Spatial Exploratory Data Analysis with R

Applied Spatial Econometrics

Lecture 2 (1.5h)

Thibault LAURENT
thibault.laurent@univ-tlse1.fr

Toulouse School of Economics

27th January of 2016

Introduction

Mapping

Interactive Analysis

Spatial Interpolation

Purposes of SEDA

One definition of SEDA : "Visualising spatial distributions and local patterns of spatial autocorrelation" (L. Anselin).

- Presenting statistical information on the map.
- Coupling traditionnal tools in Exploratory Data Analysis with a map.
- Use specific tools of Spatial Data Analysis.

Preparation of the data

- Install the following packages:

```
> install.packages(c("classInt", "RColorBrewer",
  "GeoXp", "Guerry", "spatstat"))
```

- Load the necessary packages:

```
> require("rgdal")
> require("maptools")
```

- Load some codes presented in Chapter 1:

```
> source("preambule.R")
```

```
OGR data source with driver: ESRI Shapefile
Source: "Donnees/World WGS84", layer: "Pays_WGS84"
with 251 features
It has 1 fields
```

Introduction

Mapping

Interactive Analysis

Spatial Interpolation

Preparation of the data (1)

First, we select the countries of Sub-Saharian Africa :

```
> pays.af<-c("Seychelles","Equatorial Guinea","Gabon","Botswana",
 "Mauritius","South Africa","Namibia","Angola","Swaziland","Congo",
 "Cape Verde","Ghana","Sudan","Djibouti","Nigeria","Sao Tome and Princi
 "Cameroon","Lesotho","Gambia, The","Chad","Senegal","Kenya","Ivory Coa
 "Zambia","Burkina Faso","Tanzania, United Republic of","Benin","Rwanda
 "Uganda","Comoros","Guinea-Bissau","Mali","Mozambique","Guinea","Ethio
 "Madagascar", "Malawi","Togo","Sierra Leone","Niger","Central African
 "Eritrea","Burundi","Somalia","Zimbabwe","Liberia","Zaire")
> africa.sub=world[world@data$NOM%in%pays.af,]
```

Then we create a `data.frame` of the GDP :

```
> africa.df<-data.frame(pib=c(25000,26500,18100,15700,15400,11300,7800,
 5700,4600,4400,3300,2500,2600,2700,2100,2300,2100,1900,2500,2000,1800,
 1700,1400,1600,1600,1400,1400,1300,1200,1100,1200,1100,1300,900,800,11
 1300,800,900,700,600,600,600,700,400),row.names=pays.af)
```

Preparation of the data (2)

We create the `SpatialPolygonsDataFrame` object :

```
> africa.sub<-SpatialPolygonsDataFrame(africa.sub,africa.df)
```

We merge Maghreb and Sub-Saharian Africa :

```
> africa <- spRbind(northAf, africa.sub)
```

We add a variable `region` to the country (`E`, East Af, `C`, Central Af, `N`, North Af, `Au`, Austral Af and `O`, West Af) :

```
> africa@data$region <- factor(c("N","N","N","N","N","N","O",
 "O","C","E","E","O","O","O","E","O","O","C","E","O","O",
 "O","E","O","O","C","O","O","C","O","E", "E","C","C","C",
 "C","E","E","E","E","C","E","E","E","E","E","E","Au","Au",
 "E","Au","Au","Au"))
```

Representing a categorial variable

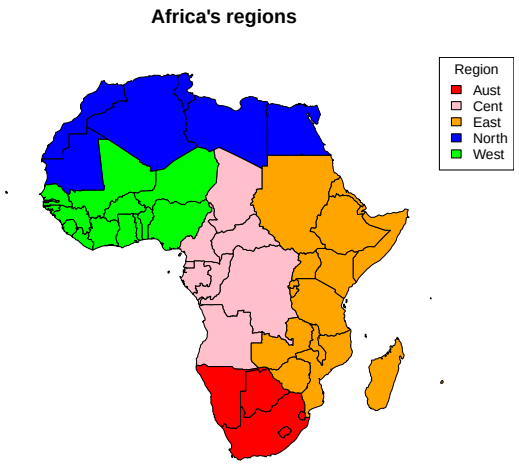
Objective : copy this map `http://fr.wikipedia.org/wiki/Afrique#mediaviewer/File:Zones_Afrique.jpg`

Choice of appropriate colors :

```
> pal.reg <- c("red","pink","orange",
 "blue","green")
```

Plotting the map :

```
> ind=as.numeric(africa@data$region)
> plot(africa, col=pal.reg[ind])
> title("Africa's regions")
> legend("topright",legend=c("Aust","Cent"
 "East","North","West"),cex=0.8,
 title="Region",fill=pal.reg)
```



Representing a numeric variable

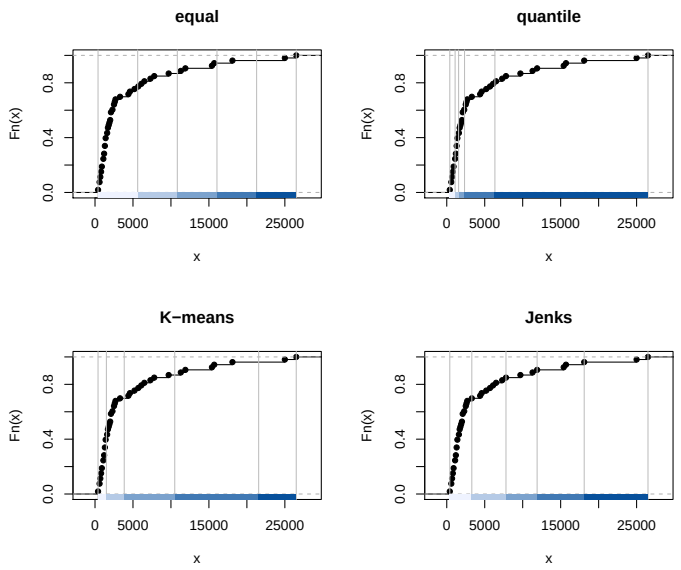
How to cut the GDP variable ? R package **classInt**.

```
> require("classInt")
> pib <- africa@data$pib
```

- The `equal` style divides the range of the variable into *n* parts.
- The `quantile` style provides quantile breaks; arguments to quantile may be passed through ....
- The `kmeans` style uses kmeans to generate the breaks.
- The `jenks` style has been imported from Jenks' Basic code, and has been checked for consistency with ArcView, ArcGIS, and MapInfo.

Application

```
> require("RColorBrewer")
> plotclr <- brewer.pal(5,"Blues")
> pal1 <- plotclr[c(1,5)]
> opar <- par(mfrow=c(2,2))
> plot(classIntervals(pib,5,"equal"),
  pal=pal1, main="equal")
> plot(classIntervals(pib,5,"quantile"),
  pal=pal1, main="quantile")
> plot(classIntervals(pib,5,"kmeans"),
  pal=pal1, main="K-means")
> plot(classIntervals(pib,5,"jenks"),
  pal=pal1, main="Jenks")
> par(opar)
```



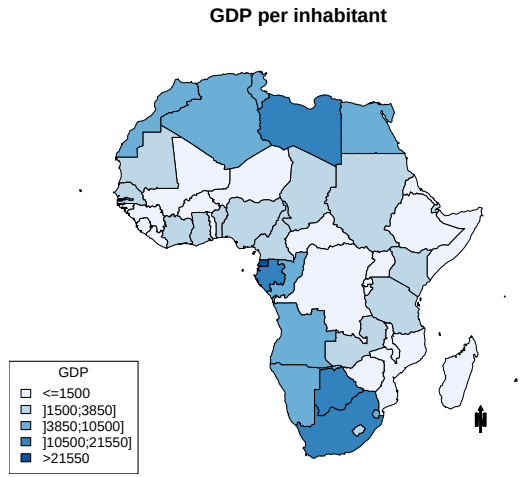
Choropleth map

The breakpoints of the bins :

```
> bk=round(classIntervals(pib, 5,"kmeans")$brks,
 digits=1)
```

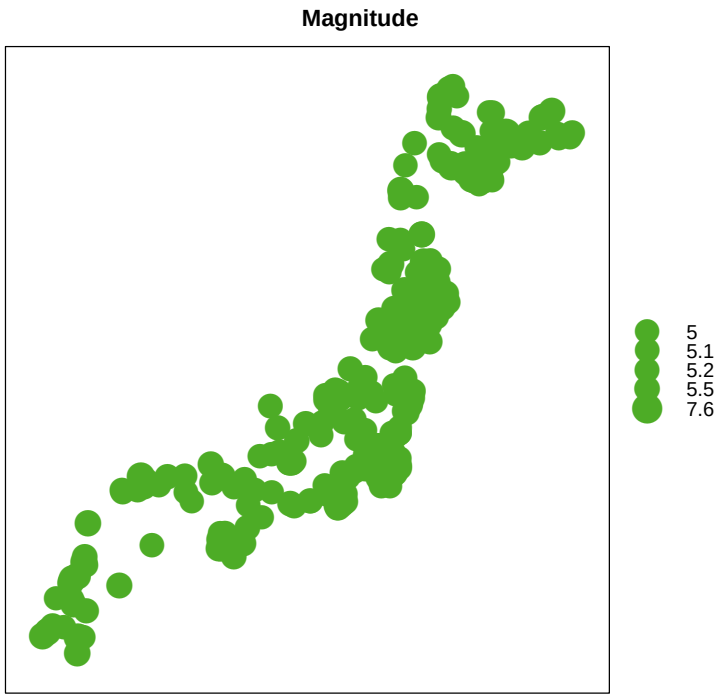
Plotting :

```
> ind=findInterval(pib, bk, all.inside=TRUE
> plot(africa, col=plotclr[ind])
> decoup <- c("<=1500","]1500;3850]",
 "]3850;10500]","]10500;21550]",">21550")
> legend("bottomleft", legend = decoup,
 cex = 0.8,title="GDP", fill=plotclr)
> SpatialPolygonsRescale(layout.north.arrow,
 offset = c(50,-30), scale = 5, plot.grid=F)
> title("GDP per inhabitant")
```

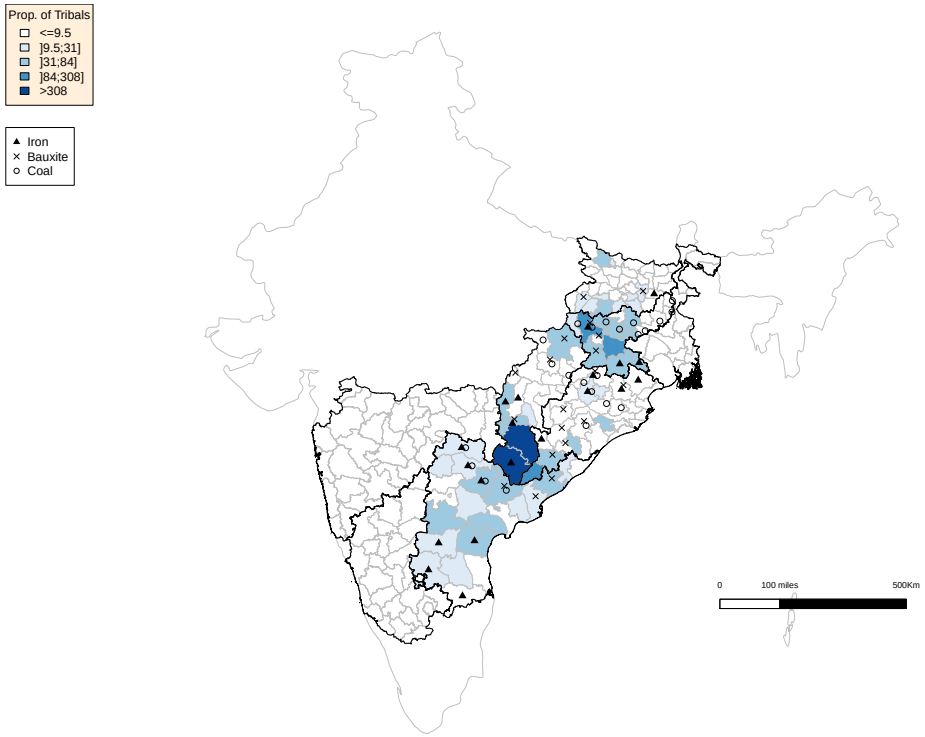


Bubble plot of spatial data

```
> bubble(seisme.jp, "Magnitude")
```



Several variables at the same time



Essays in Political Economy of Maoist Conflict in India (K. Bathia thesis, 2015)

Introduction

Mapping

Interactive Analysis

Spatial Interpolation

GeoXp: Interactive exploratory spatial data analysis

- Can be download on the CRAN: `http://cran.r-project.org`.
- A R vignette: `vignette("presentation_geoxp")`.
- **GeoXp** works with objects of class `SpatialXXXDataFrame`.
- Around 30 functions presented in T. Laurent, A. Ruiz-Gazen and C. Thomas-Agnan, "**GeoXp**: an R package for Exploratory Spatial Data Analysis", *JSS*.

Interactivity principle

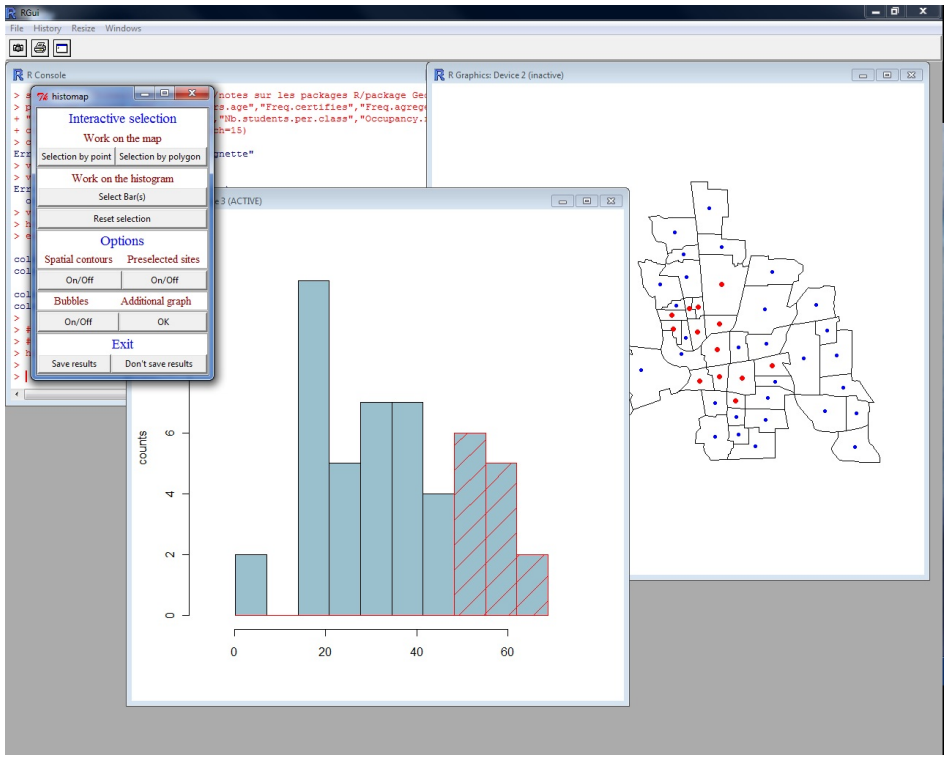
The functions of GeoXp allow an exchange between a map and a statistical graph. This exchange is bilateral :

- the user selects spatial units on the map, whose corresponding elements are then highlighted on the graph
- the user selects elements of the graph, whose corresponding spatial units are then highlighted on the map

The highlighting is done by a change of color or a change of symbol.

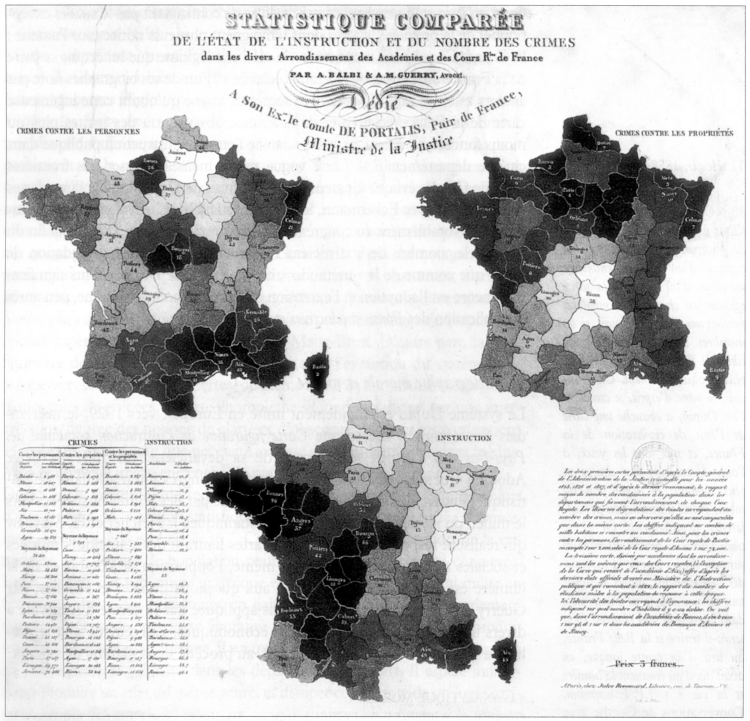
Three windows

A Tcl/Tk window for the menu, a graphical window for the statistical graph and a graphical window for the map



An application with the Guerry's data

- In 1833, André-Michel Guerry wrotes *Essai sur la statistique morale de la France*
- His goal: can we explain the violence in the French department by some social variables (education, clergy, etc.).



Initialisation

- Load the packages:

```
> require("GeoXp")
> require("Guerry")
```

- Two variables of interest : crime against persons (`Crime_pers`) and crime against property `Crime_prop`, for 1000 inhabitants :

```
> gfrance85@data$Crime_pers<-1000/gfrance85@data$Crime_pers
> gfrance85@data$Crime_prop<-1000/gfrance85@data$Crime_prop
```

- Modify the labels of the individuals :

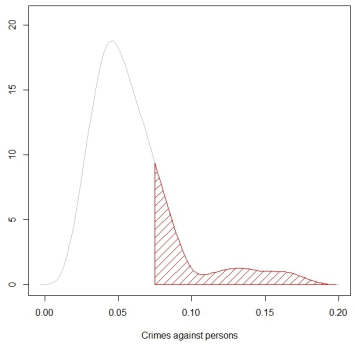
```
> row.names(gfrance85) <- as.character(gfrance85@data$Department)
```

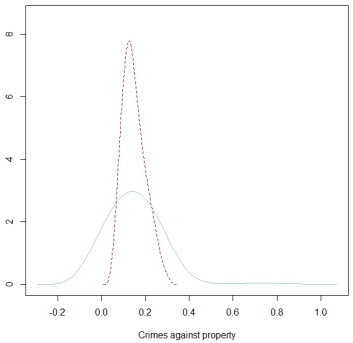
- Explanatory variables: see description in `help(Guerry)`.
- Mapping of the variables `Crime`:

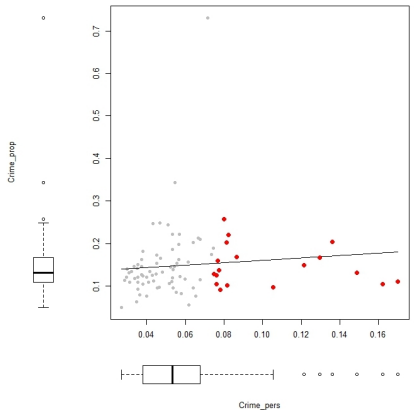
```
> spplot(gfrance85,"Crime_pers",col.regions=rev(heat.colors(16)))
> spplot(gfrance85,"Crime_prop",col.regions=rev(heat.colors(16)))
```

Non parametric estimates of the variables of interest

```
> dbledensitymap(gfrance85, c("Crime_pers", "Crime_prop"),  xlab =
  c("Crimes against persons", "Crimes against property"), identify=T)
```

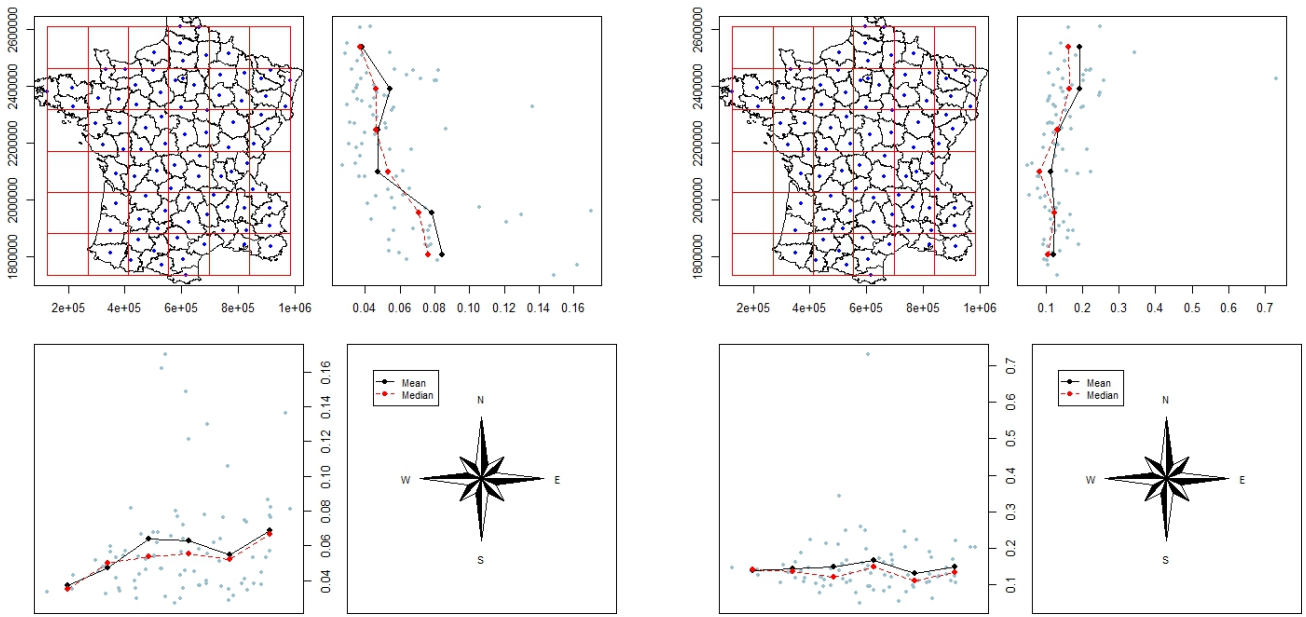






Analysis of a spatial tendency - "Driftmap"

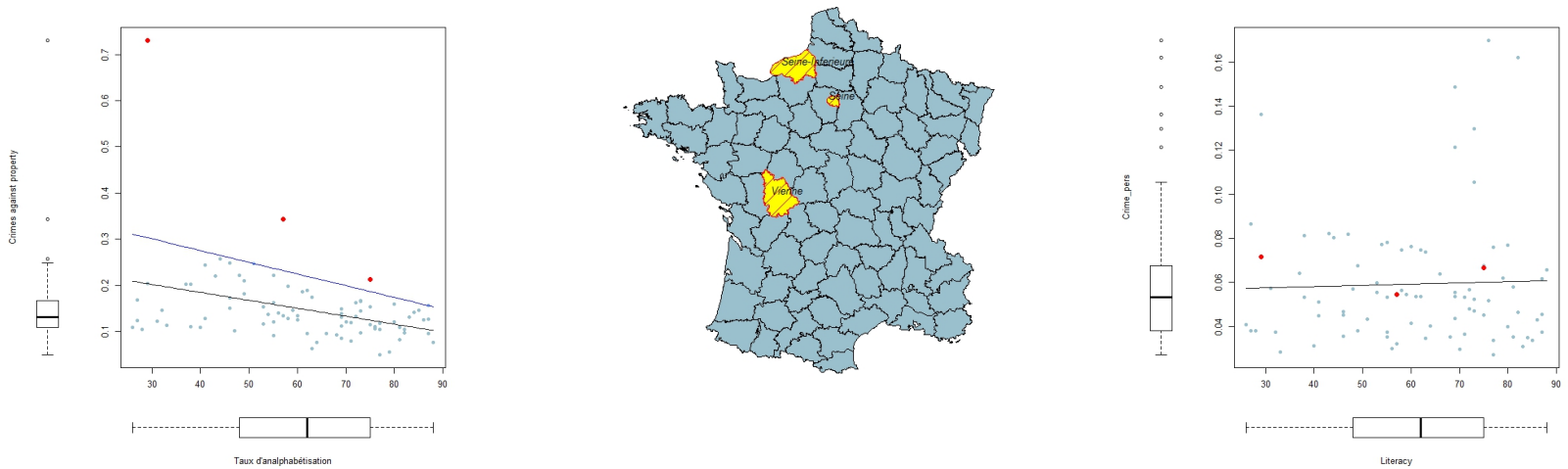
```
> driftmap(gfrance85, "Crime_pers")
> driftmap(gfrance85, "Crime_prop")
```



North/South and West-East tendencies for crimes against persons.
A small South/North tendency for crimes against property.

Relationship between crimes and education

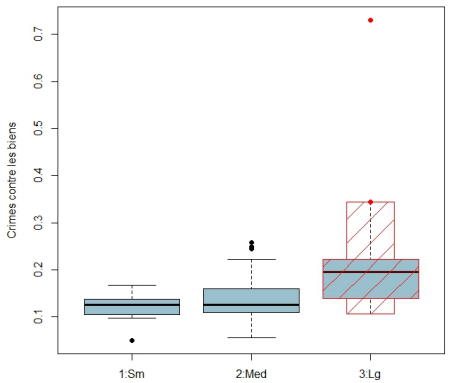
```
> scattermap(gfrance85, c("Literacy", "Crime_prop"),
  xlab="Literacy", ylab="Crimes", identify=TRUE)
```

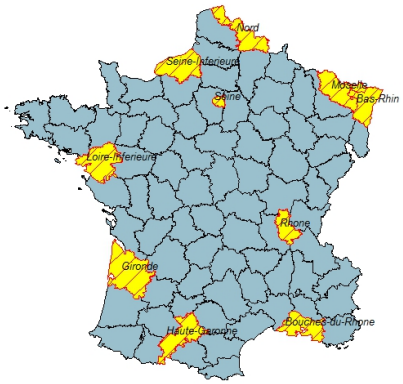


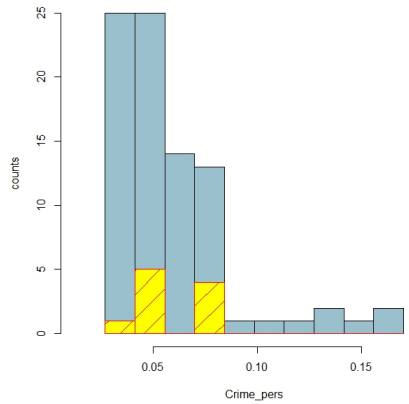
Negative effect of education on the crimes against property. Detection of outliers. No effect of education on the crimes against persons.

Relationship between crimes and presence of a huge city in the department

```
> polyboxplotmap(gfrance85, c("MainCity", "Crime_prop"), identify=TRUE)
```



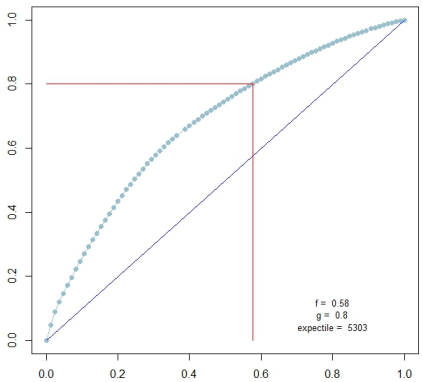


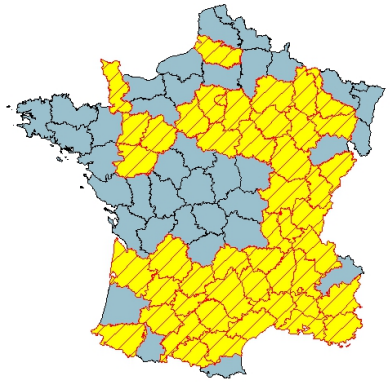


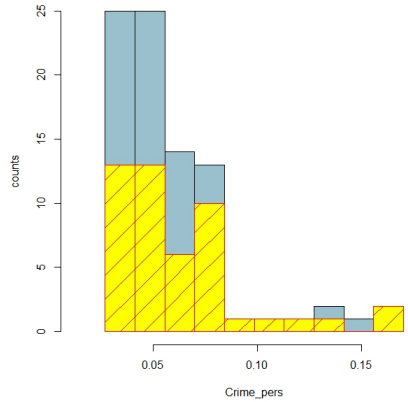
Crimes against property higher in departments with a huge city.

Relationship between crimes and donations to the poor

```
> ginimap(gfrance85, "Donations")
```





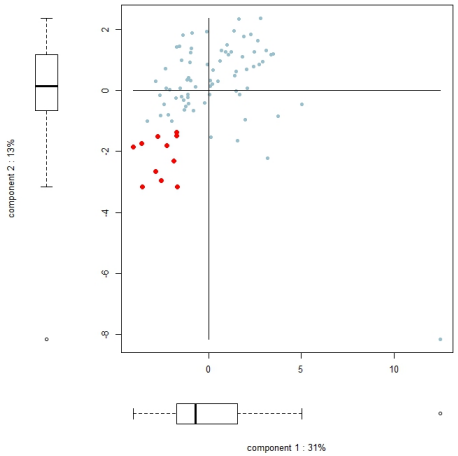


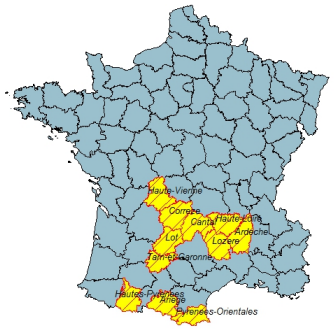
80% of the donations are done by 58% of the departments. Non obvious relationship.

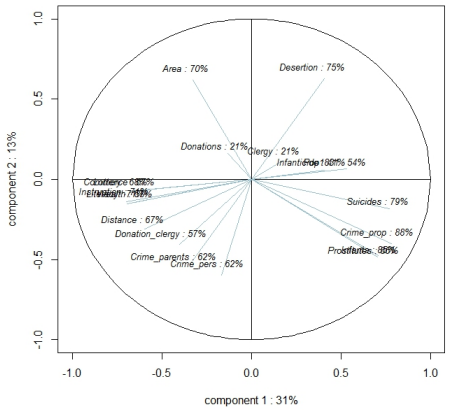
Principal Component Analysis

```
> pcamap(gfrance85, c(7:12, 14:26), identify=TRUE)
```

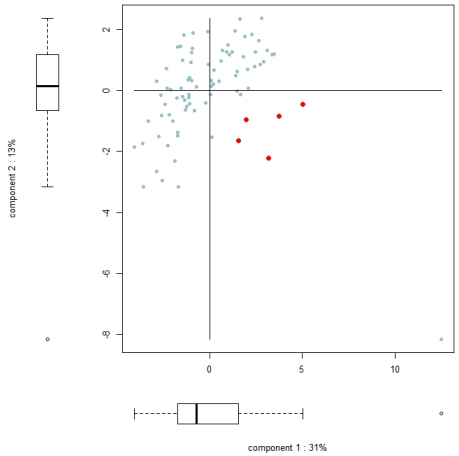
"Crime against person"

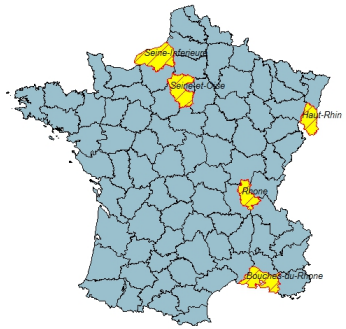






"Crime against property"





Summary

We got the same results that those obtained by A.-M. Guerry :

- The 2 variables `Crimes` are not spatially distributed in the same way. High values of crime against property are not situated in the same departement than high values of crime against person.
- Crime against property are most important in urban areas.
- Crime against person are not related to education.
- Crimes seem to be related to wealth.

Introduction

Mapping

Interactive Analysis

Spatial Interpolation

Spatial Interpolation in Spatial Point Pattern analysis

We would like to represent the intensity of the earthquake in Japan. We need to define an observation window :

```
> require("spatstat")
> bb <- bbox(japan2)
> W = owin(bb[1,], bb[2,])
```

we define the spatial point pattern process of the earhtquake located in the window W :

```
> wp.seisme <- ppp(x=seisme.jp2@coords[,1],
 y = seisme.jp2@coords[,2], window=W)
```

We mark the spatial point process

```
> marks(wp.seisme) <- seisme.jp2@data$Magnitude
```

Kernel Smoothed Intensity of Point Pattern

Choice of the bandwidth with `bw.diggle` :

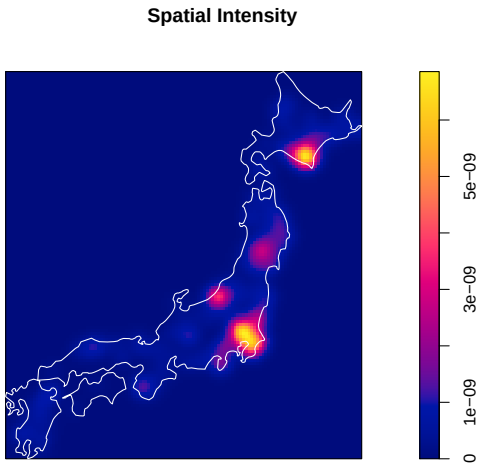
```
> bw.diggle(wp.seisme)
```

Kernel Smoothed Intensity of Point Pattern :

```
> D.seisme<- density(wp.seisme, sigma=3000(
```

Plotting :

```
> plot(D.seisme, main="Spatial Intensity")
> plot(japan2, add=TRUE, border="white")
```



Spatial Interpolation in geostatistic (1)

155 measures of biological concentrations near the river Meuse

```
> data(meuse)
> coordinates(meuse) <- ~x+y
> proj4string(meuse) <- CRS("+init=epsg:28992")
```

We create a Spatial Point pattern process :

```
> require("rgeos")
> test <- gBuffer(meuse, width=250)
> W = as(test, "owin")
> meuse.ppp <- ppp(meuse$x, meuse$y, window=W)
> marks(meuse.ppp) <- meuse$zinc
```

We would like to interpolate the concentration in Zinc. We use the Inverse-distance weighted smoothing of observations at irregular points (see details in `help(idw)`)

```
> dens.zinc <- idw(meuse.ppp, 2)
```

Spatial Interpolation in geostatistic (2)

Other available methods : Mark of nearest neighbour, non parametric spatial smoothing.

```
> op = par(mfrow=c(2,2), mar=c(0,0,3.3,0))
> plot(meuse.ppp,
  main="Concentrations in zinc")
> plot(nnmark(meuse.ppp),
  main="Mark of Nearest Neighbour")
> plot(Smooth(meuse.ppp),
  main="Spatial smoothing")
> plot(dens.zinc,
  main="Inv-dist weighted smoothing")
> par(op)
```