OpenSpat PCA example

Pr Yves Brostaux, ULiège

Openspat ULisboa

June 2019

Example II - full spatial data

Raw data : WorldClim data

- 19 bioclimatic variables, averages for 1970-2000
- 10 minutes spatial resolution ($\simeq$ 340 km^2)

Available at `http://worldclim.org/version2` or directly from within R (package `raster`)

Data content}

BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Outline

- Extract the multivariate data from spatial data
- Interpret PCA analysis and components
- Add the spatial components back

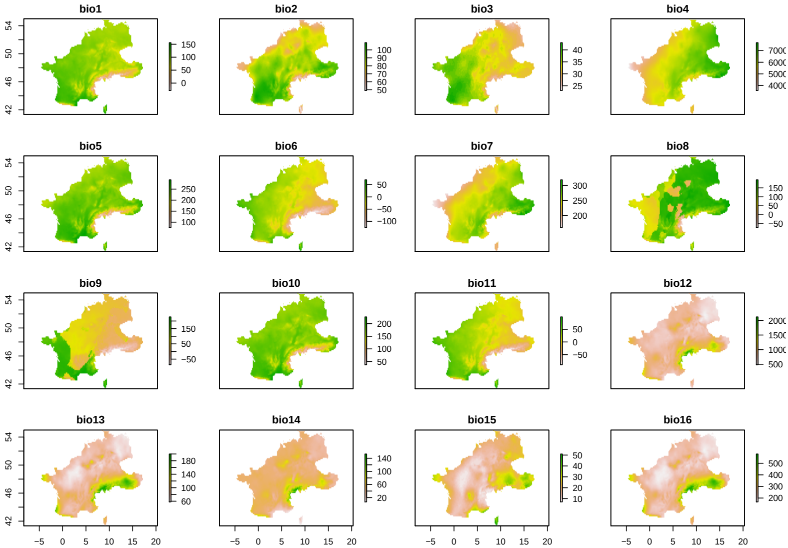
Get needed data

```
# get worldwide bioclimatic data
library(raster)
wclim <- getData('worldclim', res=10, var='bio',
                 path="../../")

# loading countries borders
library(maptools)
data(wrld_simpl) # simplified world contries borders
# selection of west european region
eur_simpl <- wrld_simpl[wrld_simpl@data$SUBREGION==155,]
```

Select only western europe

```
# check crs
crs(wclim);crs(wrld_simpl)

# crop climatic data for western europe
climEur <- crop(wclim, bbox(eur_simpl))
climEur <- mask(climEur, eur_simpl)

plot(climEur)
```

Bioclimatic data



Preparing for PCA

```r
# extract data for PCA analysis
climdat <- getValues(climEur)
climdat <- na.omit(as.data.frame(climdat))

# index of deleted NA values
nai <- as.vector(attr(na.omit(climdat), "na.action"))

# basic descriptive statistics
summary(climdat)

# correlation matrix
cor(climdat)
```

Preparing for PCA

```r
# basic descriptive statistics
summary(climdat)
```

```
##       bio1             bio2             bio3
##  Min.   :-29.00   Min.   : 49.00   Min.   :23.00   Min.
##  1st Qu.: 82.00   1st Qu.: 78.00   1st Qu.:31.00   1st
##  Median : 92.00   Median : 85.00   Median :33.00   Medi
##  Mean   : 91.39   Mean   : 84.56   Mean   :33.71   Mean
##  3rd Qu.:106.00   3rd Qu.: 92.00   3rd Qu.:36.00   3rd
##  Max.   :156.00   Max.   :109.00   Max.   :43.00   Max.
##       bio5            bio6             bio7
##  Min.   : 75.0   Min.   :-126.0   Min.   :160.0   Min.
##  1st Qu.:215.0   1st Qu.: -36.0   1st Qu.:233.0   1st Q
##  Median :228.0   Median : -19.0   Median :249.0   Media
##  Mean   :226.7   Mean   : -20.9   Mean   :247.6   Mean
##  3rd Qu.:242.0   3rd Qu.:   0.0   3rd Qu.:263.0   3rd Q
##  Max.   :292.0   Max.   :  70.0   Max.   :319.0   Max.
```

Preparing for PCA

```r
# correlation matrix
cor(climdat)
```

```
##              bio1        bio2        bio3        bio4
## bio1   1.00000000  0.36077195  0.62936050 -0.38869852
## bio2   0.36077195  1.00000000  0.65128722  0.21464899
## bio3   0.62936050  0.65128722  1.00000000 -0.57868983
## bio4  -0.38869852  0.21464899 -0.57868983  1.00000000
## bio5   0.87202364  0.66658743  0.53421843  0.05321799
## bio6   0.91580086  0.05928900  0.60738004 -0.68325669
## bio7  -0.16254745  0.64279415 -0.15505606  0.87490188
## bio8  -0.02130125  0.03734866 -0.35810075  0.56436773
## bio9   0.71482889  0.26566040  0.66156100 -0.60511944
## bio10  0.94014183  0.46856791  0.46620765 -0.05422684
## bio11  0.94953835  0.20846450  0.69523807 -0.65478139
## bio12 -0.56131165 -0.04939058 -0.06213942 -0.07607415 -
## bio13 -0.55037411 -0.01973373 -0.11893486  0.04274221 -
```

PCA with package FactoMineR

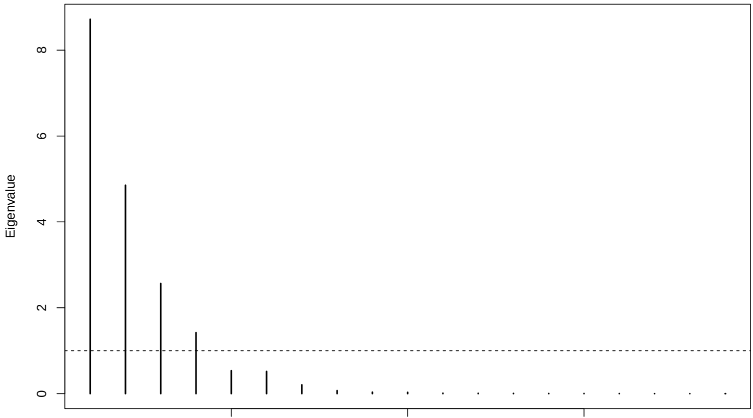
```r
library(FactoMineR)

# PCA on standardized data (default)
climdat.pca <- PCA(climdat)

# available data in PCA object
climdat.pca
```

PCA with package FactoMineR

```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 4799 individuals, describe
## *The results are available in the following objects:
##
##    name                 description
## 1  "$eig"               "eigenvalues"
## 2  "$var"               "results for the variables"
## 3  "$var$coord"         "coord. for the variables"
## 4  "$var$cor"           "correlations variables - dimensio
## 5  "$var$cos2"          "cos2 for the variables"
## 6  "$var$contrib"       "contributions of the variables"
## 7  "$ind"               "results for the individuals"
## 8  "$ind$coord"         "coord. for the individuals"
## 9  "$ind$cos2"          "cos2 for the individuals"
## 10 "$ind$contrib"       "contributions of the individuals'
## 11 "$call"              "summary statistics"
## 12 "$call$centre"       "mean of the variables"
```

Screeplot

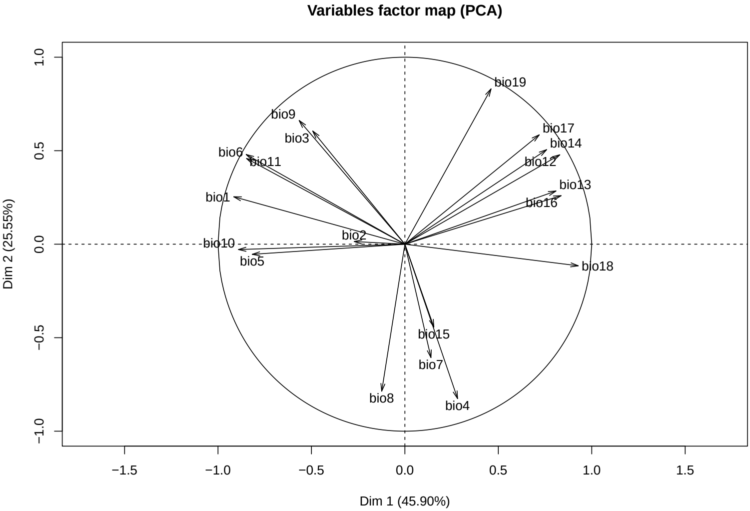
```
# screeplot
plot(climdat.pca$eig[,1], type="h", lwd=2, ylab="Eigenvalue
abline(h=1, lty="dashed")
```



Interpreting the axes

```
# circle of correlation
plot(climdat.pca, choix="var")

# significant correlations with the axes
dimdesc(climdat.pca)
```

Interpreting the axes



Interpreting the axes

```
## $Dim.1
## $Dim.1$quanti
##        correlation       p.value
## bio18    0.9278900  0.000000e+00
## bio16    0.8360953  0.000000e+00
## bio12    0.8290807  0.000000e+00
## bio13    0.8086465  0.000000e+00
## bio14    0.7586011  0.000000e+00
## bio17    0.7191266  0.000000e+00
## bio19    0.4612173 1.825133e-251
## bio4     0.2819676  2.064254e-88
## bio15    0.1550844  3.221264e-27
## bio7     0.1396126  2.558462e-22
## bio8    -0.1239376  6.887554e-18
## bio2    -0.2711174  1.260948e-81
## bio3    -0.4931424 1.315090e-292
## bio9    -0.5659367  0.000000e+00
```

Interpreting the axes

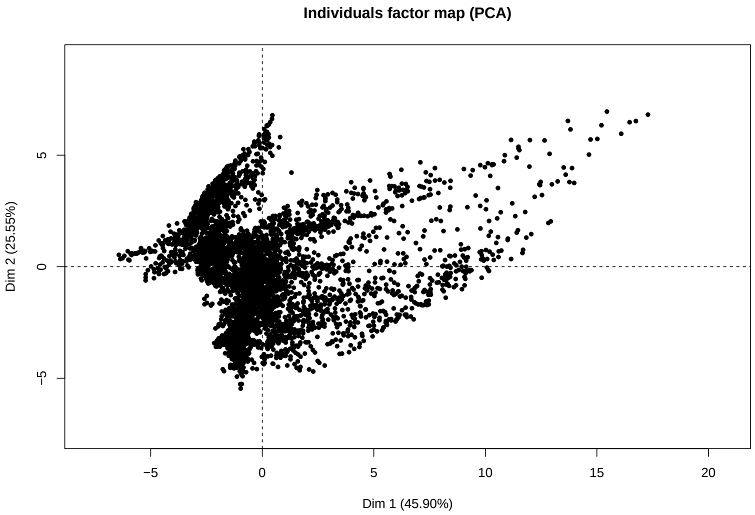
```
# highest correlations with the axes
subset(as.data.frame(dimdesc(climdat.pca)$Dim.1$quanti),
       abs(correlation)>0.5)
```

```
##        correlation p.value
## bio18    0.9278900       0
## bio16    0.8360953       0
## bio12    0.8290807       0
## bio13    0.8086465       0
## bio14    0.7586011       0
## bio17    0.7191266       0
## bio9    -0.5659367       0
## bio5    -0.8168103       0
## bio6    -0.8472543       0
## bio11   -0.8492738       0
## bio10   -0.8906935       0
## bio1    -0.9158772       0
```

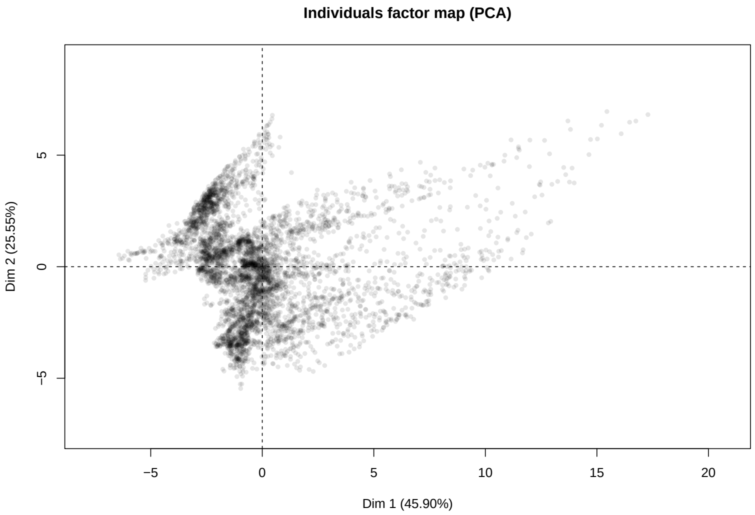
Trends and groups

```
plot(climdat.pca, choix="ind", label="none")

plot(climdat.pca, choix="ind", label="none",
     col.ind=adjustcolor("black", alpha.f = 0.1))
```

Trends and groups



Trends and groups

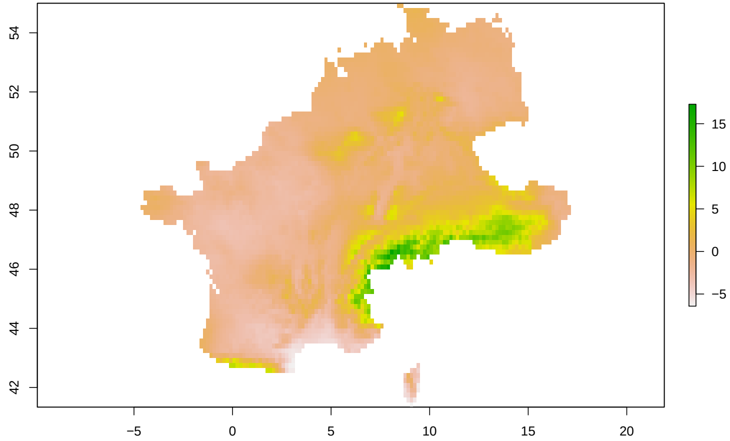


Back to spatial

```
df <- as.data.frame(climdat.pca$ind$coord)

# Spatial trend for component 1
climPCA1 <- raster(climEur)
values(climPCA1)[-nai] <- df[,1]
plot(climPCA1)
```

```
# Spatial trend for component 2
climPCA2 <- raster(climEur)
values(climPCA2)[-nai] <- df[,2]
plot(climPCA2)
```

Back to spatial



Back to spatial