

Doctoral Course (Doctoral Schools ABIES and GAIA)
Environmental Genetics

Multivariate Data Analysis:

*an introduction with a focus on Principal
Component Analysis of SNP data*

Denis Laloë, Tatiana Zerjal, Xavier Rognon

UMR GABI – INRA/AgroParisTech



Multivariate Data Analysis

Terminology

Factorial methods

Geometric data analysis

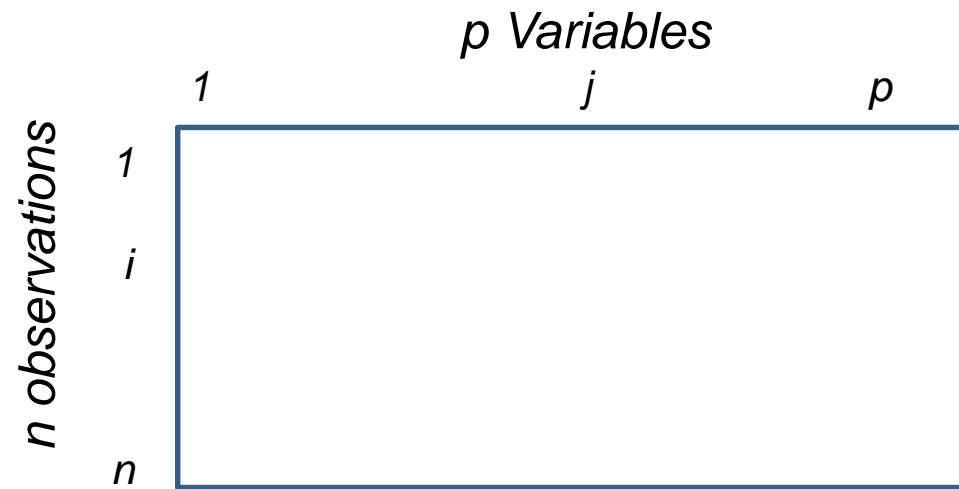
Data analysis “*à la française*”

Multivariate Data Analysis

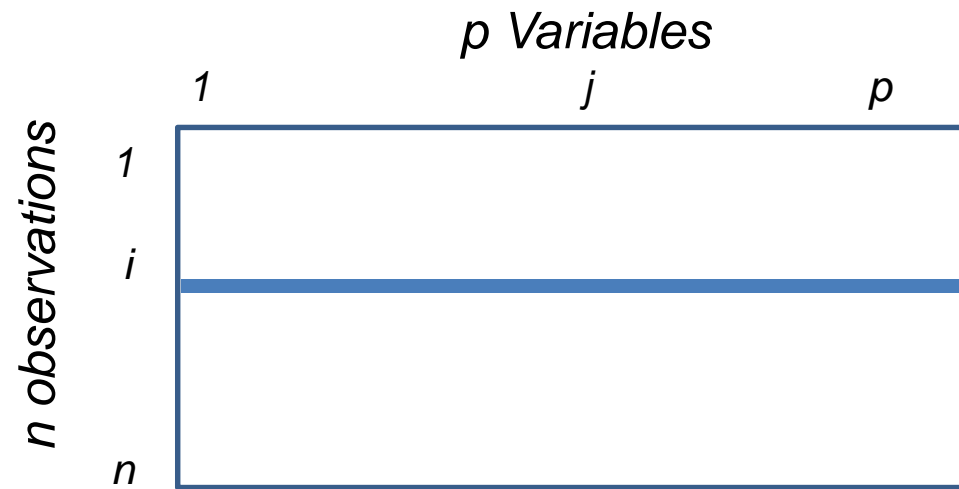
Context and aims

Multidimensional data
Exploratory/Descriptive
Reduction of dimensionality
Graphics

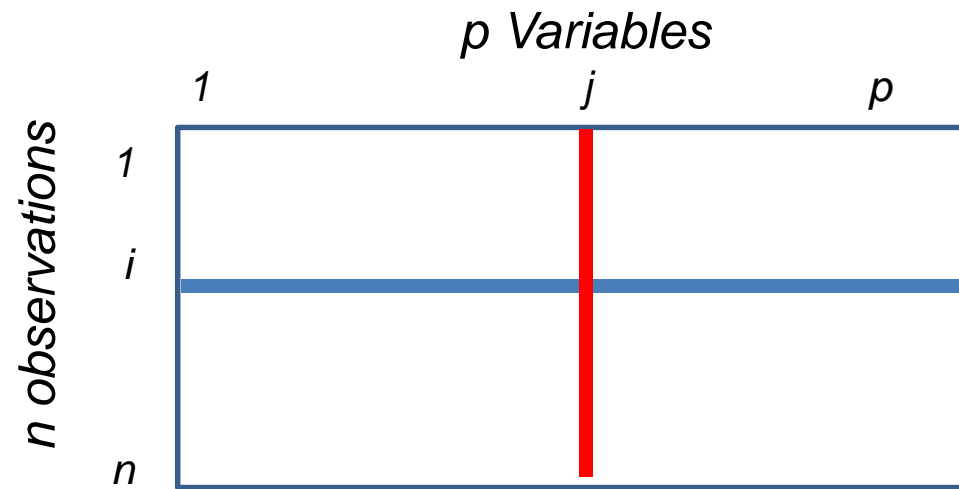
A data table of quantitative variables



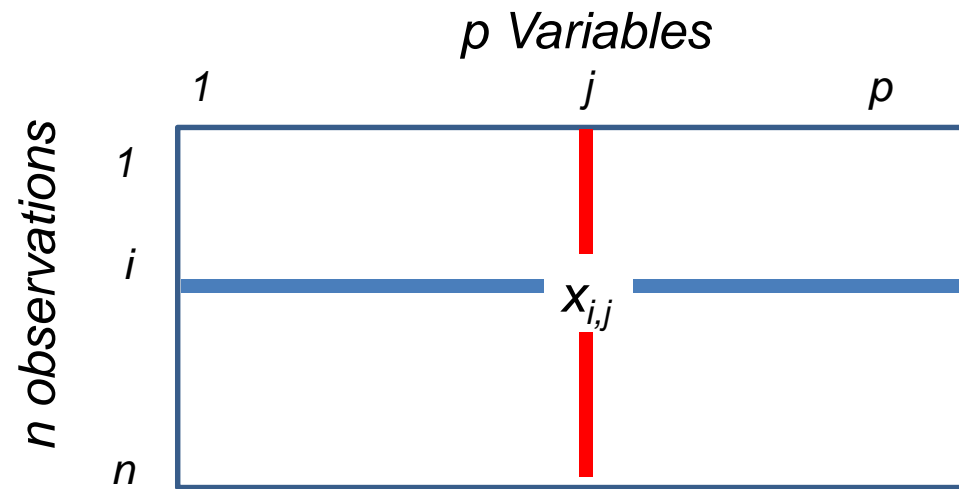
A data table of quantitative variables



A data table of quantitative variables

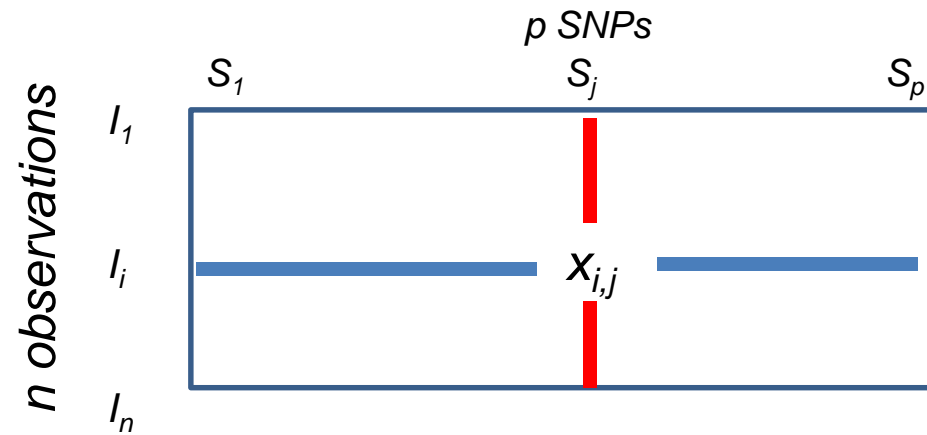


A data table of quantitative variables



An example of data representation

$$\mathbf{X} = \begin{bmatrix} SNP_1 & SNP_2 & \dots & SNP_p \\ 0 & 2 & \dots & 2 \\ 2 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 2 & \dots & 2 \end{bmatrix}$$



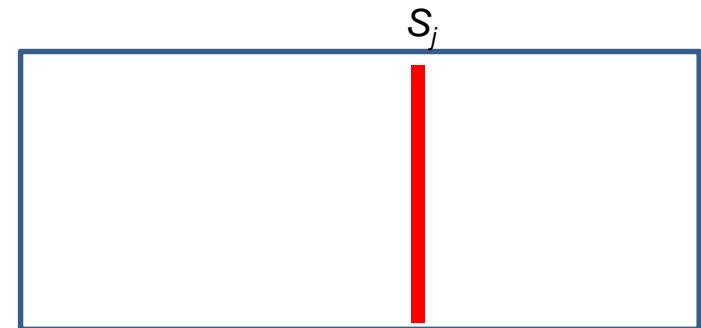
Two points of views:

The observations (the individuals)



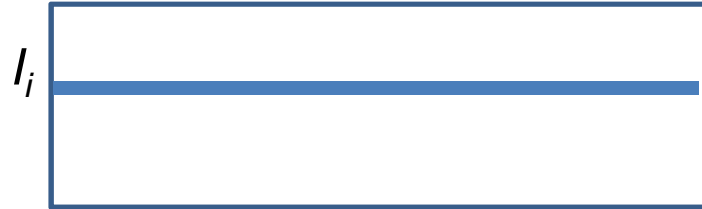
The i^{th} observation is defined by p SNPs

The variables (the SNPs)

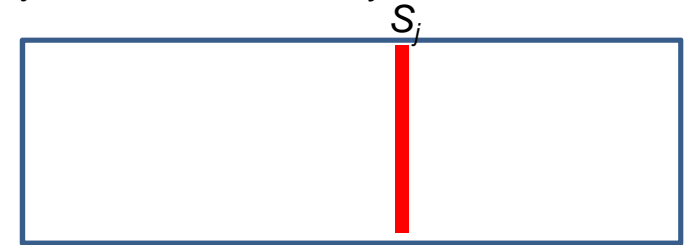


The j^{th} SNP is defined by n observations

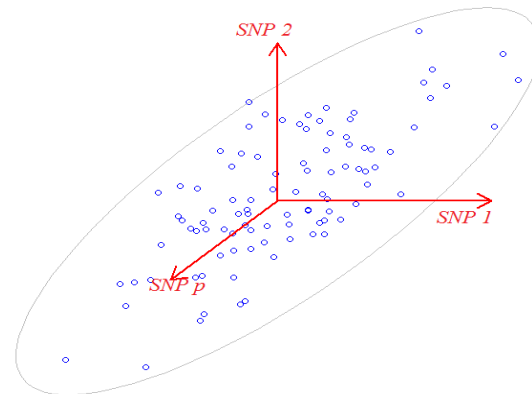
The i^{th} observation is defined by p SNPs



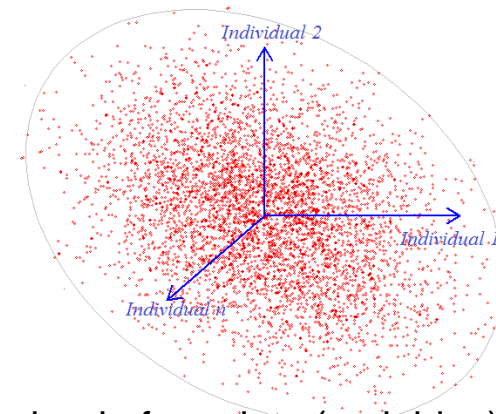
The j^{th} SNP is defined by n observations



Two points of views
Two geometric representations



A cloud of n points (observations) in an hyperspace with p dimensions (variables)

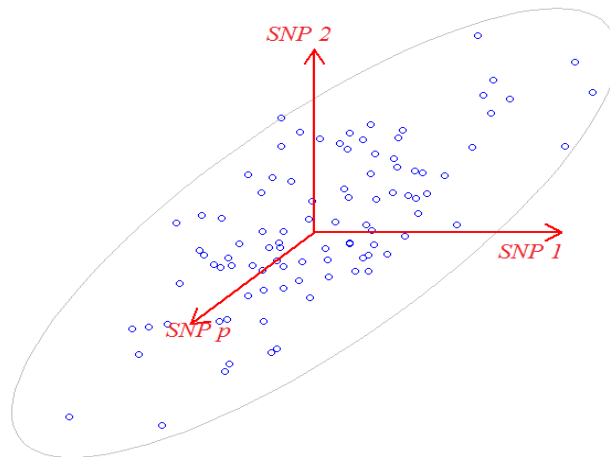


A cloud of p points (variables) in an hyperspace with n dimensions (observations)

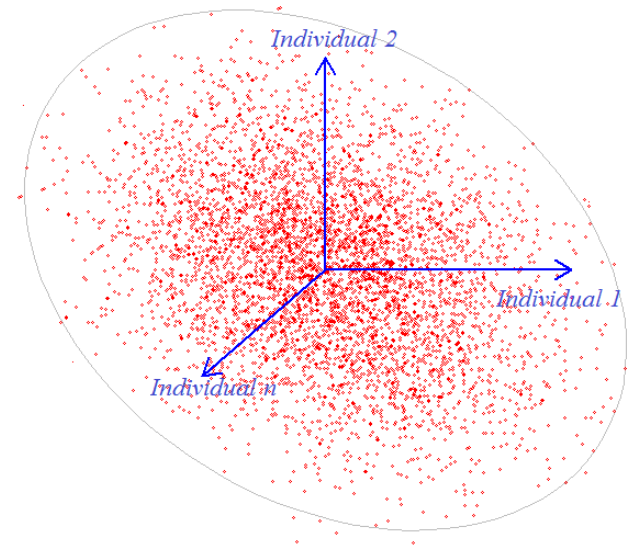
Two points of views

Two geometric representations

HyperSpace of observations
 n points in \mathbb{R}^p



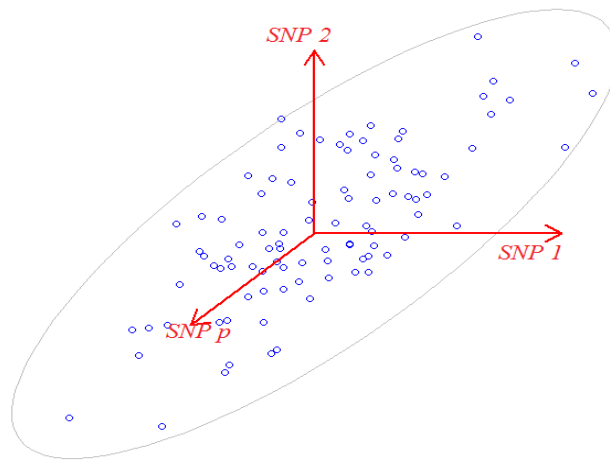
HyperSpace of variables
 p points in \mathbb{R}^n



Two points of views

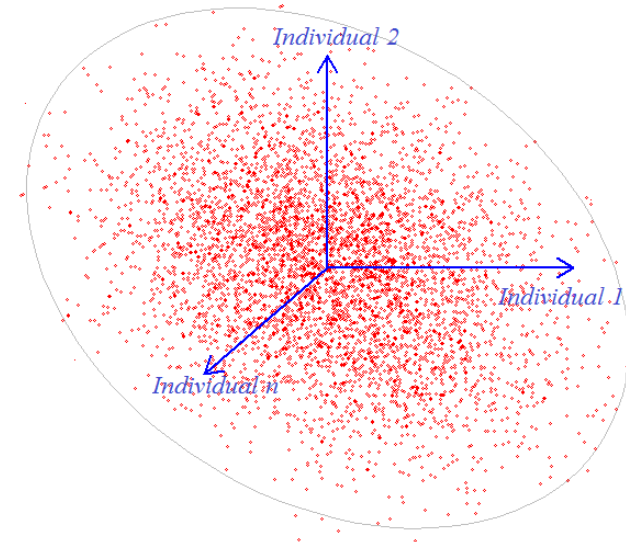
Two geometric representations

HyperSpace of observations
 n points in \mathbb{R}^p



**Similarities / differences
among observations**

HyperSpace of variables
 p points in \mathbb{R}^n

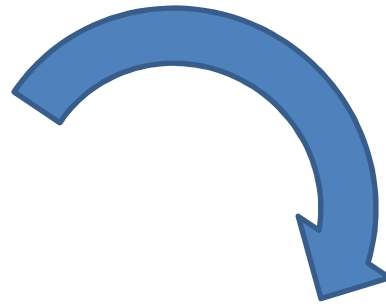
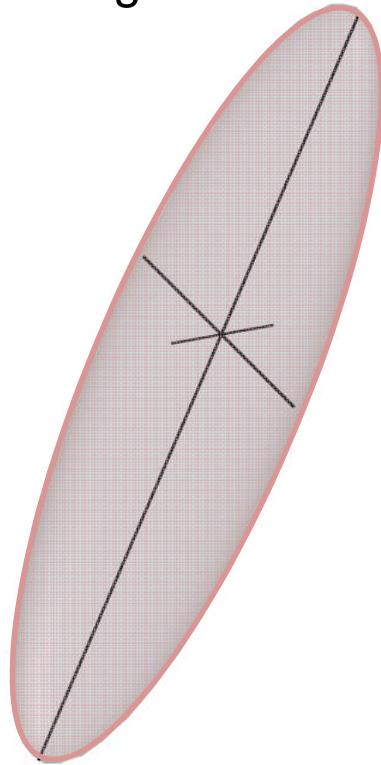


**Relationships among
variables**

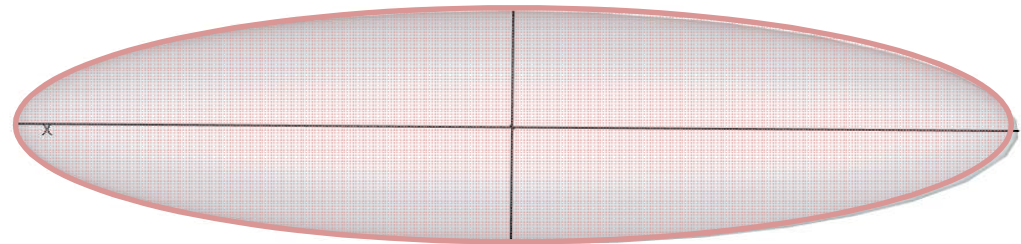
Data transformation

-Rotation and Dimension Reduction

Original data



Transformed data

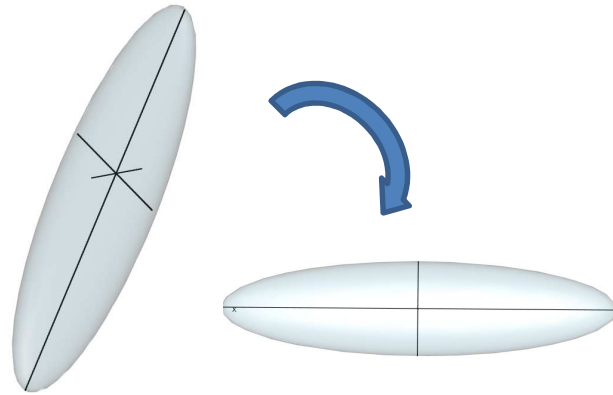


We obtain a small dimension hyperspace where the data representation is as close as possible to the original ones

Data transformation

-Rotation and dimension reduction

Hyperspace of observations
 n points in \mathbb{R}^p



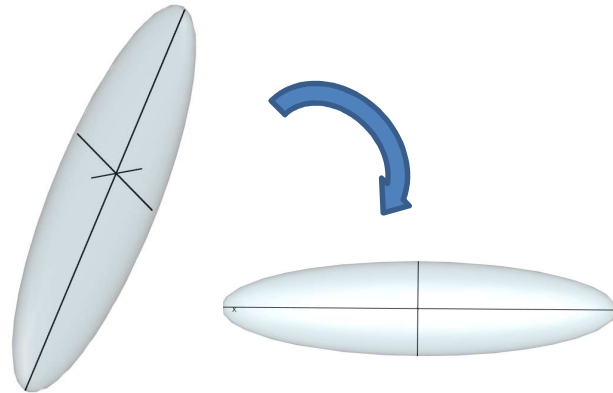
A measure of distance between
observations

The same transformation mechanism

- Rotation

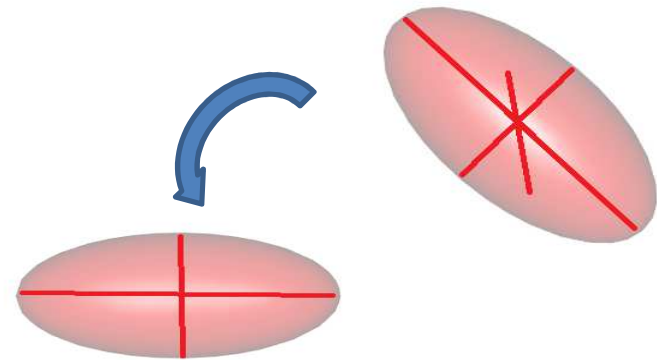
- Dimension reduction

Hyperspace of observations
 n points in \mathbb{R}^p



A measure of distance between
observations

Hyperspace of variables
 p points in \mathbb{R}^n



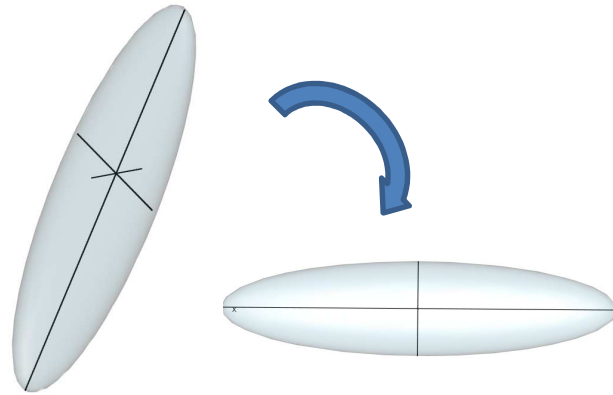
A measure of distance between
variables

The same transformation mechanism

- Rotation

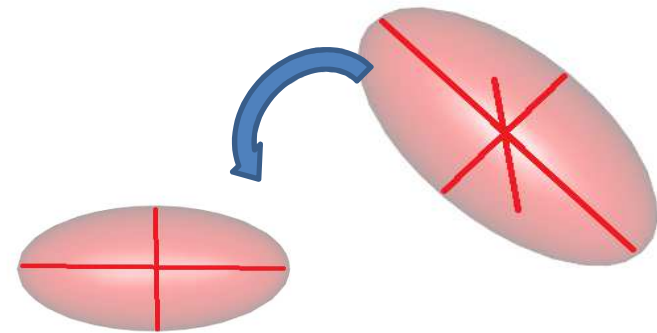
- Dimension reduction

Hyperspace of observations
 n points in \mathbb{R}^p



A measure of distance between
observations

Hyperspace of variables
 p points in \mathbb{R}^n



A measure of distance between
variables

A mathematical formalisation
The duality diagram theory

A mathematical formalisation

The duality diagram

