:28	Fichier PRJ	1 Ko
LIMITE_DEPARTEMENT.shp	06/07/2016 14:28	Fichier SHP	2 469 Ko
LIMITE_DEPARTEMENT.shx	06/07/2016 14:28	Fichier SHX	71 Ko
ShapeIdF	14/11/2020 16:54	Dossier compressé	1 821 Ko

```
pathToShp <- "C:/.../DPTShp"
DPT <- readOGR(dsn = pathToShp, layer = "DEPARTEMENT", stringsAsFactors = FALSE)
```

- Attention : les numéros de départements sont variables
  - Par ex. 2A et 2B

  ⇒ nécessité d'adapter tout "character" ou tout "factor"

4. Cartographie

```
DataSet$DptN2 <- as.character(DataSet$CODE_DEPT)
DataSet$DptN2[DataSet$DptN2 == 1] <- "01"
DataSet$DptN2[DataSet$DptN2 == 2] <- "02"
DataSet$DptN2[DataSet$DptN2 == 3] <- "03"
DataSet$DptN2[DataSet$DptN2 == 4] <- "04"
DataSet$DptN2[DataSet$DptN2 == 5] <- "05"
DataSet$DptN2[DataSet$DptN2 == 6] <- "06"
DataSet$DptN2[DataSet$DptN2 == 7] <- "07"
DataSet$DptN2[DataSet$DptN2 == 8] <- "08"
DataSet$DptN2[DataSet$DptN2 == 9] <- "09"

DataSet$DptN3 <- DataSet$CODE_DEPT
DataSet$DptN3[DataSet$DptN3 == "2B"] <- "202"
DataSet$DptN3[DataSet$DptN3 == "2A"] <- "201"
DataSet$DptN3 <- factor(DataSet$DptN3)
```

4. Cartographie

- Regrouper les 2 objets

```
DPT@data <- data.frame(DPT@data, DataSet[match(DPT@data[,"CODE_DEPT"], DataSet[,"DptN2"]),])
```

- « map » ↔ « dataframe » pour utiliser ggplot2

```
DptLine <- fortify(DPT, region = "CODE_DEPT")
DPT@data$id <- DPT@data$CODE_DEPT
DptLineD <- join(DptLine, DPT@data, by = "id")
```

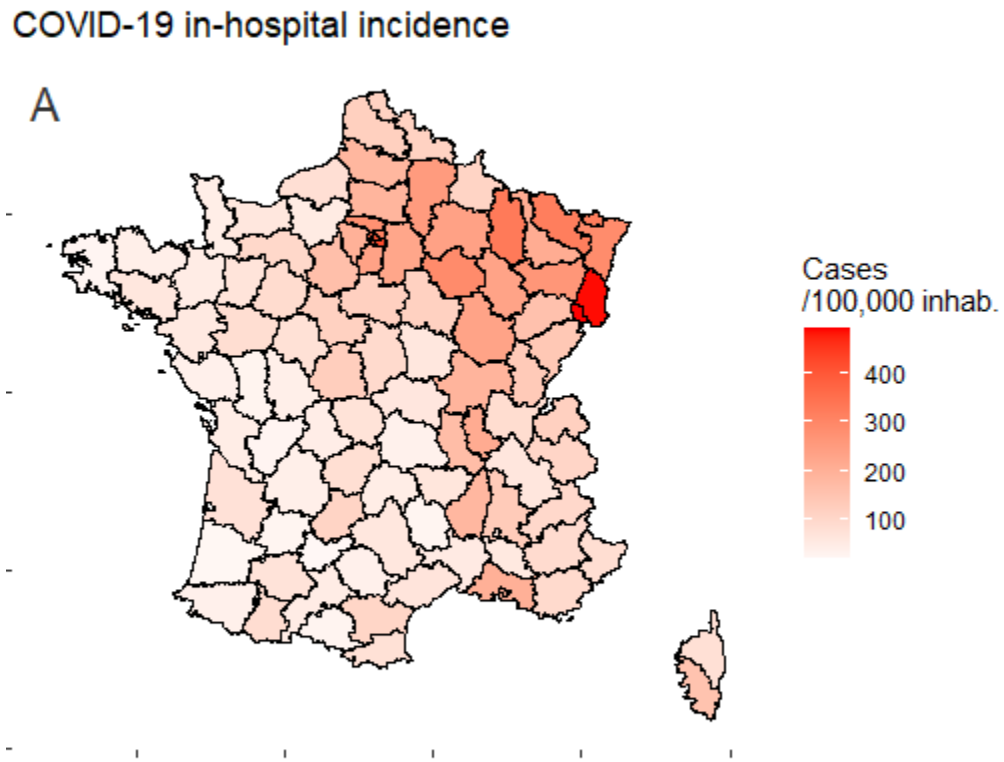
4. Cartographie

- HIncidMap <-

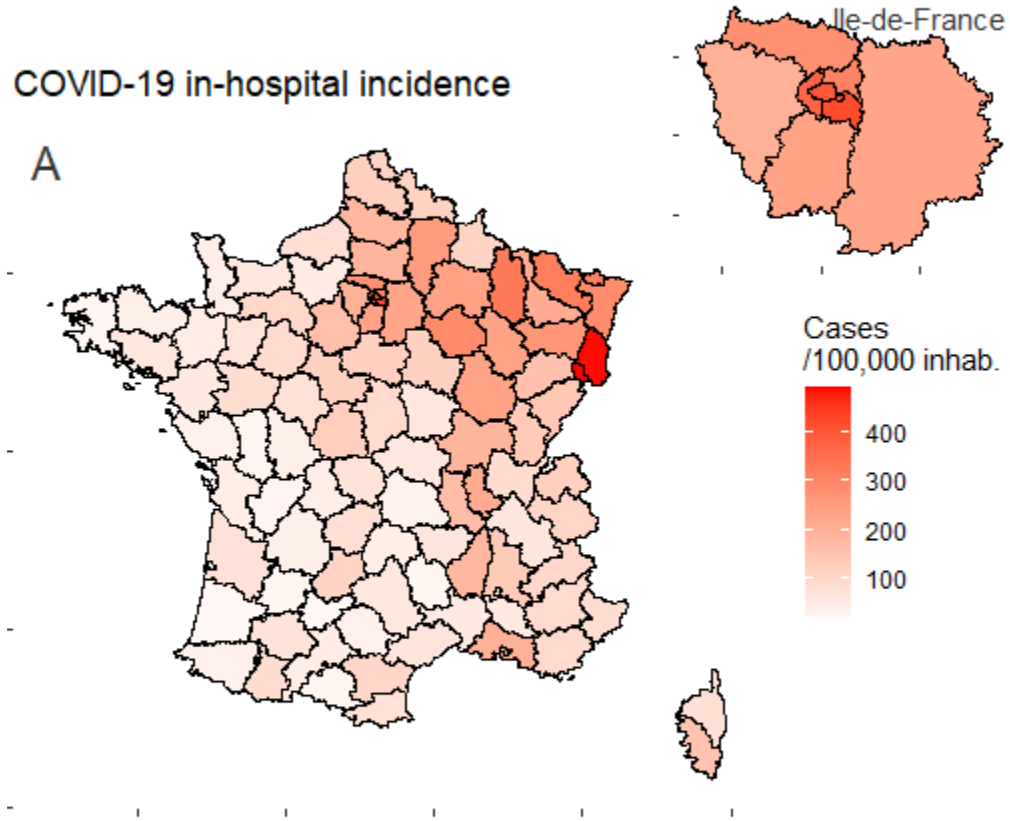
```
ggplot(data = DptLineD) +
aes(long, lat, group = group, fill = HIncid) +
geom_polygon() +
geom_path(color = "black") +
coord_equal()+
scale_fill_gradient2(low = muted("green"),
                     mid = "white",
                     high = "red",
                     midpoint = 0,
                     space = "Lab",
                     na.value = "grey50",
                     guide = "colourbar",
                     aesthetics = "fill") +
labs(title = "COVID-19 in-hospital incidence", fill = "Cases \n/100,000 inhab.") +
scale_x_continuous(name = NULL, labels = NULL) +
scale_y_continuous(name = NULL, labels = NULL) +
annotate("text", x = 100000, y = 7090000, label = "A", size = 6, alpha = 0.8) +
theme(plot.margin = margin(0, 0, 0, 0, "cm"), panel.background = element_rect(fill = "white"))
```

4. Cartographie

HIncidMap



HIncidMap2



5. Auto-corrélation spatiale

- Coefficient de Moran (1917-1988)

  Rappel Coefficient de corrélation de Pearson

$$r = \frac{\sum_{i}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

5. Auto-corrélation spatiale

- Coefficient de Moran (1917-1988)
  - Coefficient de corrélation
  - Pondéré par les distances
  - **Similarités** entre régions :

    écart à la moyenne de la région ***i*** ⇔ écart à la moyenne de la région ***j***

$$I = \frac{K \times \sum_{i,j}^{K} w_{ij}\left(Y_i - \bar{Y}\right)\left(Y_j - \bar{Y}\right)}{w_+ \times \sum_{i=1}^{K}\left(Y_i - \bar{Y}\right)^2}$$

Plus les zones i et j sont éloignées, moins le poids est important

$$w_{ij} = e^{\left(-\frac{d_{ij}}{\tau}\right)}$$

$$w_{ij} = \begin{cases} 1, si\ d_{ij} < \delta \\ 0, si\ non \end{cases}$$

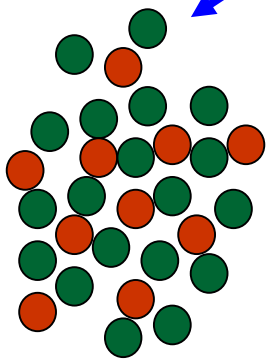
5. Auto-corrélation spatiale

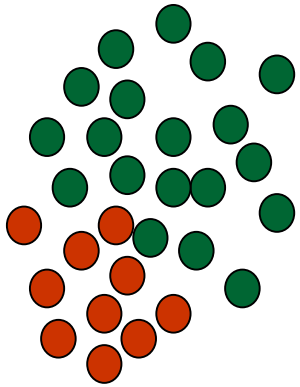
- Interprétation

  ***I* >0**: régions voisines: mêmes écarts à la moyenne = pattern sous forme de clusters

  ***I* <0**: régions voisines: ≠ écarts à la moyenne, = pattern régulier

  ***I* =0**: aucune corrélation spatiale





Mesure de l'écart à la moyenne générale :

pas d'interprétation locale possible

5. Auto-corrélation spatiale

- Matrice de voisinage

Liste des polygones

Nb points frontières

Une seule frontière suffit

```
nbDPT <- poly2nb(pl = DPT, row.names = DPT$CODE_DEPT, snap = 5, queen = TRUE)
```

- Estimation du coefficient et test

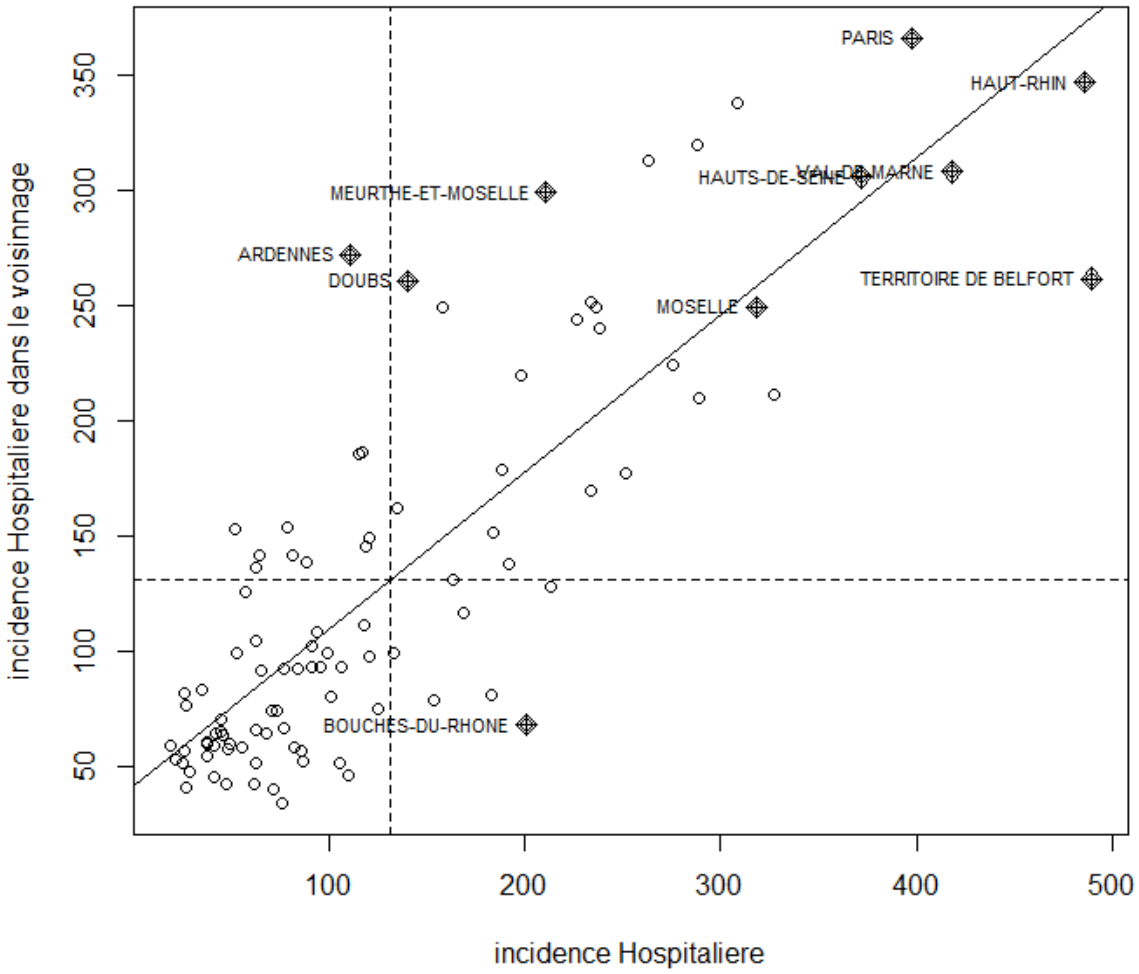
```
moran.test(x = DPT$HIncid, listw = nb2listw(nbDPT, style = "W", zero.policy = TRUE))
```

```
        Moran I test under randomisation

data:  DPT$HIncid
weights: nb2listw(nbDPT, style = "W", zero.policy = TRUE)

Moran I statistic standard deviate = 10.323, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.680923587      -0.010526316       0.004486116
```

5. Auto-corrélation spatiale

- Graphique des voisins

```
moran.plot(x = DPT$HIncid,
labels = DPT$NOM_DEPT,
listw = nb2listw(nbDPT,  style = "W",
zero.policy = TRUE),
xlab = "incidence Hospitaliere",
ylab = "incidence Hospitaliere dans le
voisinnage")
```



6. Réduction de dimension

- Nombreux co-Facteurs
  - ⇒ Fléau de la dimension (1961, Richard Bellman)
  - ⇒ Colinéarités

- Solutions
  - Extraction de caractéristiques
  - Sélection de variables => ACP + CAH {FactoMineR}

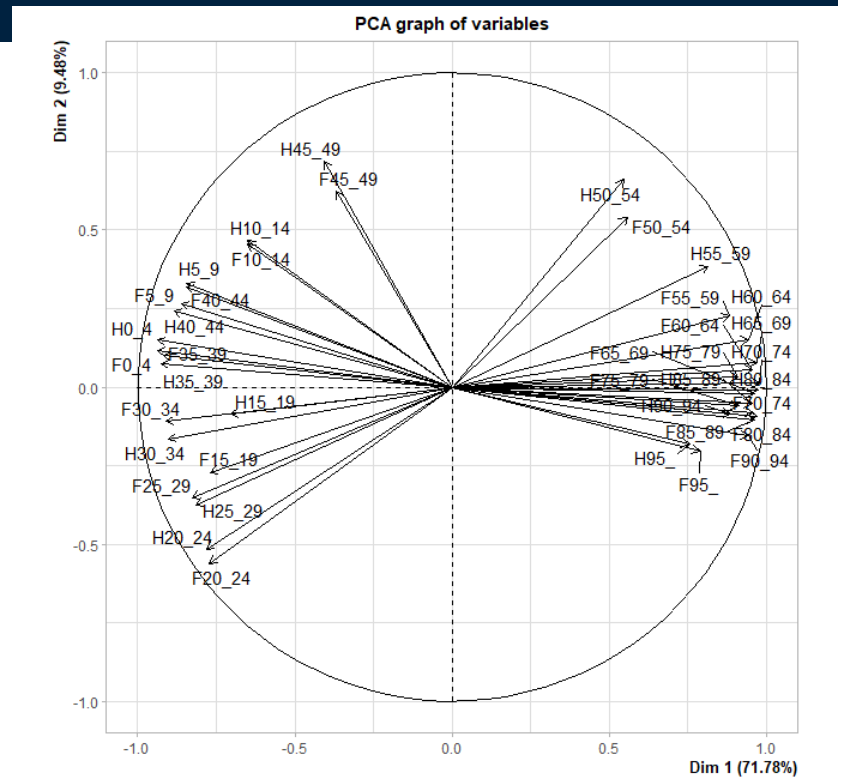
Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw. 2008; 25: 1-18

6. Réduction de dimension

colnames(DataSet[53:92])

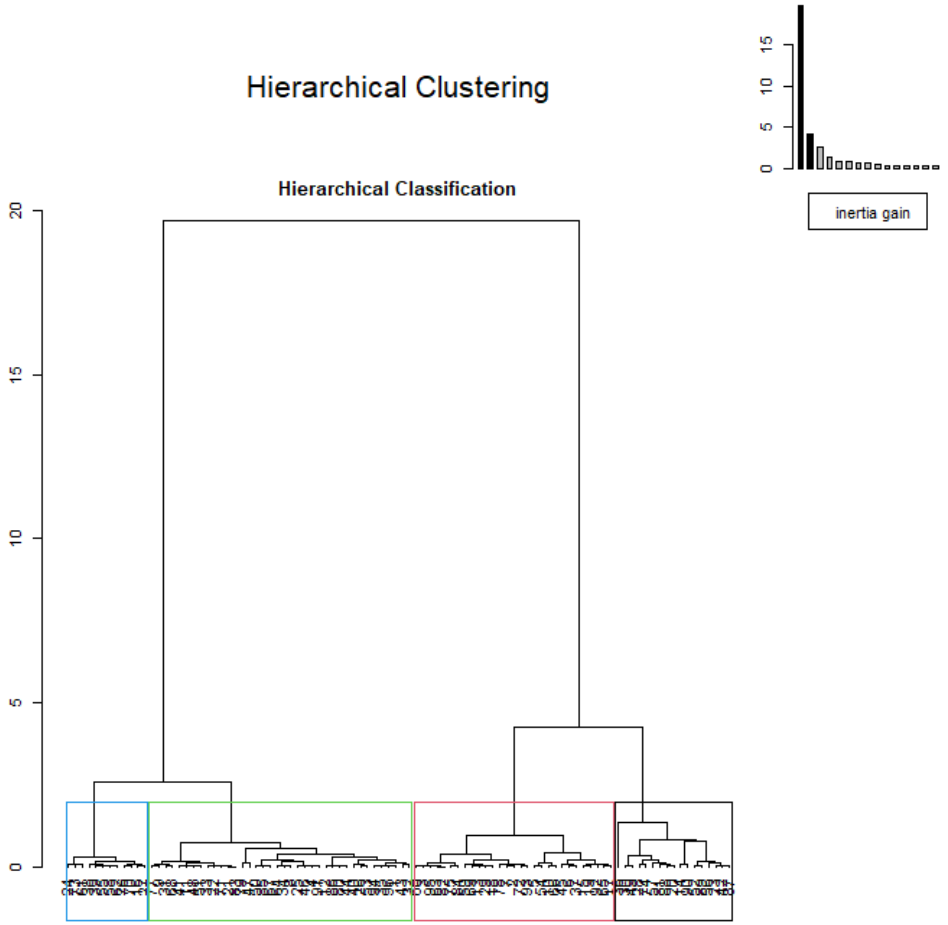
```
> colnames(DataSet[53:92])
 [1] "H0_4"    "H5_9"    "H10_14" "H15_19" "H20_24" "H25_29" "H30_34" "H35_39" "H40_44"
[10] "H45_49" "H50_54" "H55_59" "H60_64" "H65_69" "H70_74" "H75_79" "H80_84" "H85_89"
[19] "H90_94" "H95_"   "F0_4"   "F5_9"   "F10_14" "F15_19" "F20_24" "F25_29" "F30_34"
[28] "F35_39" "F40_44" "F45_49" "F50_54" "F55_59" "F60_64" "F65_69" "F70_74" "F75_79"
[37] "F80_84" "F85_89" "F90_94" "F95_"
```

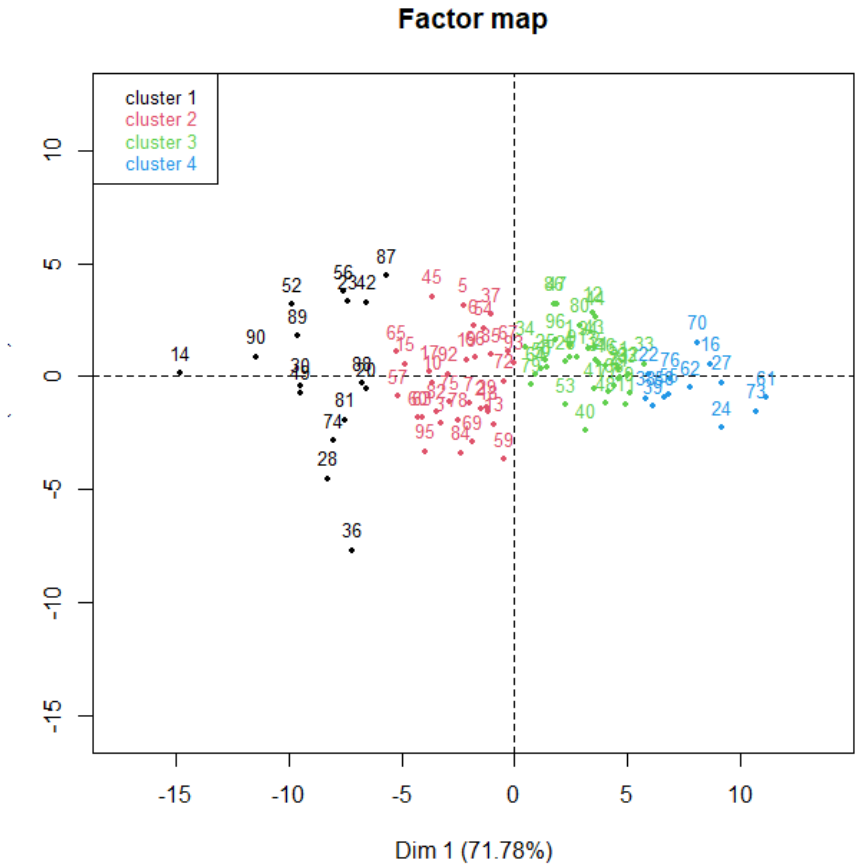
PCAAge <- PCA(DataSet[, c(53:92)], graph = T, ncp = 25) #graph = F



6. Réduction de dimension

classiffAge <- HCPC(PCAAge, nb.clust = 4)





6. Réduction de dimension

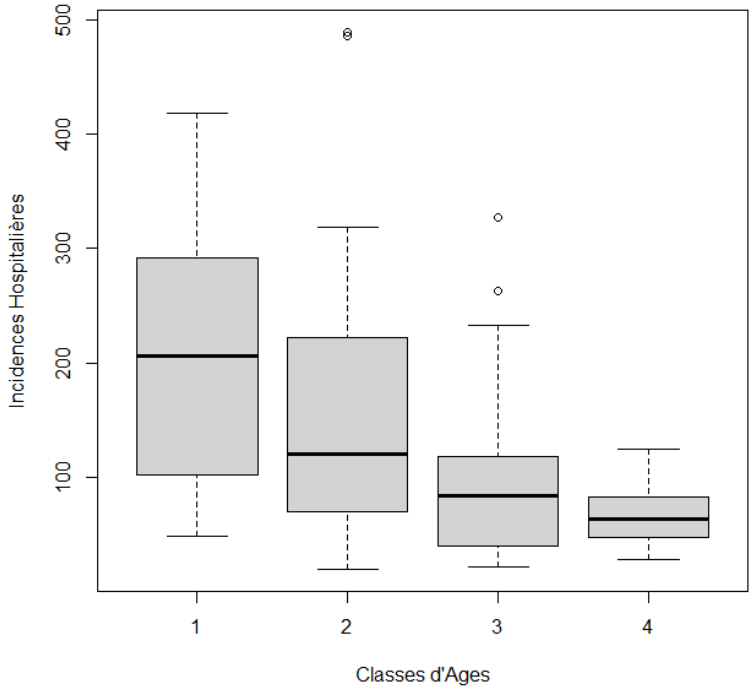
classiffAge$desc.var

```
Description of each cluster by quantitative variables
=====================================================
$`1`
          v.test Mean in category Overall mean sd in category   Overall sd      p.value
F30_34  7.085839     0.0354277141 0.0287190496   0.0033324439 0.0041268777 1.382045e-12
F40_44  6.937557     0.0332953353 0.0297334499   0.0015337819 0.0022379491 3.989393e-12
H30_34  6.934758     0.0330627334 0.0274486760   0.0030743743 0.0035287618 4.069172e-12
H40_44  6.924006     0.0323847980 0.0289247380   0.0014834072 0.0021782264 4.390483e-12
H0_4    6.904897     0.0316647511 0.0255755574   0.0034647462 0.0038439645 5.023978e-12
F0_4    6.891696     0.0303924391 0.0244594371   0.0034257115 0.0037525389 5.513128e-12
H35_39  6.890800     0.0337612486 0.0294277498   0.0018090274 0.0027412321 5.547933e-12
F35_39  6.843547     0.0355383998 0.0306974008   0.0018212988 0.0030834046 7.725570e-12
F25_29  6.761540     0.0327950708 0.0250013744   0.0060288365 0.0050242893 1.365322e-11
H25_29  6.370004     0.0313172301 0.0250472358   0.0049728973 0.0042904639 1.890235e-10
F5_9    5.479277     0.0322779386 0.0282349560   0.0035238529 0.0032162917 4.270678e-08
H5_9    5.421765     0.0335637141 0.0294863522   0.0036701366 0.0032780482 5.901331e-08
F20_24  4.959448     0.0311026599 0.0242849314   0.0058227634 0.0059921569 7.069383e-07
H20_24  4.675717     0.0309784465 0.0260001881   0.0038069785 0.0046409404 2.929279e-06
F10_14  3.928359     0.0318555047 0.0295358784   0.0032772532 0.0025738529 8.552761e-05
F15_19  3.851300     0.0311505287 0.0283275661   0.0023128820 0.0031950277 1.174926e-04
H10_14  3.474863     0.0331685631 0.0310032567   0.0033824343 0.0027161800 5.111152e-04
H45_49  3.466429     0.0342965792 0.0333287117   0.0013397669 0.0012170557 5.274217e-04
H15_19  3.028240     0.0324574339 0.0305460291   0.0024873233 0.0027513079 2.459828e-03
```

6. Réduction de dimension

```
DataSet <- cbind(DataSet, classiffAge$data.clust$clust)
dim(DataSet)
colnames(DataSet)[c(97)]
colnames(DataSet)[c(97)] <- c("ClAge")
DataSet$ClAge <- factor(DataSet$ClAge, levels = c(1, 2, 3, 4))
boxplot(DataSet$Hincid ~ DataSet$ClAge, xlab = "Classes d'Ages", ylab = "Incidences Hospitalières")
```

```
> summary(DataSet$ClAge)
 1  2  3  4
16 32 35 13
```



7. Approche régressive

- Rappel du cadre general de la modélisation statistique

$$g\big(E(Y/X_1, X_2)\big) = \alpha + \beta_1 X_1 + \beta_2 X_2$$

Fonction de lien

- Normale : $g(z) = z$
- Binomiale : $g(z) = logit(z) = log\left(\frac{z}{1-z}\right)$
- Poisson : $g(z) = log(z)$

7. Approche régressive

- Loi de Poisson (Denis Siméon)

$$\mathrm{p}(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

7. Approche régressive

- Loi de Poisson (Denis Siméon)

$$p(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Nombre d'occurences

7. Approche régressive

- Loi de Poisson (Denis Siméon)

$$p(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Nombre d'occurences

Nombre moyen d'occurence

7. Approche régressive

- Loi de Poisson (Denis Siméon)

$$\mathrm{p}(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$Y \sim P(\lambda)$$

$$E(Y) = Var(Y) = \lambda$$

7. Approche régressive

- Modèle de Poisson

$$log\big(E(Y/X_1, X_2)\big) = \alpha + \beta_1 X_1 + \beta_2 X_2$$

$$e^{\beta_1} = IR_1$$

Rapport d'incidences

7. Approche régressive

- Modèle de Poisson

$$log\big(E(Y/X_1,X_2)\big) = \alpha + \beta_1 X_1 + \beta_2 X_2$$

$$e^{\beta_1} = IR_1$$

Rapport d'incidences

$$log\big(E(Y/X_1,X_2)\big) = offset(\log(pop)) + \beta_1 X_1 + \beta_2 X_2$$

$$e^{\beta_1} = SIR_1$$

Rapport d'incidences standardisées

7. Approche régressive

- Modèle de Poisson

```
Mod1 <- glm(HCases ~ del, data = DataSet, family = "poisson")
summary((Mod1))
```

```
Call:
glm(formula = HCases ~ del, family = "poisson", data = DataSet)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-45.576  -24.577  -17.274   -5.197  146.210

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 7.0452798  0.0031539  2233.8   <2e-16 ***
del         0.0626632  0.0004172   150.2   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 138023  on 95  degrees of freedom
Residual deviance: 114046  on 94  degrees of freedom
AIC: 114823

Number of Fisher Scoring iterations: 6
```

7. Approche régressive

- Modèle de Poisson

```
Mod2 <- glm(HCases ~ offset(log(Pop2020)) + del, data = DataSet, family = "poisson")
summary(Mod2)
```

```
Call:
glm(formula = HCases ~ offset(log(Pop2020)) + del, family = "poisson",
    data = DataSet)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-38.032  -18.408   -6.752    2.573   79.544

Coefficients:
              Estimate Std. Error  z value Pr(>|z|)
(Intercept) -6.4553212  0.0031566 -2045.04   <2e-16 ***
del          0.0380085  0.0004893    77.68   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 51311  on 95  degrees of freedom
Residual deviance: 45340  on 94  degrees of freedom
AIC: 46117

Number of Fisher Scoring iterations: 5
```

7. Approche régressive

- Modèle Additif Généralisé (GAM) package {mgcv} (Simon Wood)

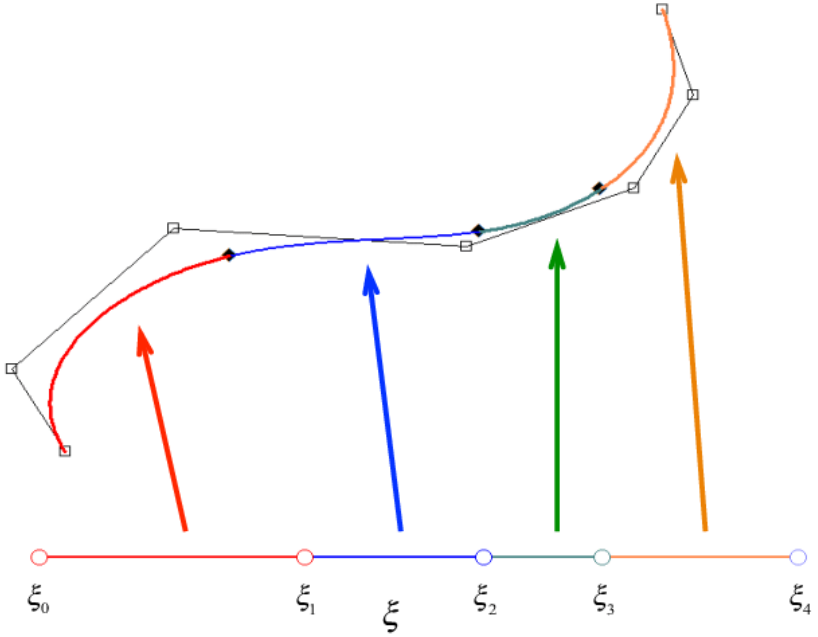
$$g\big(E(Y/X_1, X_2)\big) = \alpha + \beta_1 X_1 + s(X_2)$$

7. Approche régressive

- Modèle Additif Généralisé (GAM) package {mgcv} (Simon Wood)

$$g\big(E(Y/X_1, X_2)\big) = \alpha + \beta_1 X_1 + s(X_2)$$

Fonction de lissage
Ex. Spline

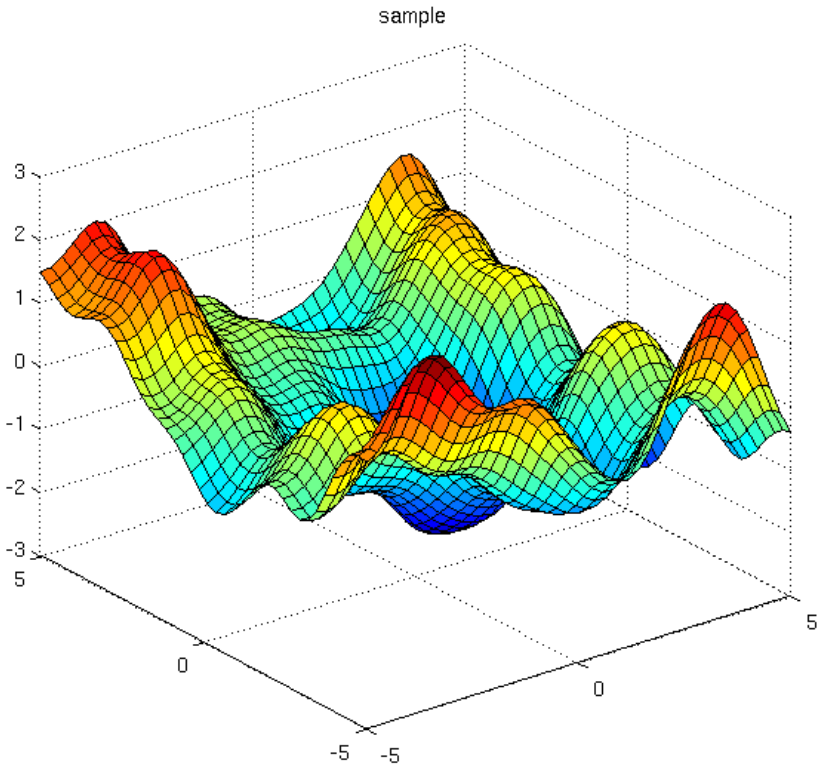


7. Approche régressive

- Modèle Additif Généralisé (GAM) package {mgcv} (Simon Wood)

$$g\big(E(Y/X_1, X_2)\big) = \alpha + \beta_1 X_1 + s(Long, Lat)$$

Gaussian Process



7. Approche régressive

- Modèle Additif Généralisé (GAM)

```
Mod3 <- gam(HCases ~ offset(log(Pop2020)) + del + s(X_CENTROID, Y_CENTROID, bs="gp", k=75, m=2),
data = DataSet, family = "poisson")
summary(Mod3)
```

```
Family: poisson
Link function: log

Formula:
HCases ~ offset(log(Pop2020)) + del + s(X_CENTROID, Y_CENTROID,
    bs = "gp", k = 75, m = 2)

Parametric coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.865128   0.007584 -905.25  < 2e-16 ***
del          0.013243   0.001713    7.73 1.07e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                           edf Ref.df Chi.sq p-value
s(X_CENTROID,Y_CENTROID) 73.95     74  39873  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.957   Deviance explained = 97.6%
UBRE = 13.368  Scale est. = 1         n = 96
```

7. Approche régressive

- Modèle Additif Généralisé (GAM)

```
Uni1 <- gam(HCases ~ offset(log(Pop2020)) + del + s(X_CENTROID, Y_CENTROID, bs="gp", k=75, m=2),
data = DataSet, family = nb())
summary(Uni1)
```

Negative Binomiale

```
Family: Negative Binomial(8.297)
Link function: log

Formula:
HCases ~ offset(log(Pop2020)) + del + s(X_CENTROID, Y_CENTROID,
    bs = "gp", k = 75, m = 2)

Parametric coefficients:
             Estimate Std. Error  z value Pr(>|z|)
(Intercept) -6.828643   0.041039 -166.393   <2e-16 ***
del          0.014907   0.005844    2.551   0.0107 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                           edf Ref.df Chi.sq p-value
s(X_CENTROID,Y_CENTROID) 29.62  43.03  250.5  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.857   Deviance explained = 85.3%
-REML = 690.53  Scale est. = 1         n = 96
```

7. Approche régressive

```
CaseModel <- gam(HCases ~ offset(log(Pop2020)) +
HCQCQn + del + ClAge + ClClim +  ClUrb + ClSys  +
s(X_CENTROID, Y_CENTROID, bs = "gp", k = 75, m = 2),
data = DataSet, family = nb())
summary(CaseModel)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.4235133  0.3347618 -19.188  < 2e-16 ***
HCQCQn       0.0003482  0.0002500   1.393  0.16366
del          0.0205824  0.0069985   2.941  0.00327 **
ClAge2      -0.1385602  0.1585908  -0.874  0.38228
ClAge3      -0.3810439  0.2360696  -1.614  0.10650
ClAge4      -0.0891266  0.3033668  -0.294  0.76892
ClClim1     -0.2066128  0.1840411  -1.123  0.26159
ClClim3      0.0762824  0.1817499   0.420  0.67470
ClClim4      0.5395285  0.2937994   1.836  0.06630 .
ClUrb2      -0.4946955  0.2625301  -1.884  0.05952 .
ClUrb3      -0.1671657  0.3074695  -0.544  0.58666
ClUrb4      -0.0629445  0.3549255  -0.177  0.85924
ClSys2       0.0664156  0.1312747   0.506  0.61291
ClSys3      -0.0367546  0.1558073  -0.236  0.81351
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                            edf Ref.df Chi.sq  p-value
s(X_CENTROID,Y_CENTROID) 20.23  30.81  71.51 4.67e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.904   Deviance explained = 86.8%
-REML = 693.13  Scale est. = 1         n = 96
```

7. Approche régressive

exp(CaseModel$coefficients[2:14])
exp(CaseModel$coefficients[2:14] - 1.96*summary(CaseModel)$se[2:14])
exp(CaseModel$coefficients[2:14] + 1.96*summary(CaseModel)$se[2:14])

```
> exp(CaseModel$coefficients[2:14])
   HCQCQn       del    clAge2    clAge3    clAge4   clClim1   clClim3   clClim4    clUrb2    clUrb3    clUrb4
1.0003483 1.0207956 0.8706109 0.6831479 0.9147297 0.8133345 1.0792673 1.7151980 0.6097566 0.8460594 0.9389956
   clSys2    clSys3
1.0686707 0.9639127
> exp(CaseModel$coefficients[2:14] - 1.96*summary(CaseModel)$se[2:14])
   HCQCQn       del    clAge2    clAge3    clAge4   clClim1   clClim3   clClim4    clUrb2    clUrb3    clUrb4
0.9998582 1.0068888 0.6380120 0.4300989 0.5047330 0.5670355 0.7558235 0.9643332 0.3644906 0.4631029 0.4683223
   clSys2    clSys3
0.8262294 0.7102510
> exp(CaseModel$coefficients[2:14] + 1.96*summary(CaseModel)$se[2:14])
  HCQCQn      del   clAge2   clAge3   clAge4  clClim1  clClim3  clClim4   clUrb2   clUrb3   clUrb4   clSys2
1.000839 1.034895 1.188008 1.085079 1.657768 1.166617 1.541124 3.050713 1.020062 1.545696 1.882705 1.382252
  clSys3
1.308168
```

7. Approche régressive

- Coefficient d'auto-corrélation de Moran sur les résidus

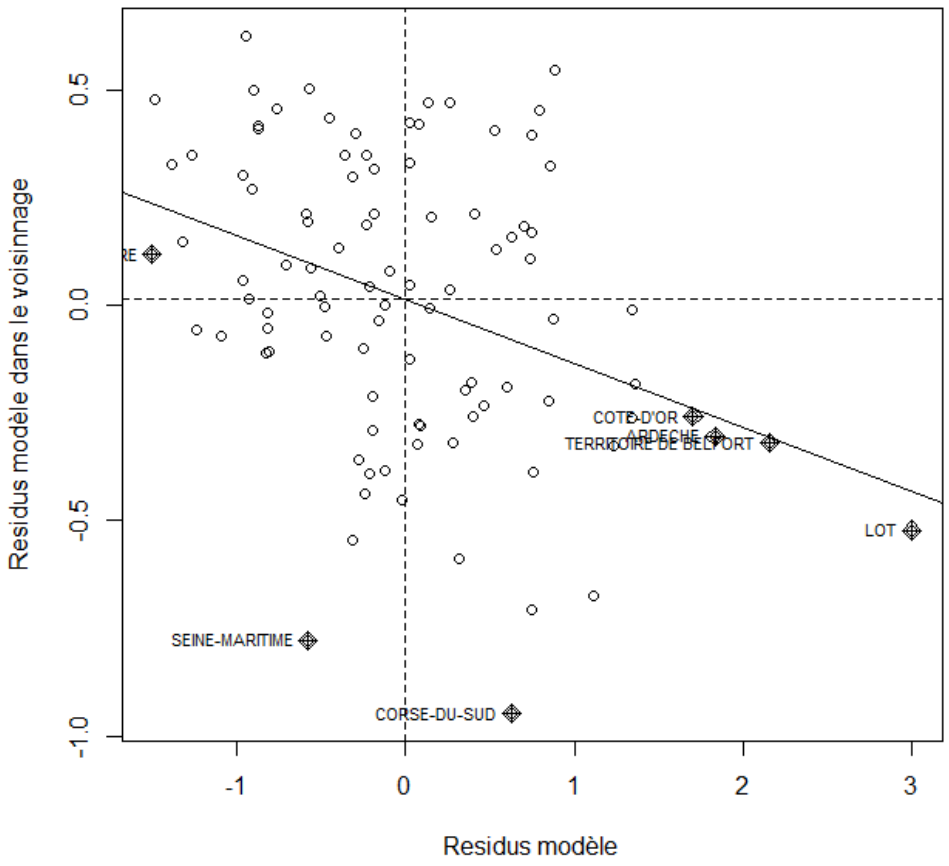
```
DPT@data <- cbind(DPT@data, residuals(CaseModel, type = "scaled.pearson"))
dim(DPT@data)
colnames(DPT@data)[109]
colnames(DPT@data)[109] <- "HRes"

nbDPT <- poly2nb(pl = DPT, row.names = DPT$CODE_DEPT, snap = 5, queen = TRUE)
moran.test(x = DPT$HRes, listw = nb2listw(nbDPT, style = "W", zero.policy = TRUE))
```

```
        Moran I test under randomisation

data:  DPT$HRes
weights: nb2listw(nbDPT, style = "W", zero.policy = TRUE)

Moran I statistic standard deviate = -2.0471, p-value = 0.9797
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
     -0.148254004      -0.010526316       0.004526319
```

7. Approche régressive

- Coefficient d'auto-corrélation de Moran sur les résidus

moran.plot(x = DPT$HRes, labels = DPT$NOM_DEPT, listw = nb2listw(nbDPT, style = "W"), xlab = "Residus modèle", ylab = "Residus modèle dans le voisinnage")



8. Analyse de sensibilité

- Garder le même modèle,

mais changer d'Outcome : nombre d'Allergies aux urgences

```
DataAllerg <- read.csv2("C:/.../DataAllergies.csv")
colnames(DataAllerg)
summary(DataAllerg$DepCode)
DataSet <- data.frame(DataSet, DataAllerg[match(DataSet[, "DptN2"], DataAllerg[, "DepCode"]),])
```

```
> colnames(DataAllerg)
[1] "DepCode"     "DepName"     "NbPass"      "NbAllergies" "NbAllergH"
> summary(DataAllerg$DepCode)
   Length     Class      Mode
       96 character character
```

8. Analyse de sensibilité

```
Allerg1 <- gam(NbAllergies ~ offset(log(Pop2020)) +
HCQCQn + del + ClAge + ClClim +  ClUrb + ClSys +
s(X_CENTROID, Y_CENTROID, bs = "gp", k = 74, m = 2),
data = DataSet, family = nb())
summary(Allerg1)
```

```
Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.830e+00  3.073e-01 -28.735   <2e-16 ***
HCQCQn      -2.273e-05  2.651e-04  -0.086   0.9317
del          4.925e-03  7.834e-03   0.629   0.5296
ClAge2       2.414e-02  1.656e-01   0.146   0.8841
ClAge3       1.722e-01  2.469e-01   0.697   0.4857
ClAge4       1.870e-01  3.223e-01   0.580   0.5617
ClClim1     -1.527e-01  1.541e-01  -0.991   0.3218
ClClim3     -2.793e-01  1.577e-01  -1.771   0.0765 .
ClClim4      5.260e-02  2.437e-01   0.216   0.8291
ClUrb2       1.972e-02  2.701e-01   0.073   0.9418
ClUrb3      -2.049e-02  3.148e-01  -0.065   0.9481
ClUrb4      -1.634e-01  3.784e-01  -0.432   0.6659
ClSys2       8.512e-03  1.356e-01   0.063   0.9499
ClSys3       1.756e-01  1.557e-01   1.128   0.2593
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                            edf Ref.df Chi.sq p-value
s(X_CENTROID,Y_CENTROID) 2.003  2.005  2.266   0.323

R-sq.(adj) =  0.788   Deviance explained = 19.2%
-REML = 466.03  Scale est. = 1         n = 87
```

9. Références

- Simon Wood; Generalized Additive Models: and introduction with R. CRC Press
- P. Cowperwait; Introductory time series with R. Springer
- R. Bivand, EJ. Pebesma, V. Gomez-Rubio; Applied spatial data analysis with R. Springer
- L. Waller, C. Gotway; Applied Spatial Statistics for Public Health Data. Wiley