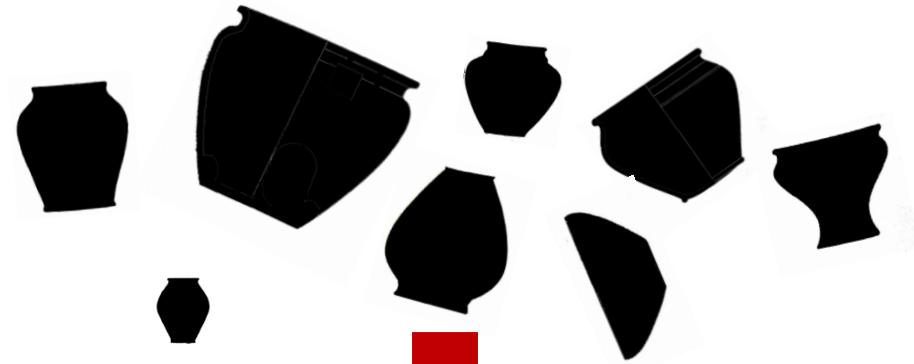


M2 AGES

# **Introduction à la Morphométrie**

Josef WILCZEK



**1) Collecter des données**

**2) Standardisation**

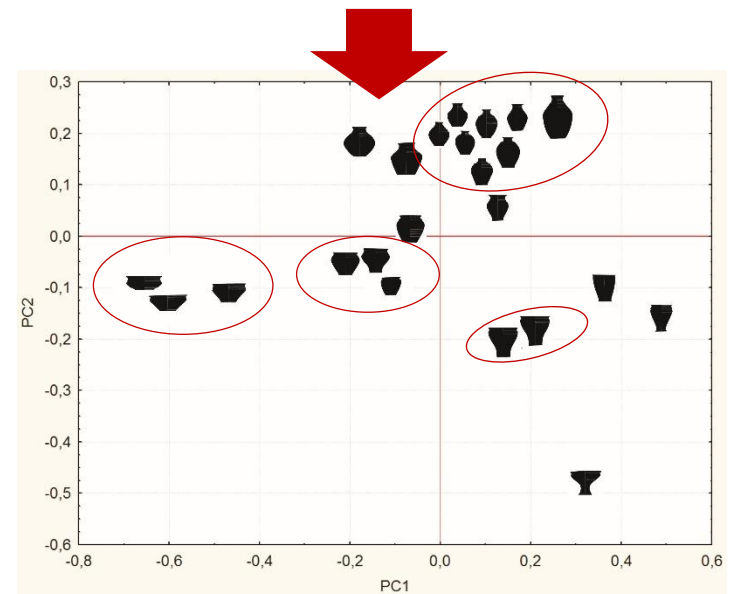
(position, taille et orientation)

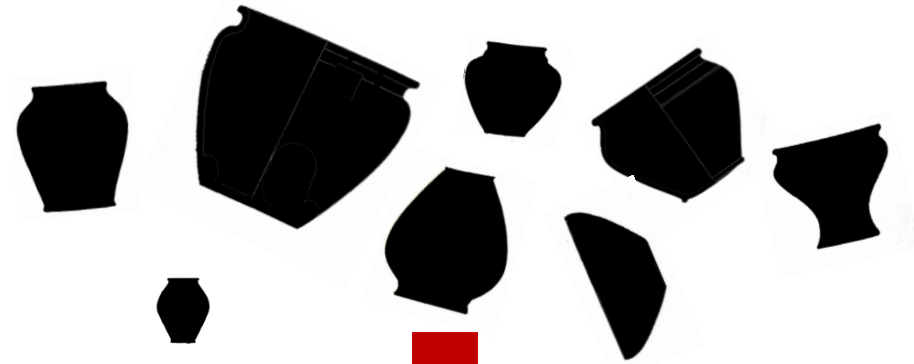


**3) Calcul des variables des formes**

(« shape variables »)

**4) Traitement des données et visualisation**





**1) Collecter des données**

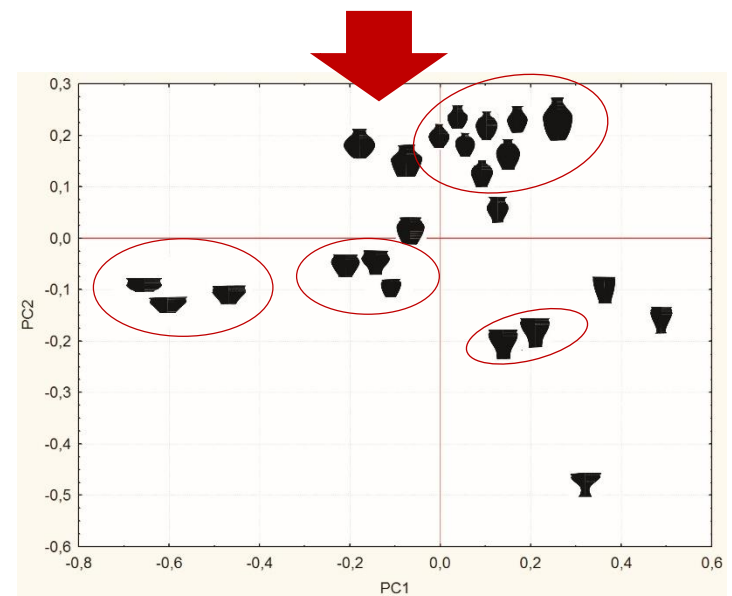
**2) Standardisation**

(position, taille et orientation)

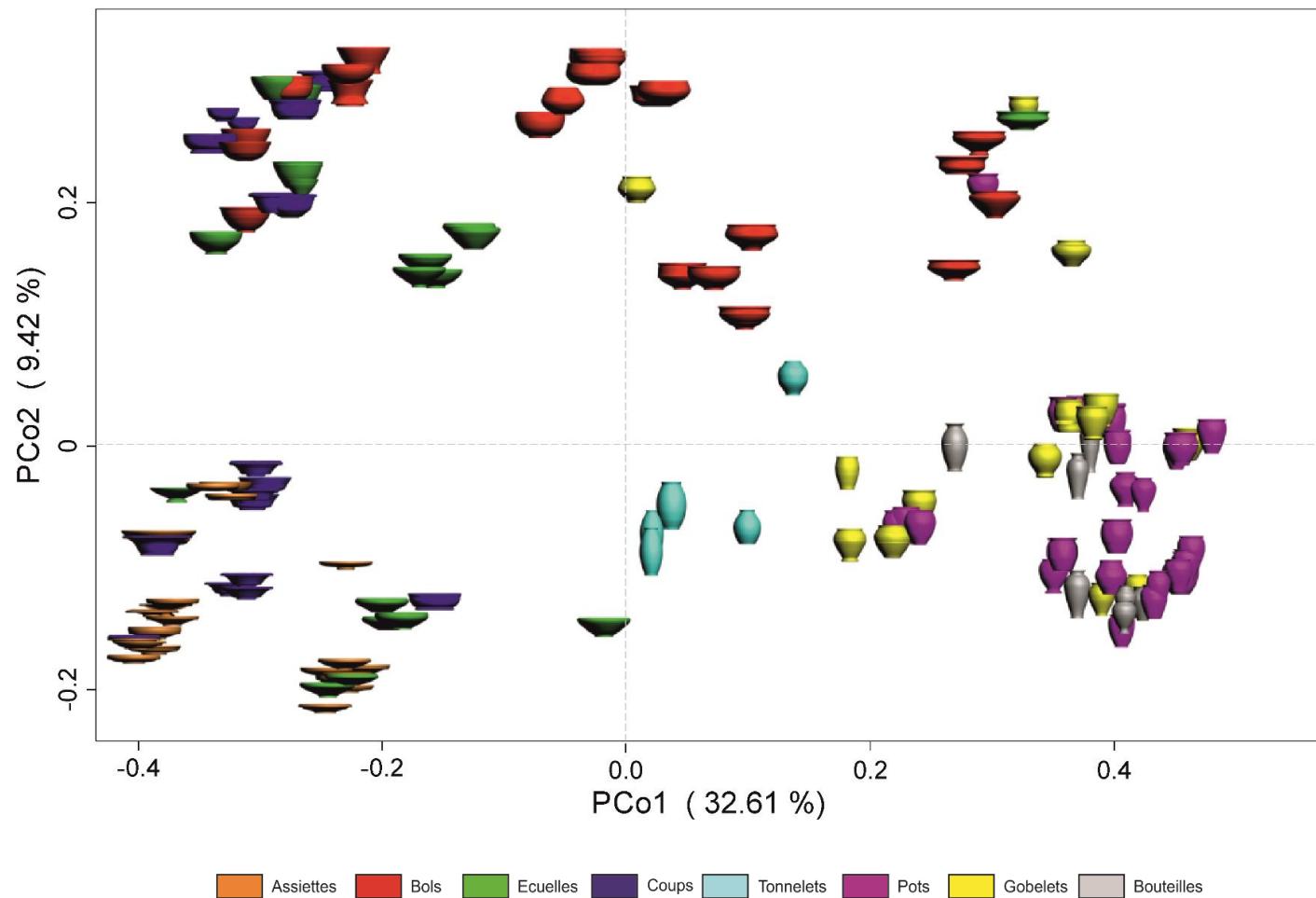
**3) Calcul des variables des formes**

(« shape variables »)

**4) Traitement des données et visualisation**



On peut visualiser l'assemblage dans un graphique (MORPHOSPACE)

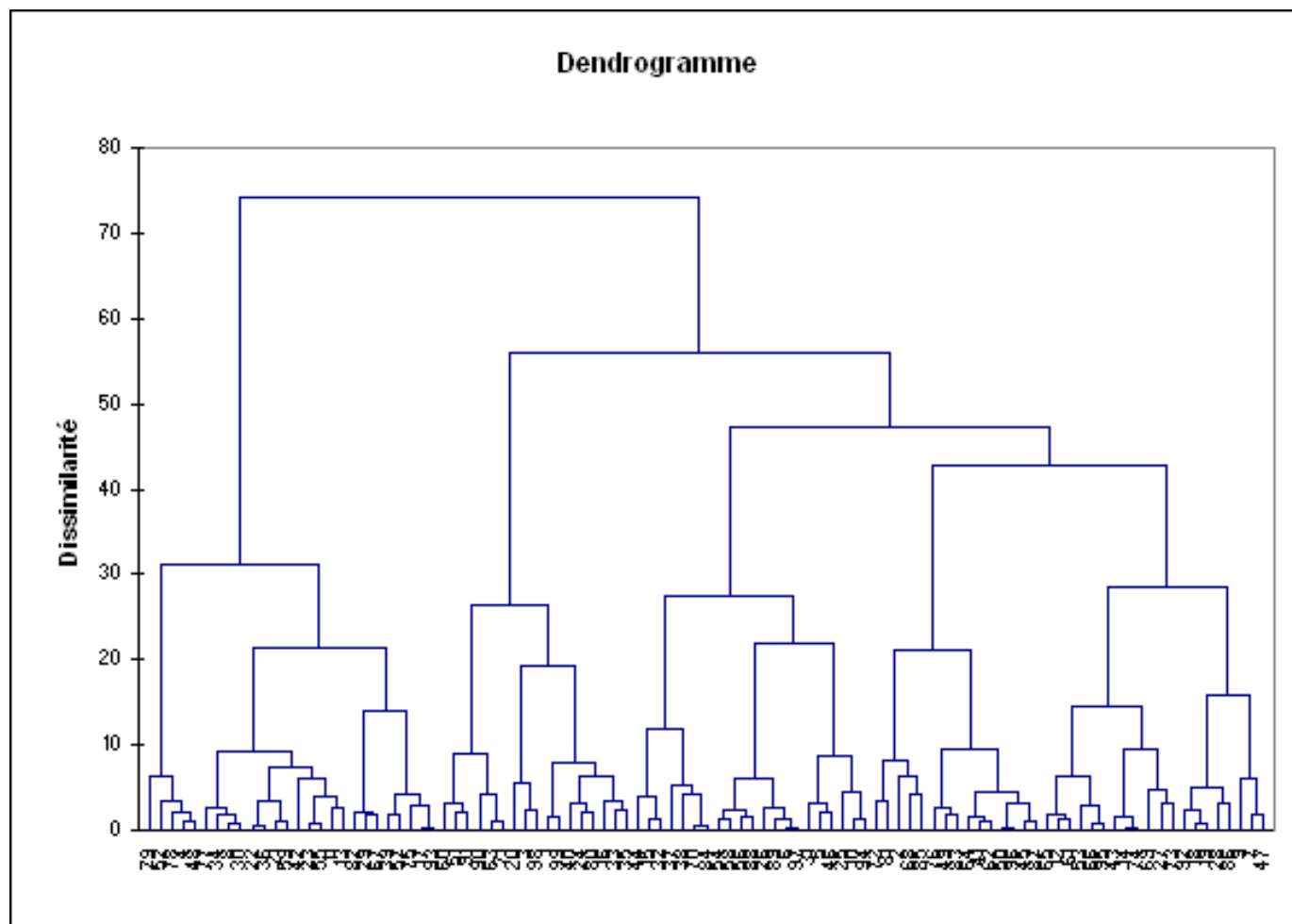


**On peut visualiser l'assemblage dans un graphique (MORPHOSPACE)**

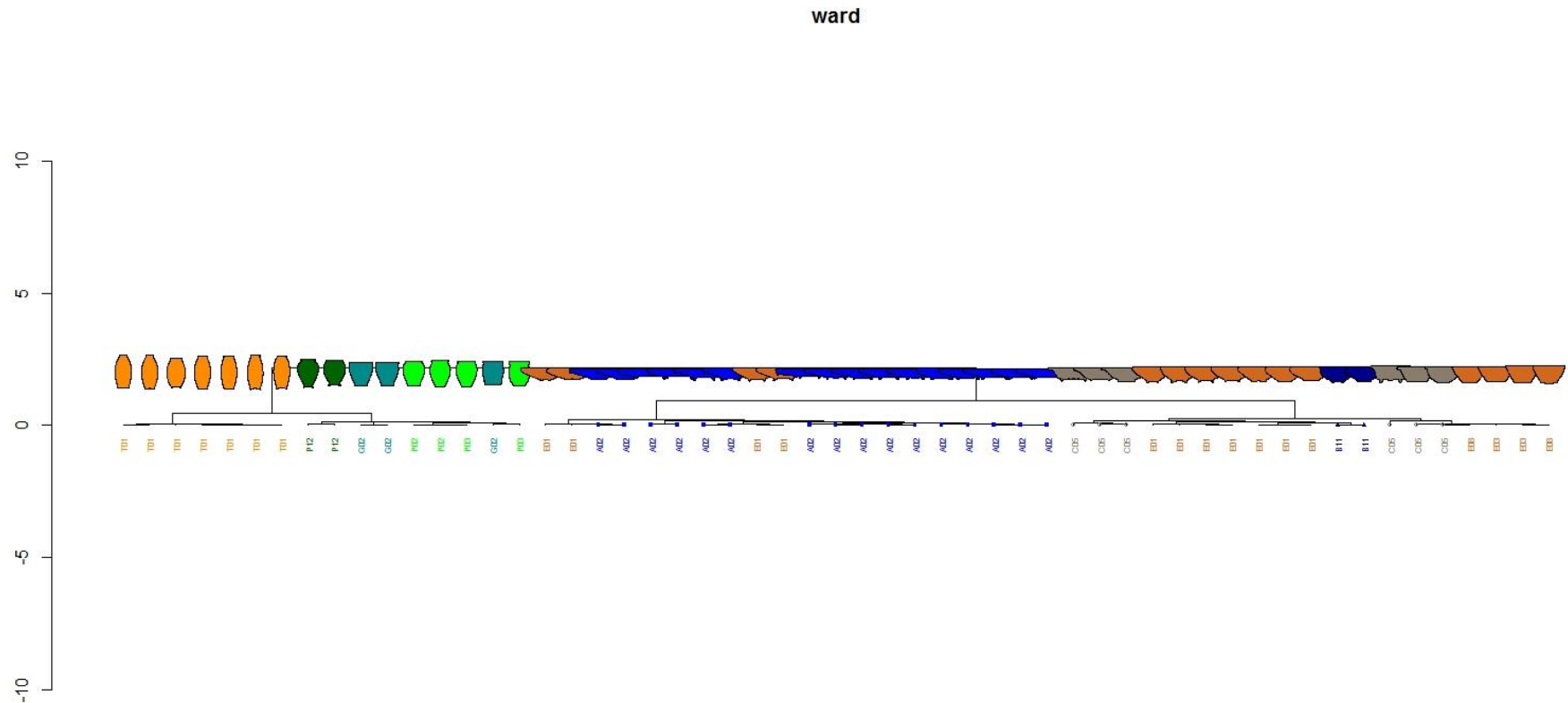
3D

 Assiettes  Bols  Ecuelles  Coups  Tonnelets  Pots  Gobelets  Bouteilles

On peut créer la classification (CA)



On peut créer la classification (CA)



**On peut créer la classification (MBCA)**





**On peut créer la classification (MBCA)**



**On peut créer la classification (MBCA)**



**On peut créer la classification (MBCA)**



**On peut créer la classification (MBCA)**



**On peut créer la classification (MBCA)**



## On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6


**On peut créer la classification (MBCA)**



G1	G2	G3	G4	G5	G6

## On peut créer la classification (MBCA)





G1	G2	G3	G4	G5	G6
		 <p>POI</p>			






## On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
		 POI		 NEWS	





## On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
		 POI	 VSR2	 NEPS	






On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
 SR		 POI	 VSR2	 NEVS	







## On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
  94T 76T		 POI	 VSR2	 HEV3	








## On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
 9F  9BT		 POI	 VSR2	 NEY3	 VIT1





































## On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
 <p>SF</p>  <p>VIT</p>	 <p>VIT</p>	 <p>POI</p>	 <p>VSR2</p>	 <p>NEV3</p>	 <p>VIT1</p>





































On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
					
SA*	FA*	POI	VSR2	NEV3	VIT1
					
SM*	LO*		VSR1		VSA
					
SM*	CS*	PL23	VSR	NEQ	ELS
					
SE*		PCNE	VSR		ERP1
					
SM*		NSL	VSR2	NE1	
					
ROT	ME	POI	VSR	NEV2	ERP2

On peut créer la classification (MBCA)



G1	G2	G3	G4	G5	G6
					
SA*	FA*	POI	VSR2	NEV3	VIT1
					
SM*	LO*		VSR1		VSA
					
SM*	CS*	PL23	VSR	NEQ	ELS
					
SE		PCNE	VSR		ERP1
					
SM*		NSL	VSR2	NE1	
					
ROT	ME	POI	VSR	NEV2	ERP2

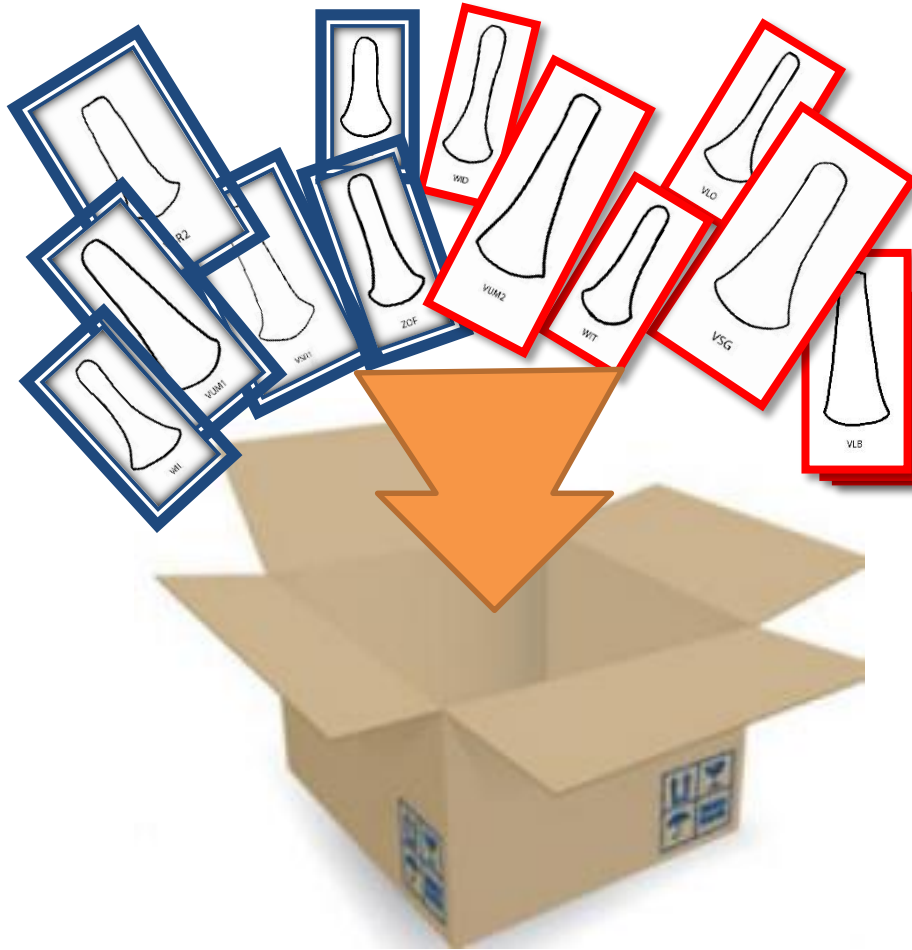


**On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)**

On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)



On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)



**On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)**

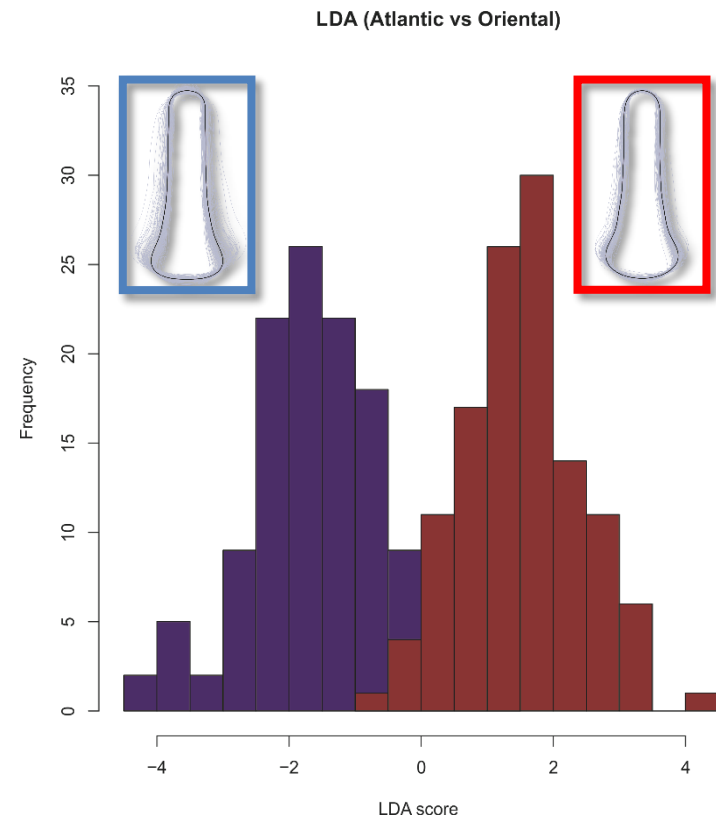


**On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)**



## 4) Traitement des donn  s

On peut voir s'il y a des diff  rences entre les groupes (DA, MBDA, MANOVA)

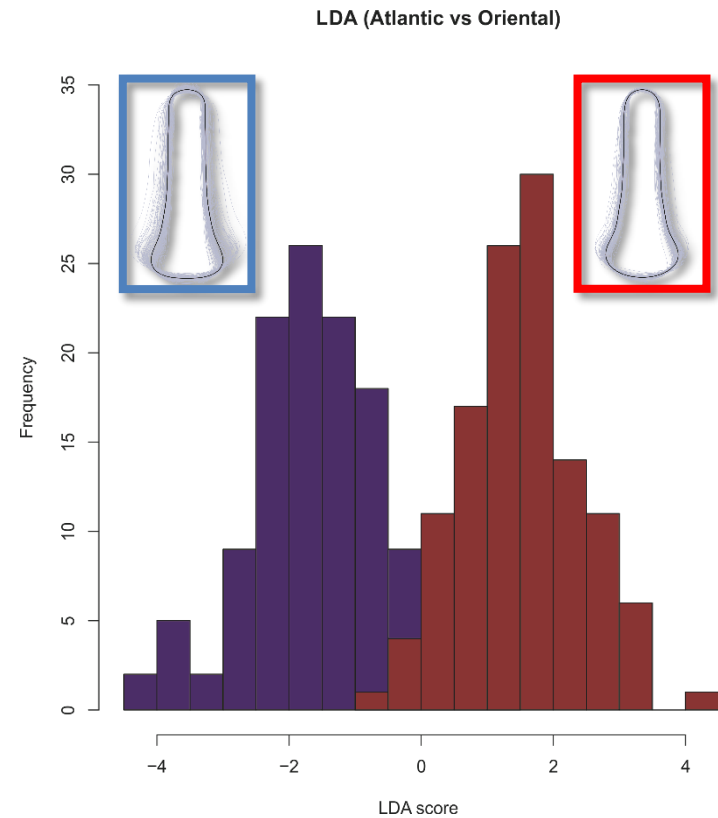


```
          Df Hotelling-Lawley approx F num Df den Df      Pr(>F)
GR2.groups  1             2.3211  11.605   41   205 < 2.2e-16 ***
Residuals 245
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 4) Traitement des données

On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

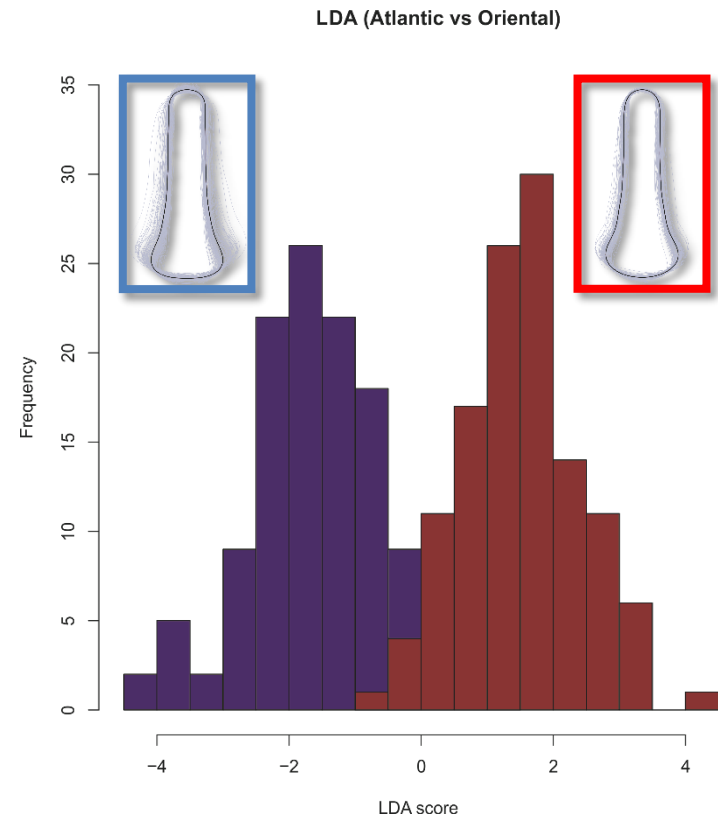
On peut prédire l'appartenance d'un individu



## 4) Traitement des données

On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

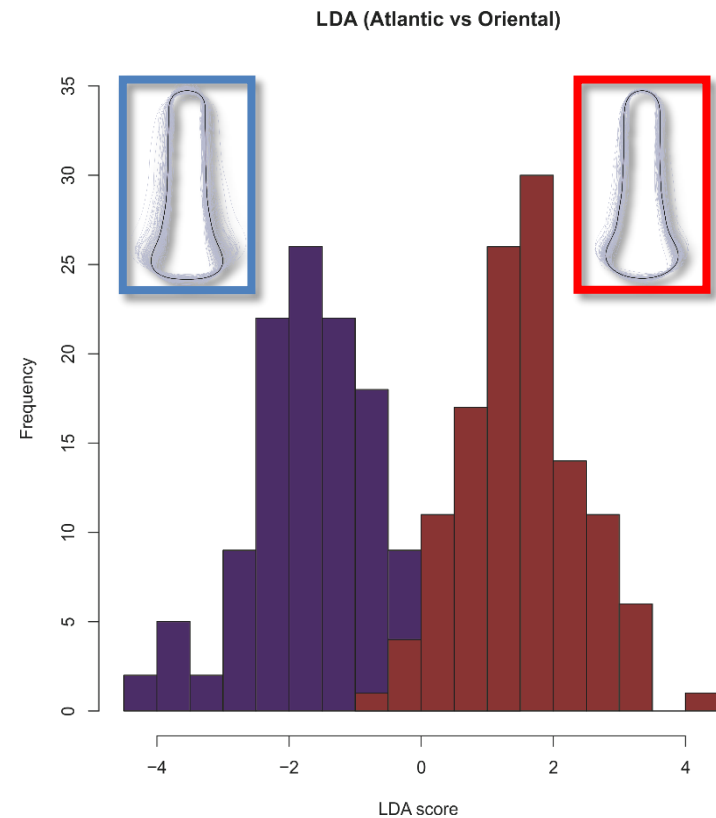
On peut prédire l'appartenance d'un individu





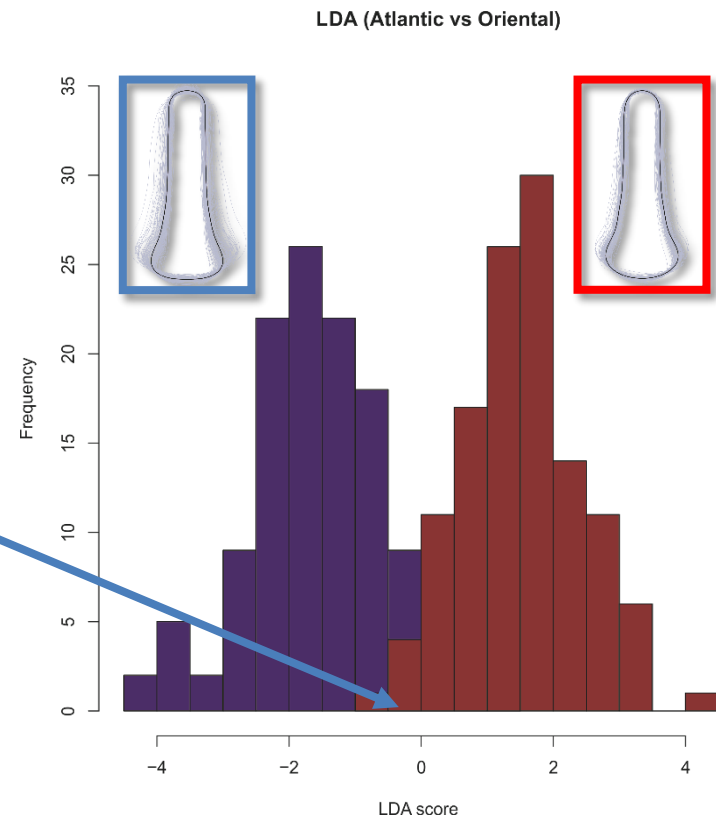
On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

On peut prédire l'appartenance d'un individu



On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

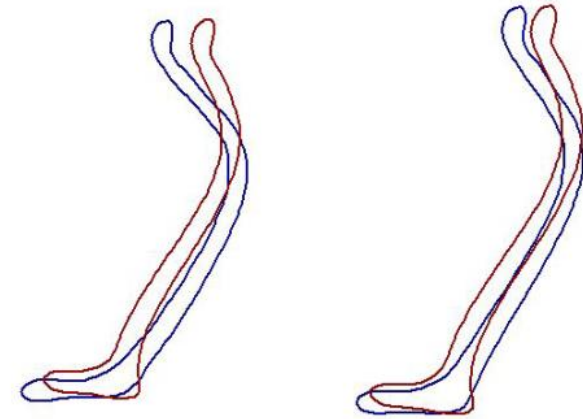
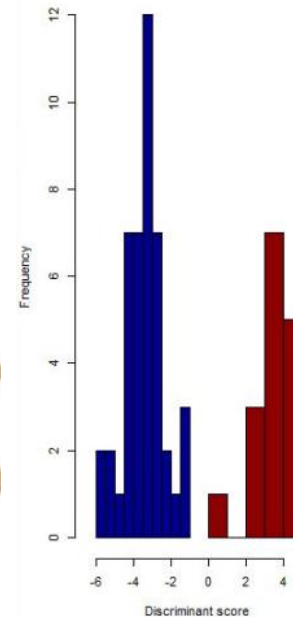
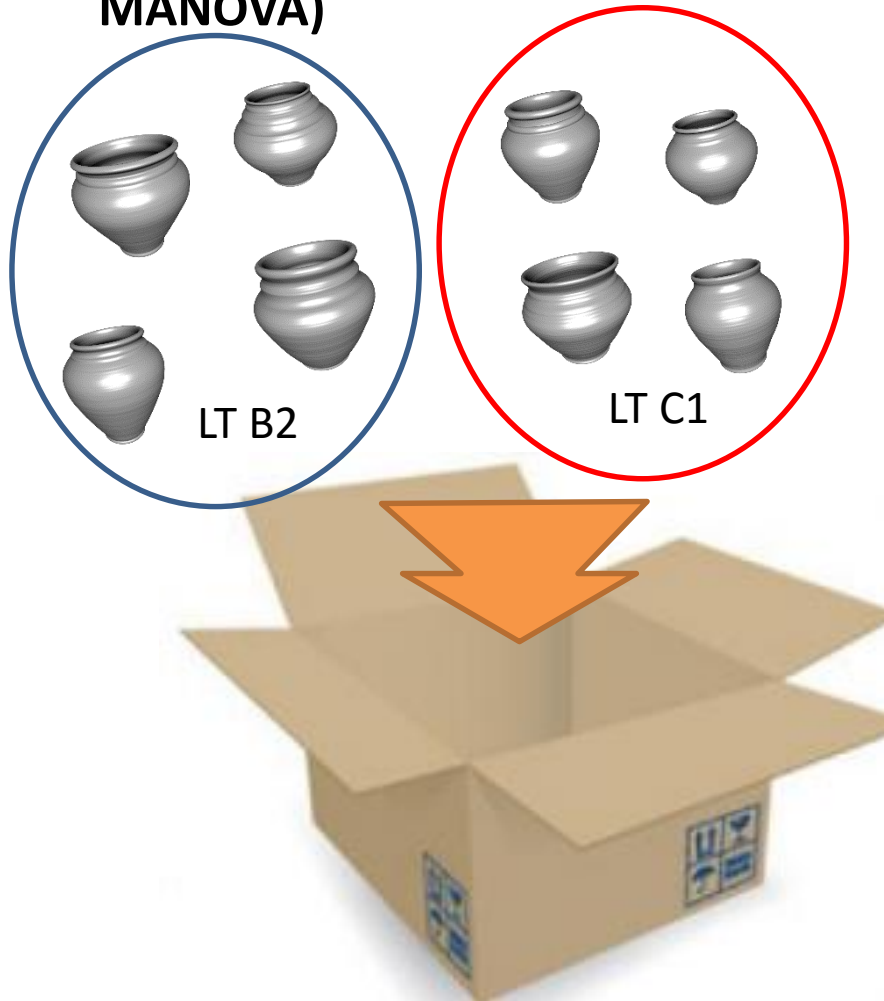
On peut prédire l'appartenance d'un individu



## 4) Traitement des données

On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

L'autre exemple...



On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)  
Ca marche avec plus que deux groupes...



#### 4) Traitement des données

On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

Ca marche avec plus que deux groupes...



On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

Ca marche avec plus que deux groupes...



On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)  
**Ca marche avec plus que deux groupes...**



Indiv	G1	G2	G3	G4	G5	G6
1	0.9	0.1	0	0	0	0
2	0.8	0.0	0.2	0	0	0
3	0	0	0.2	0	0.1	0.7
4	0.5	0.3	0.1	0.1	0	0
5	0.7	0.1	0	0.1	0	0.1
...	...	...	...	...	...	...
N	0					

On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

Ca marche avec plus que deux groupes...

On peut prédire l'appartenance d'un individu



Indiv	G1	G2	G3	G4	G5	G6
1	0.9	0.1	0	0	0	0
2	0.8	0.0	0.2	0	0	0
3	0	0	0.2	0	0.1	0.7
4	0.5	0.3	0.1	0.1	0	0
5	0.7	0.1	0	0.1	0	0.1
...	...	...	...	...	...	...
N	0					



On peut voir s'il y a des différences entre les groupes (DA, MBDA, MANOVA)

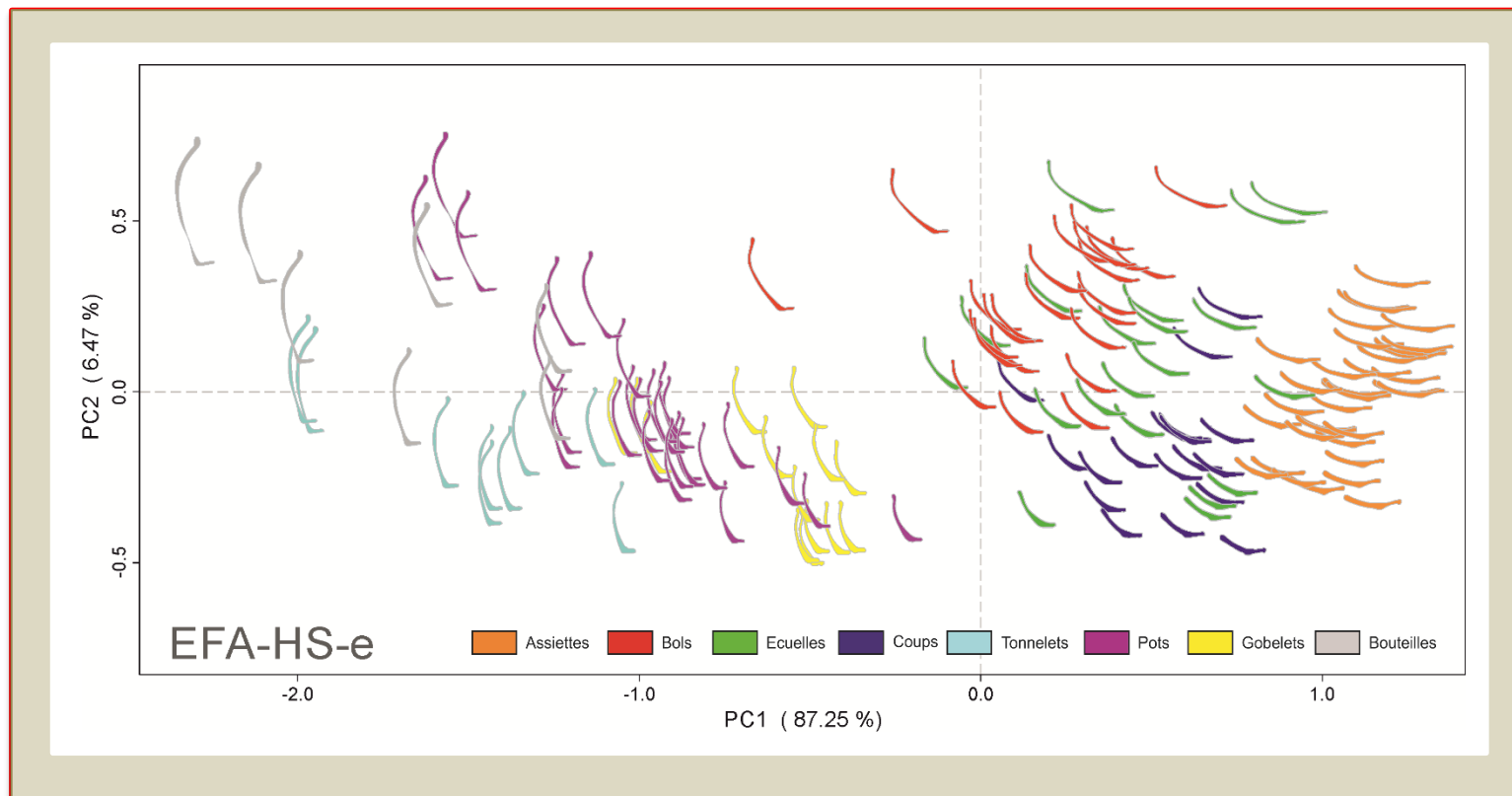
Ca marche avec plus que deux groupes...

On peut prédire l'appartenance d'un individu



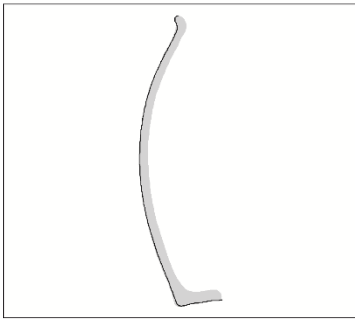
Indiv	G1	G2	G3	G4	G5	G6
1	0.9	0.1	0	0	0	0
2	0.8	0.0	0.2	0	0	0
3	0	0	0.2	0	0.1	0.7
4	0.5	0.3	0.1	0.1	0	0
5	0.7	0.1	0	0.1	0	0.1
...	...	...	...	...	...	...
New	0.1	0.1	0.8	0	0	0

On peut visualiser les formes moyennes des groupes

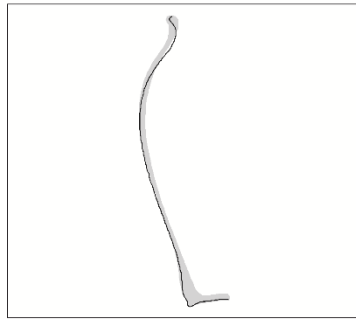


On peut visualiser les formes moyennes des groupes

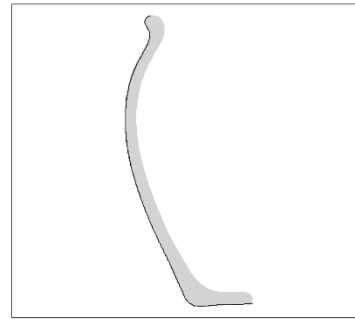
Meanshape of *Tonnelets*



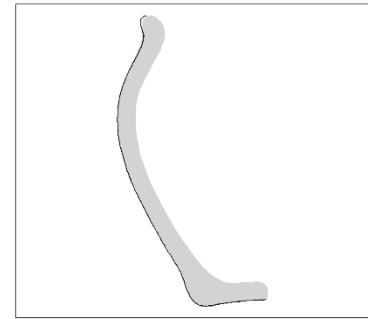
Meanshape of *Bouteilles*



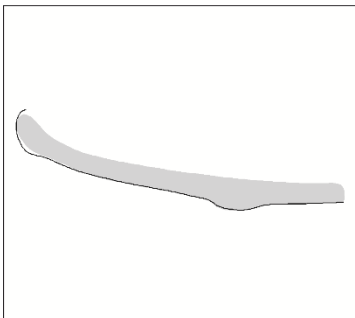
Meanshape of *Pots*



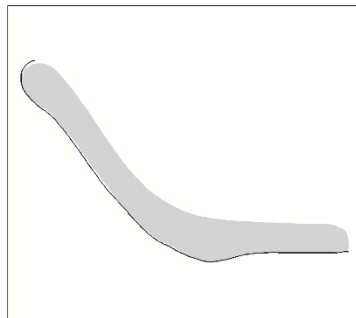
Meanshape of *Gobelets*



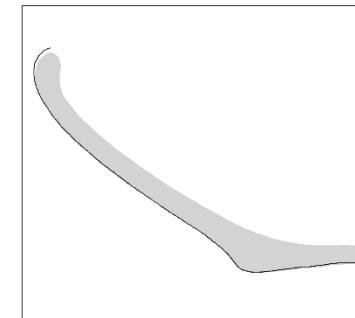
Meanshape of *Assiettes*



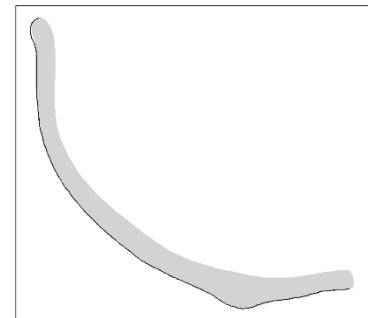
Meanshape of *Coups*



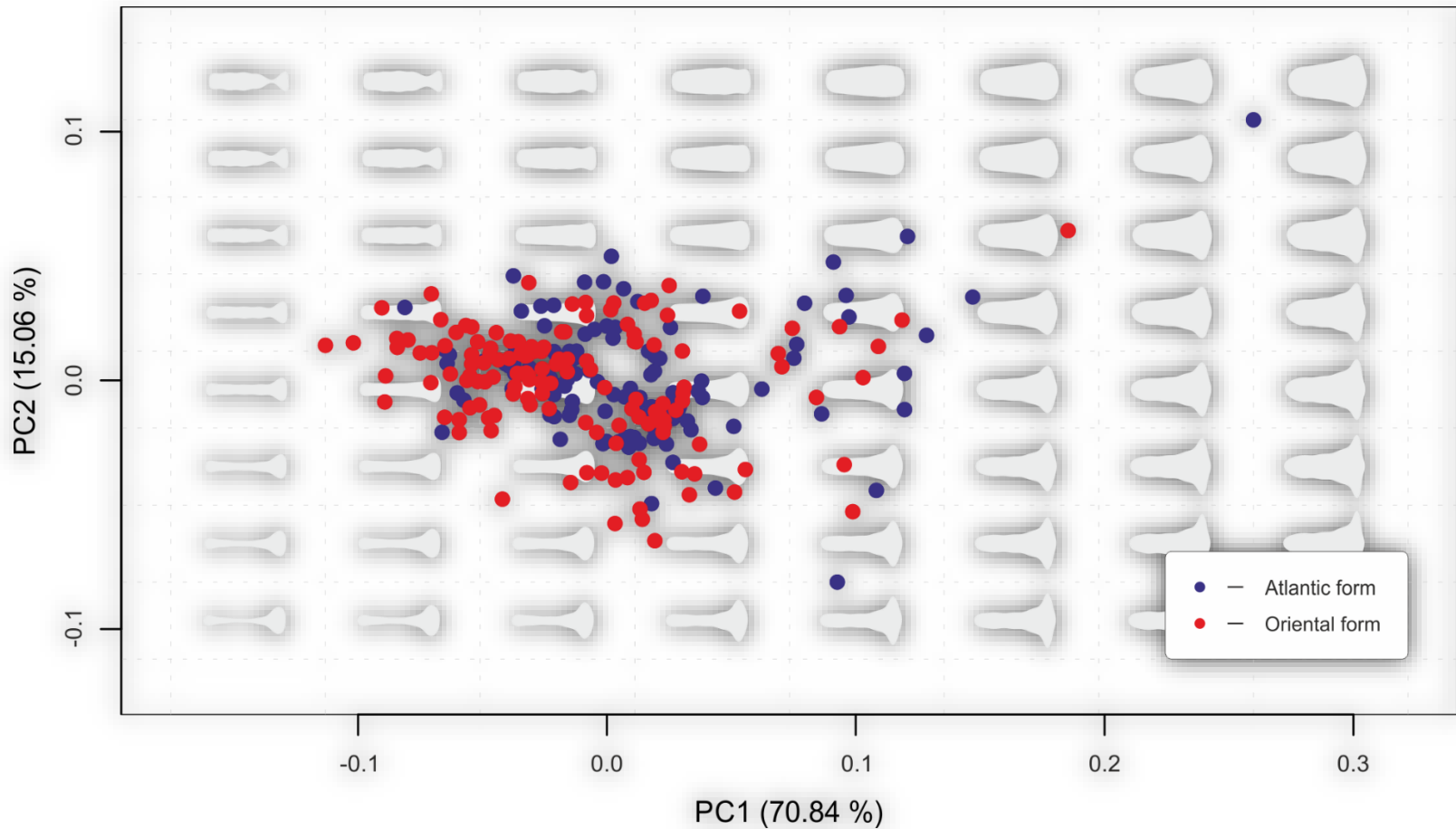
Meanshape of *Ecuelles*



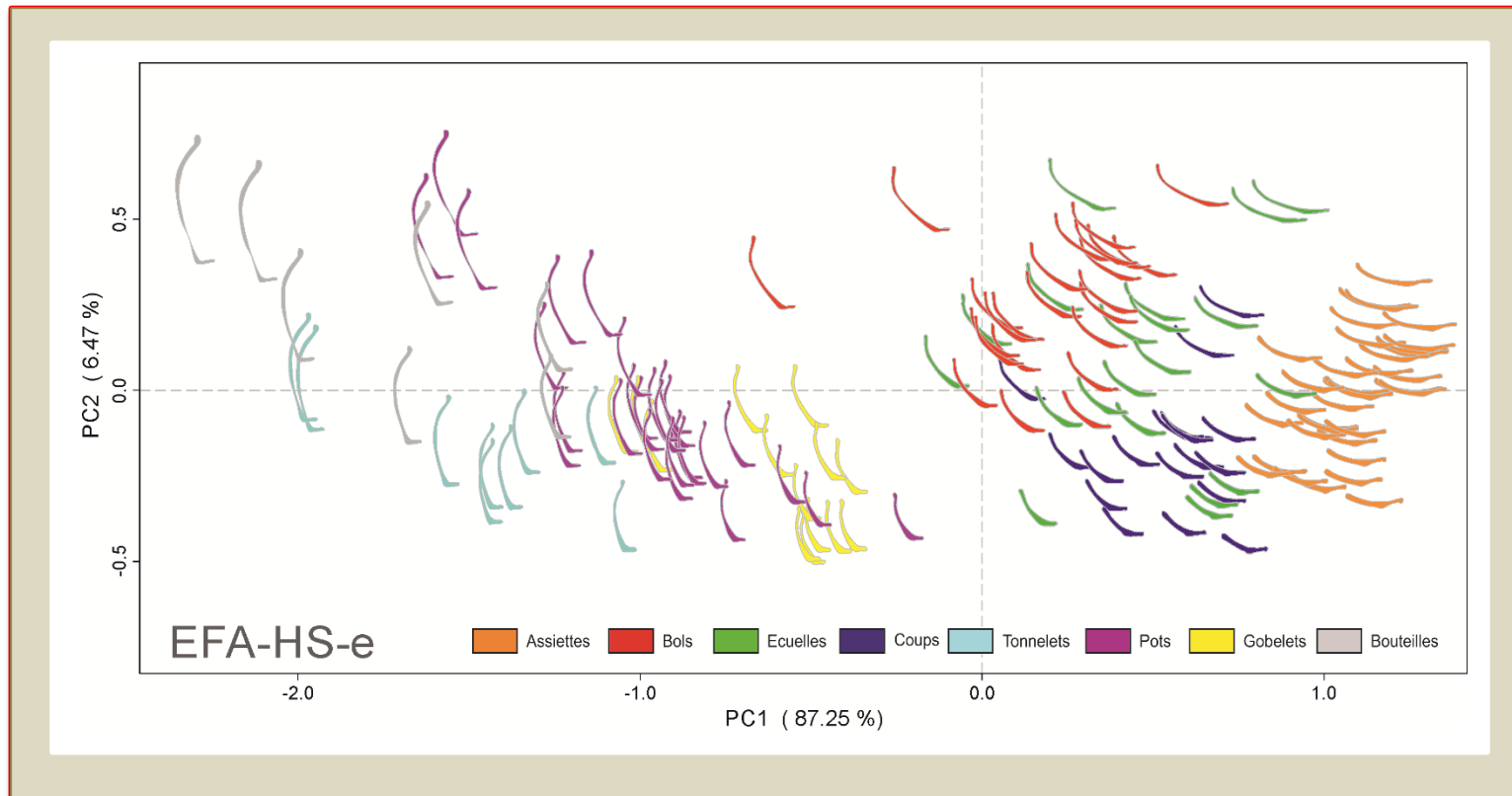
Meanshape of *Bols*



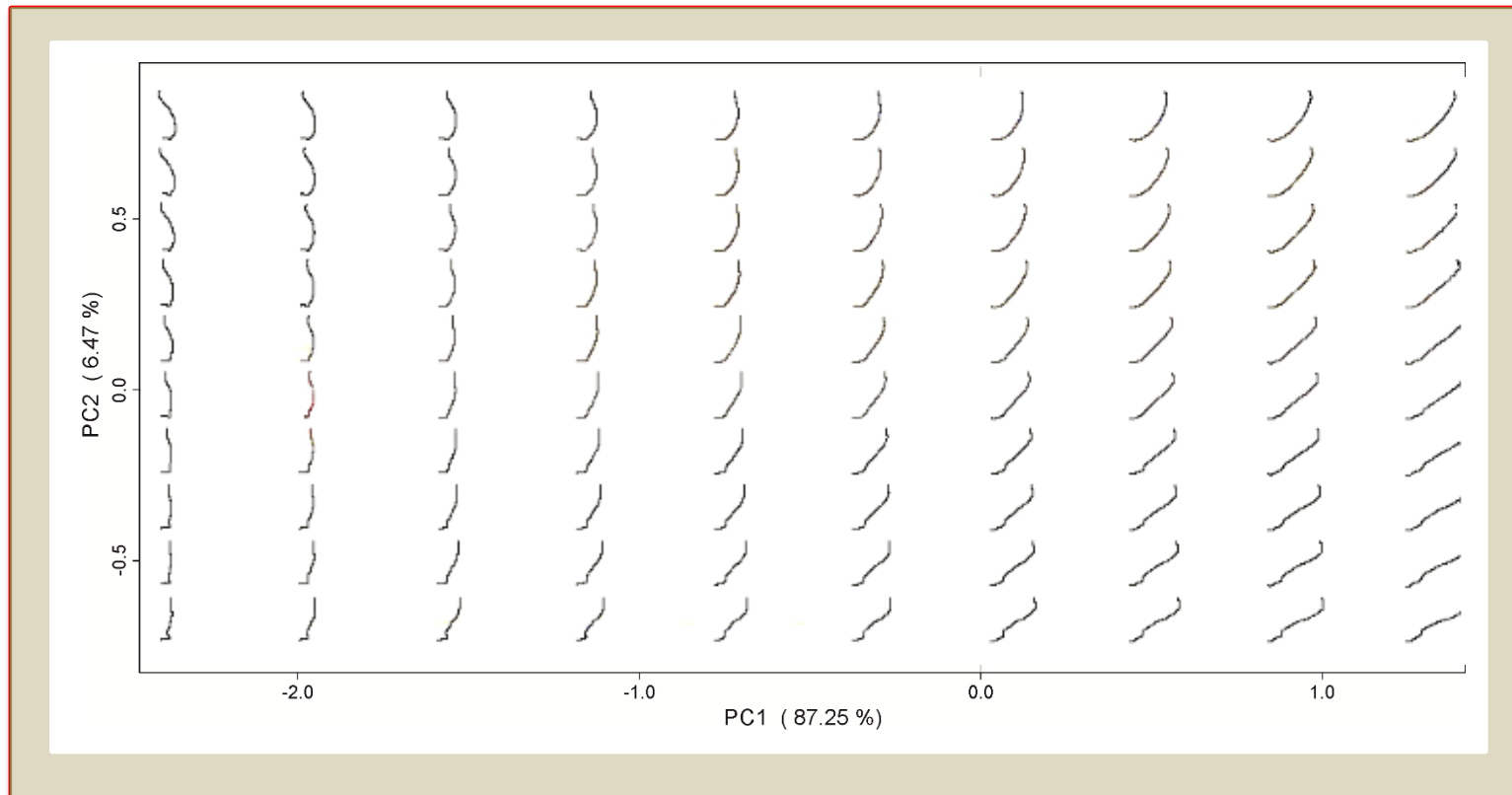
On peut visualiser n'importe quel individu dans le morphospace



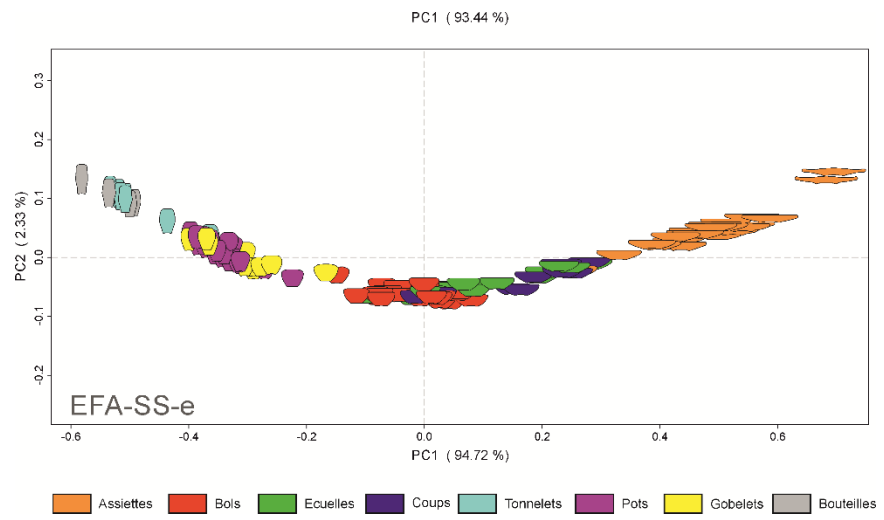
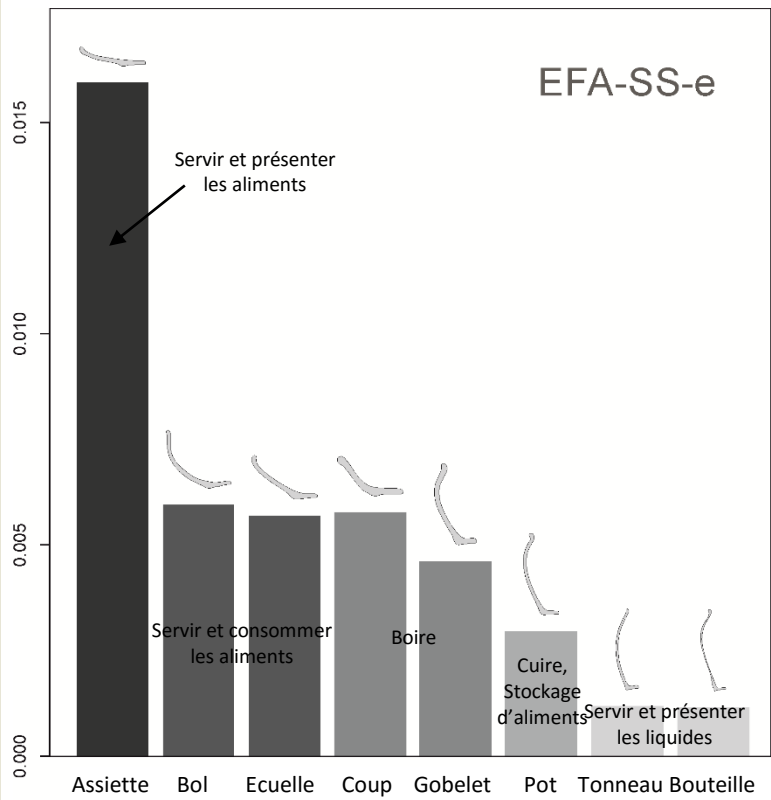
On peut visualiser n'importe quel individu dans le morphospace



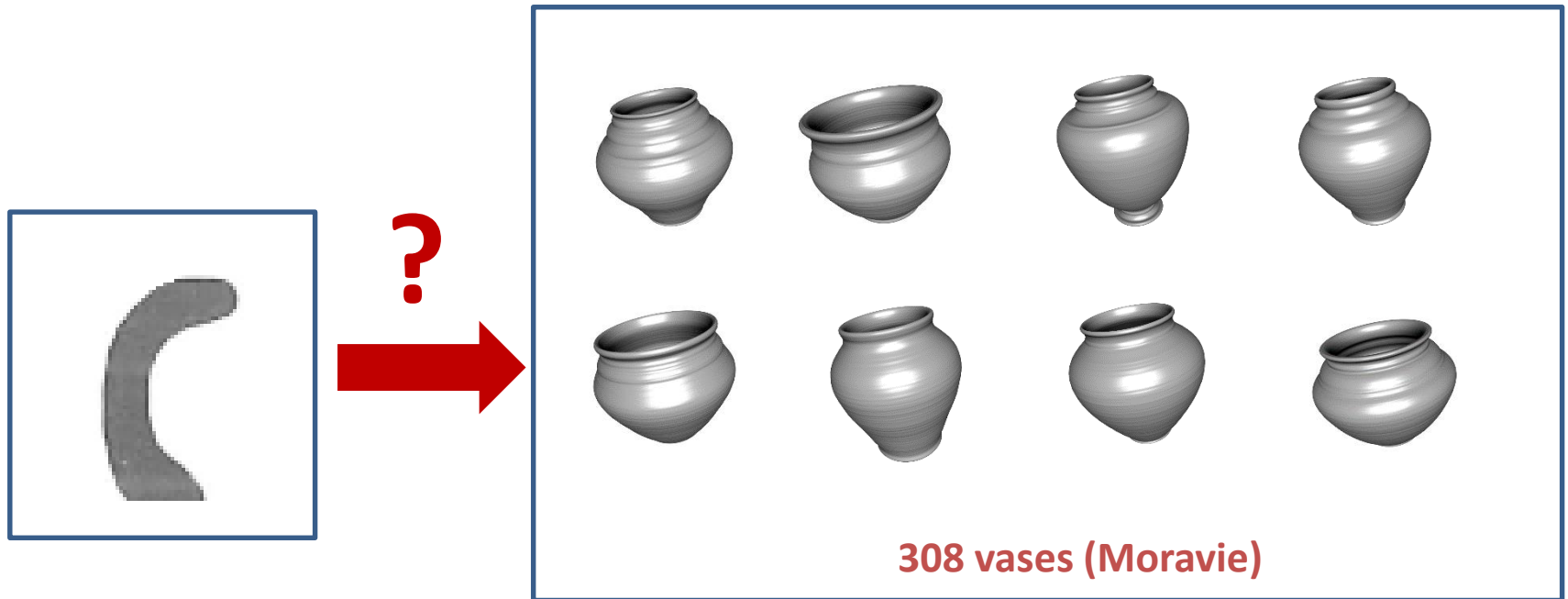
**On peut visualiser n'importe quel individu dans le morphospace**



On peut estimer la diversité des formes (niveau de la standardisation des production)

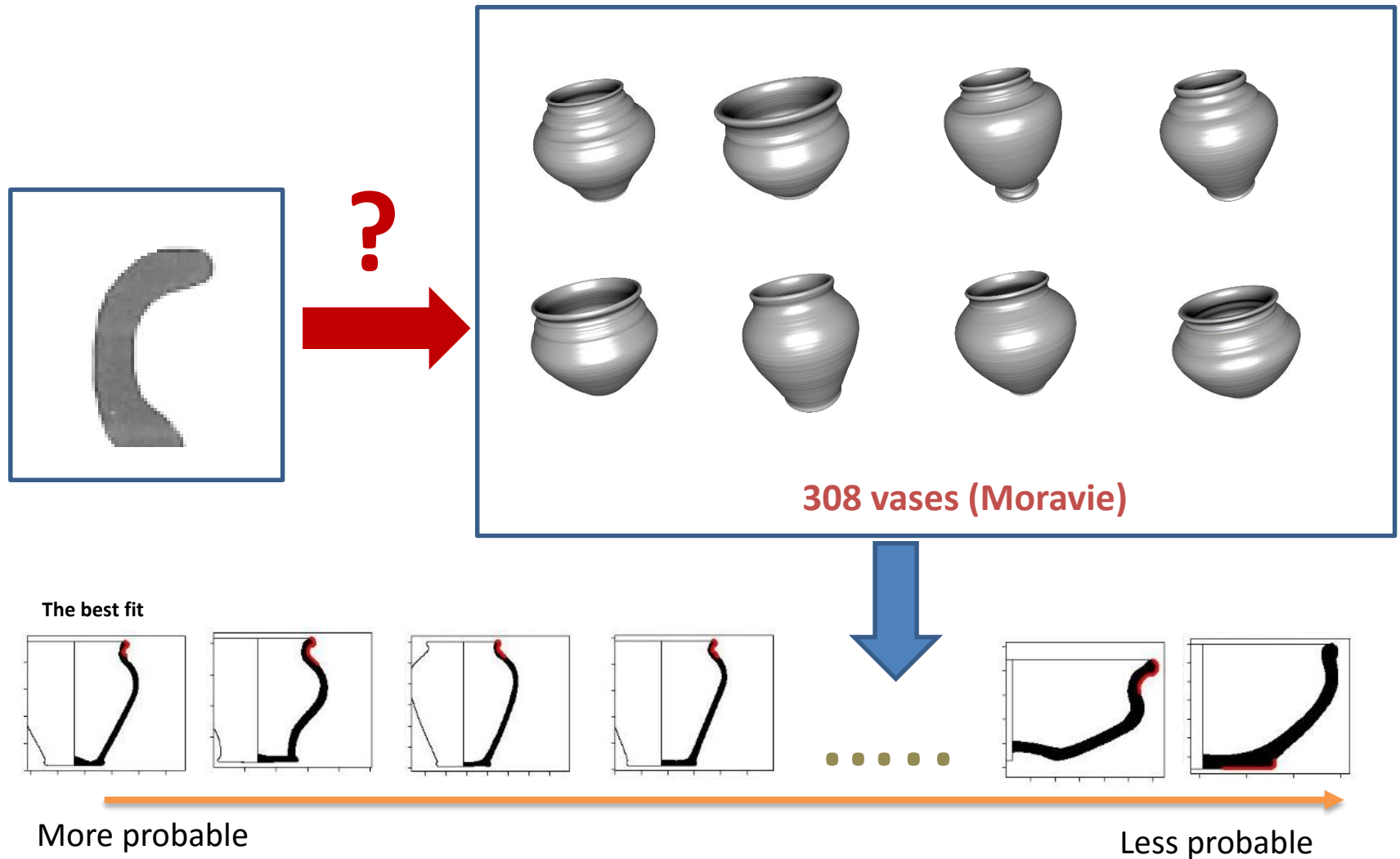


On peut prédire les formes complètes a partir d'un tesson





On peut prédire les formes complètes a partir d'un tesson



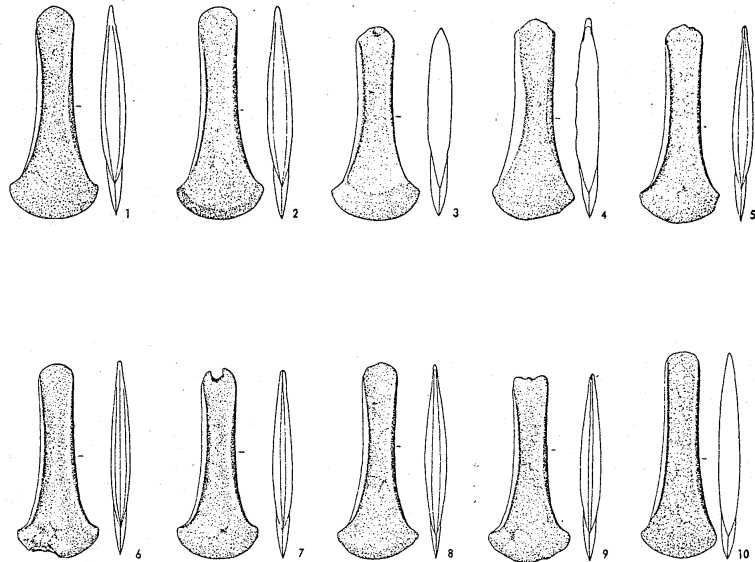
# Unsupervised models for typological classification of archaeological artefacts

Wilczek<sup>1,2</sup>, Monna<sup>1</sup>, Gabillot<sup>1</sup>, Navarro<sup>3</sup>, Rusch<sup>1</sup>, Chateau<sup>4</sup>

1. ArTéHiS, UMR 6298 CNRS-Université de Bourgogne
2. ÚAM FF, Masarykova univerzita
3. Laboratoire PALEVO, UMR uB/CNRS 6282-Biogéosciences, Université de Bourgogne
4. Université de Bourgogne, UFR SVTE

# Unsupervised models for typological classification

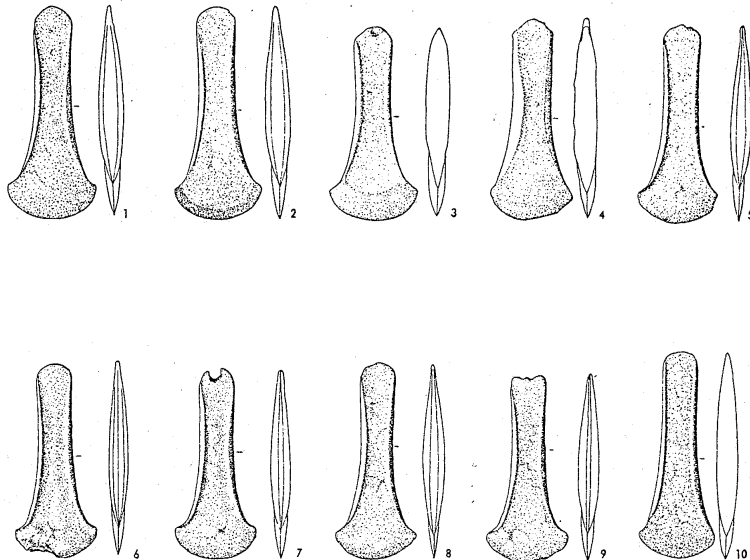
## Bronze Age axes (1650-1350 BC)



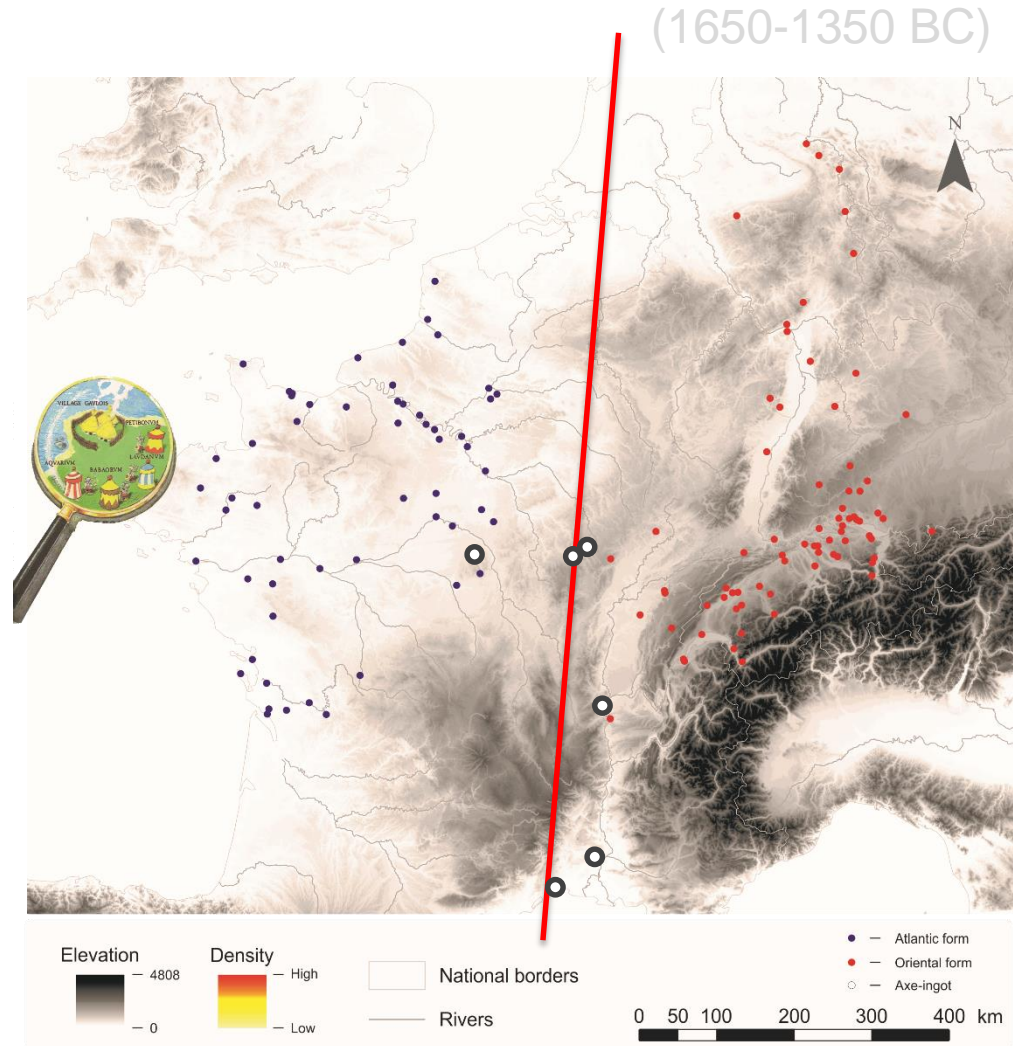
# Unsupervised models for typological classification

## Checking the current typology of flanged axes

- Specific types (Neyruz, Salez, Möhlin, ...)
- Atlantic vs Oriental type, „Axes-ingots“
- Geometric morphometry (EFA)



Hache a rebords



# Unsupervised models for typological classification

## Choice of corpus

- Entirely preserved
- Not much used
- A-priori classified (Atlantic/Oriental type, Axe-ingot)

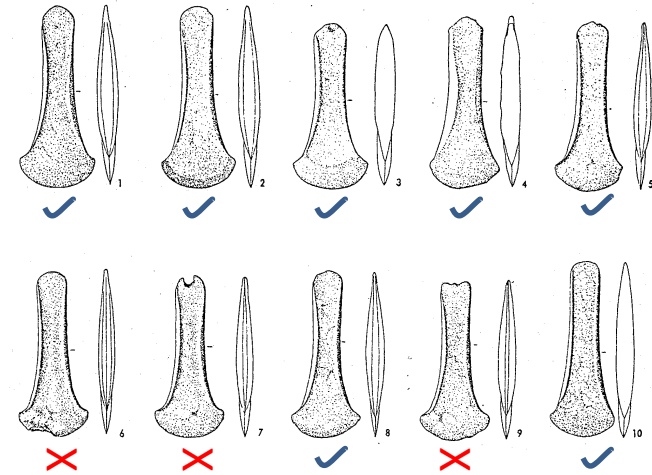
## Studied corpus

- 286 axes from 132 sites
- 121 oriental forms
- 126 occidental forms
- 21 axes-ingots

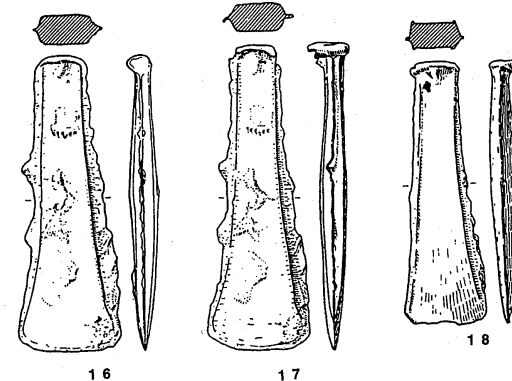
## Axes-ingots

- Similar to flanged axes
- Residual parts on borders

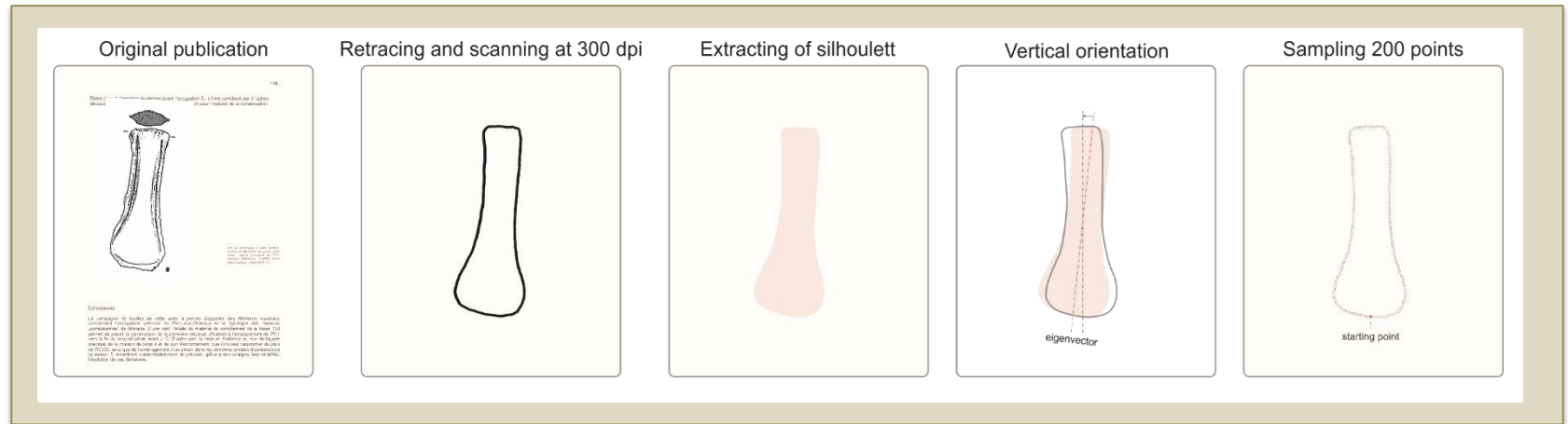
Flanged axes



Axes-ingots



# Unsupervised models for typological classification



## Elliptic Fourier Analysis (EFA)

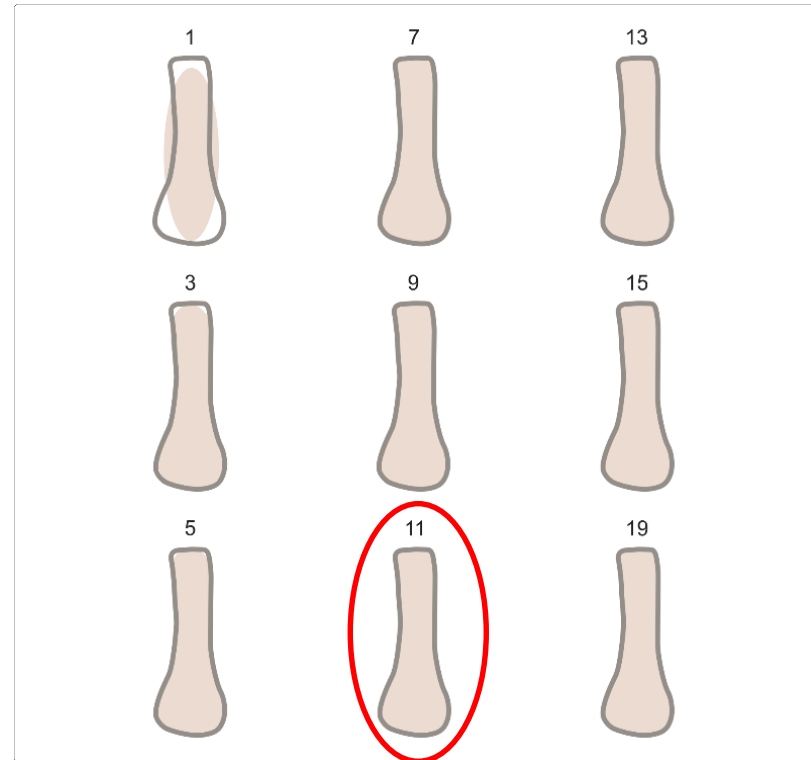
$$x(t) = a_0 + \sum_{j=1}^k a_j \cos\left(\frac{2j\pi t}{T}\right) + \sum_{j=1}^k b_j \sin\left(\frac{2j\pi t}{T}\right)$$

$$y(t) = c_0 + \sum_{j=1}^k c_j \cos\left(\frac{2j\pi t}{T}\right) + \sum_{j=1}^k d_j \sin\left(\frac{2j\pi t}{T}\right)$$

# Unsupervised models for typological classification

## Determination of harmonic number

- Visualisation of inverse Fourier transformation

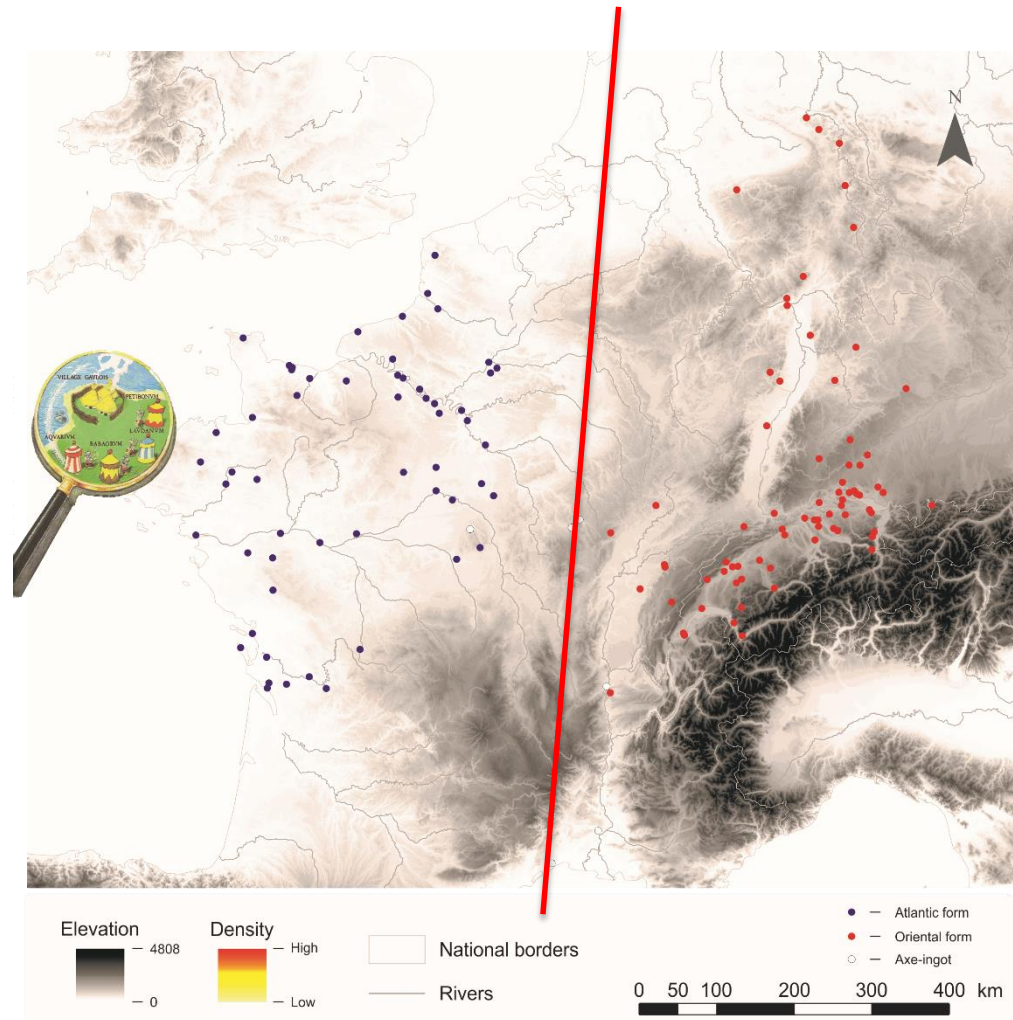




# Unsupervised models for typological classification

## Checking the current typology

- **Oriental vs. Atlantic forms**
- **Principal Component Analysis (PCA)**
- **Multidimensional Analysis of Variance (MANOVA)**
- **Linear Discriminant Analysis (LDA)**

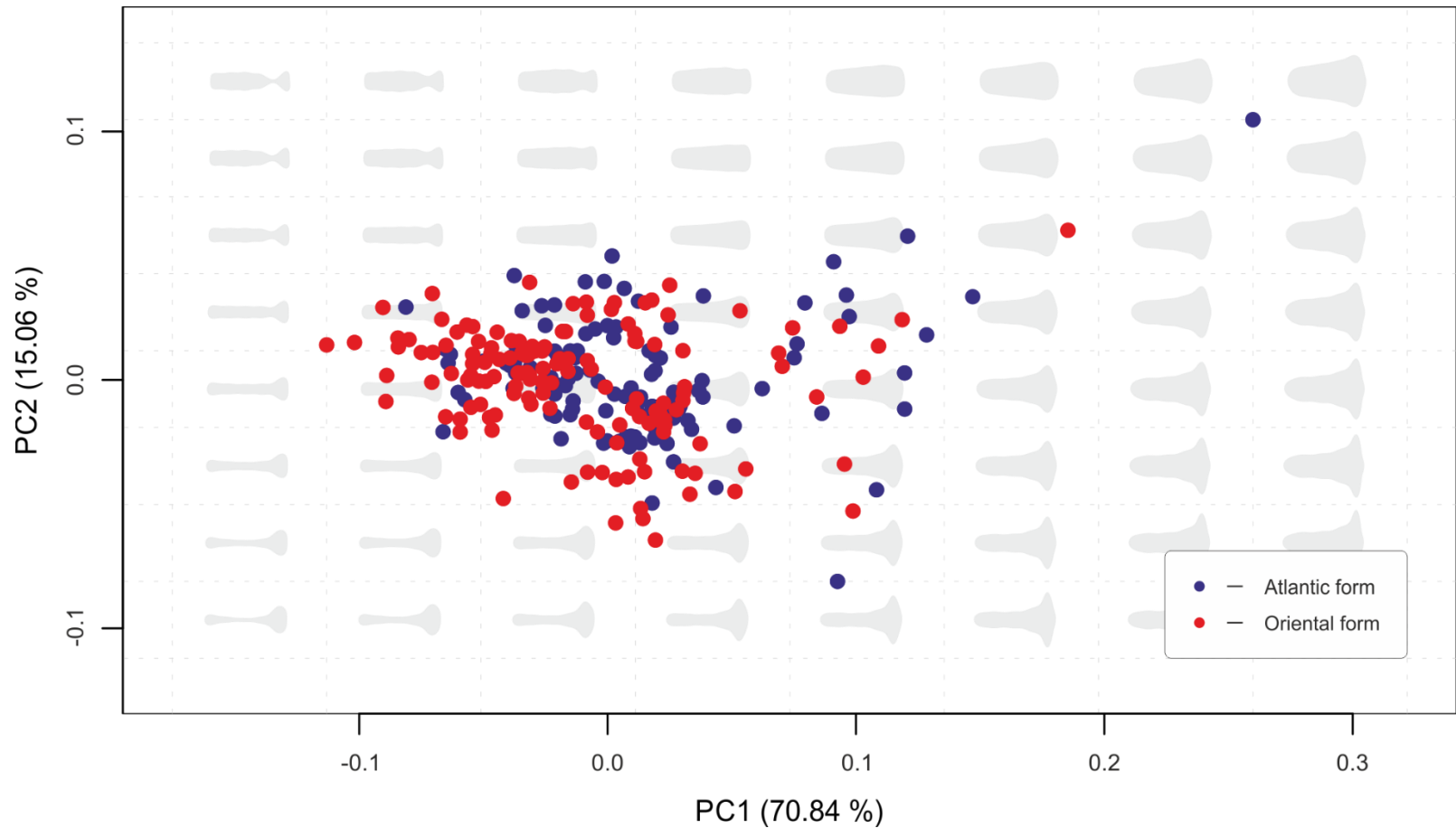




# Unsupervised models for typological classification

## Checking the current typology

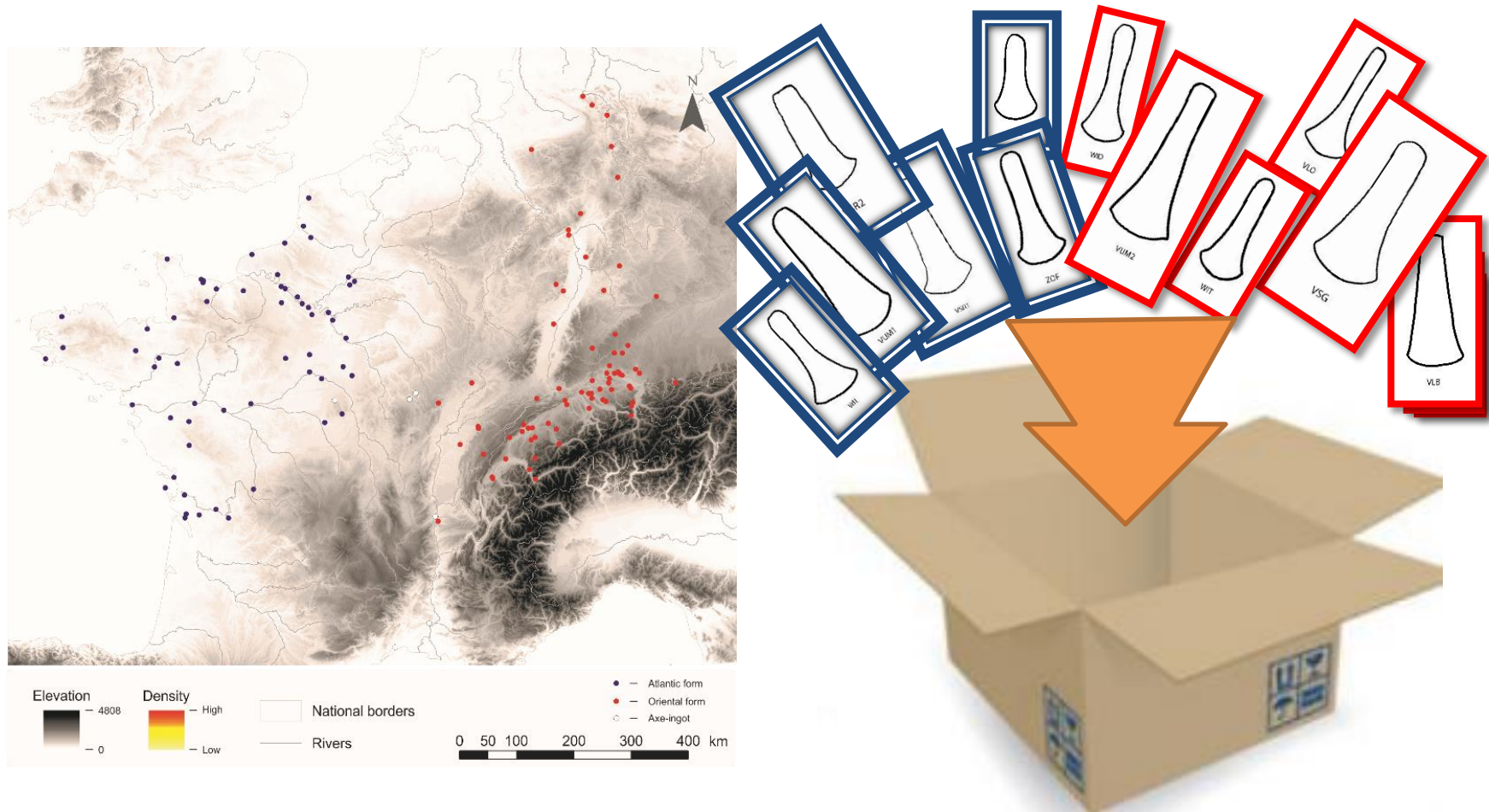
### - PCA



# Unsupervised models for typological classification

## Checking the current typology

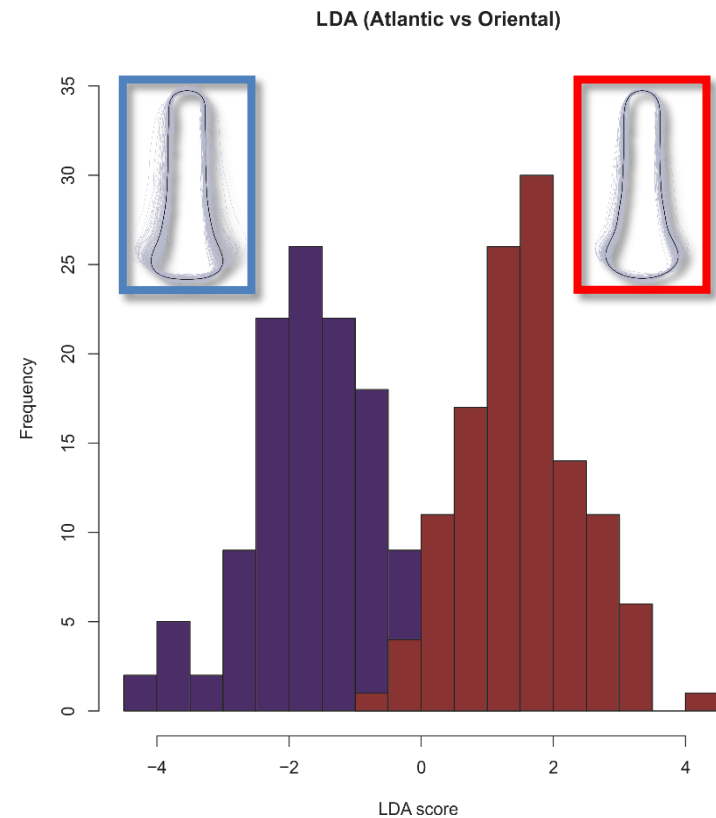
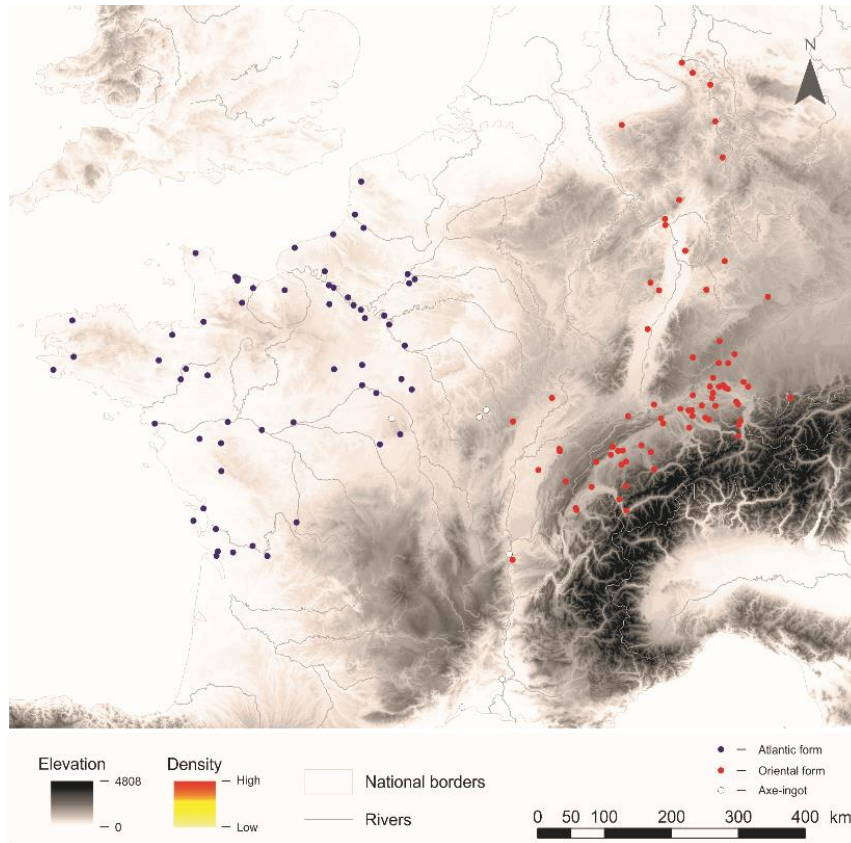
- MANOVA ( $p < 2.2 * 10^{-16}$ ) + Linear discriminant analysis (posteriori classification)



# Unsupervised models for typological classification

## Checking the current typology

- MANOVA ( $p < 2.2 * 10^{-16}$ ) + Linear discriminant analysis (posteriori classification)



```
Df Hotelling-Lawley approx F num Df den Df Pr(>F)
GR2.groups 1 2.3211 11.605 41 205 < 2.2e-16 ***
Residuals 245
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Unsupervised models for typological classification

## Creation of new typology

### - Model-Based Cluster Analysis







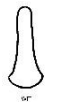



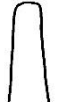
















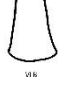







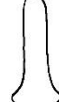








# Unsupervised models for typological classification

## Creation of new typology

- Model-Based Cluster Analysis
- we have found 6 groups



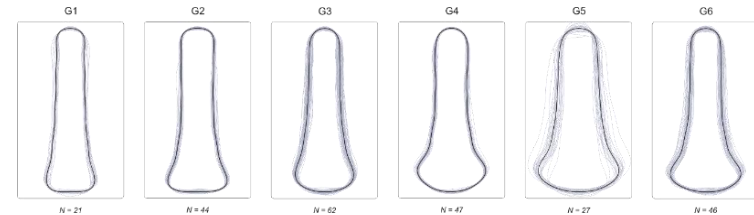
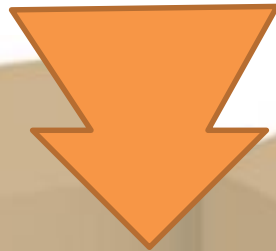
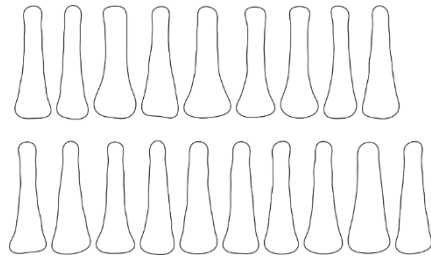
G1	G2	G3	G4	G5	G6
 N = 21	 N = 44	 N = 62	 N = 47	 N = 27	 N = 46
G1	G2	G3	G4	G5	G6
 S4C	 S4E	 POI	 VSR2	 NEV3	 VIT1
 S4E	 L2M	 P1A3	 VSR1	 N2V2	 VSA
 S4E	 C13	 P1B1	 V16	 N2V1	 ELS
 S4E	 P1C1	 P1C2	 VSR3	 N1V1	 ERP1
 S4E	 P1D1	 P1D2	 VSR4	 N1V2	 ERP2
 ROT	 P1E1	 P1E2	 V18	 N1V3	 ERP3
21	44	62	47	27	46

# Unsupervised models for typological classification

## Creation of new typology

### - Classification of new individuals (axes-ingots)

21 axes-ingots



G1	G2	G3	G4	G5	G6
64"	64"	POI	VSR2	NEV3	VIT1
WIT	CON	HAL	VSR1	NEV2	VSA
WIT	C37				
W					
		PCNE			ELS
WIT					ERP1
ROT	IM0	NOG			
21	44	62	47	27	46






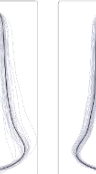
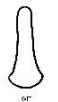



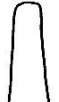
















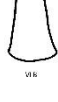







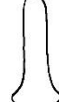






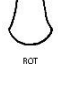







# Unsupervised models for typological classification

## Creation of new typology

- Classification of new individuals (axes-ingots)



G1	G2	G3	G4	G5	G6
 N = 21	 N = 44	 N = 62	 N = 47	 N = 27	 N = 66
G1	G2	G3	G4	G5	G6
 SAC	 SAC	 POI	 VSR2	 NEV3	 VIT1
 SAC	 SAC	 SAC	 SAC	 SAC	 SAC
 SAC	 SAC	 SAC	 SAC	 SAC	 SAC
 SAC	 SAC	 SAC	 SAC	 SAC	 SAC
 SAC	 SAC	 SAC	 SAC	 SAC	 SAC
 SAC	 SAC	 SAC	 SAC	 SAC	 SAC
 SAC	 SAC	 SAC	 SAC	 SAC	 SAC
3	0	16	0	1	1

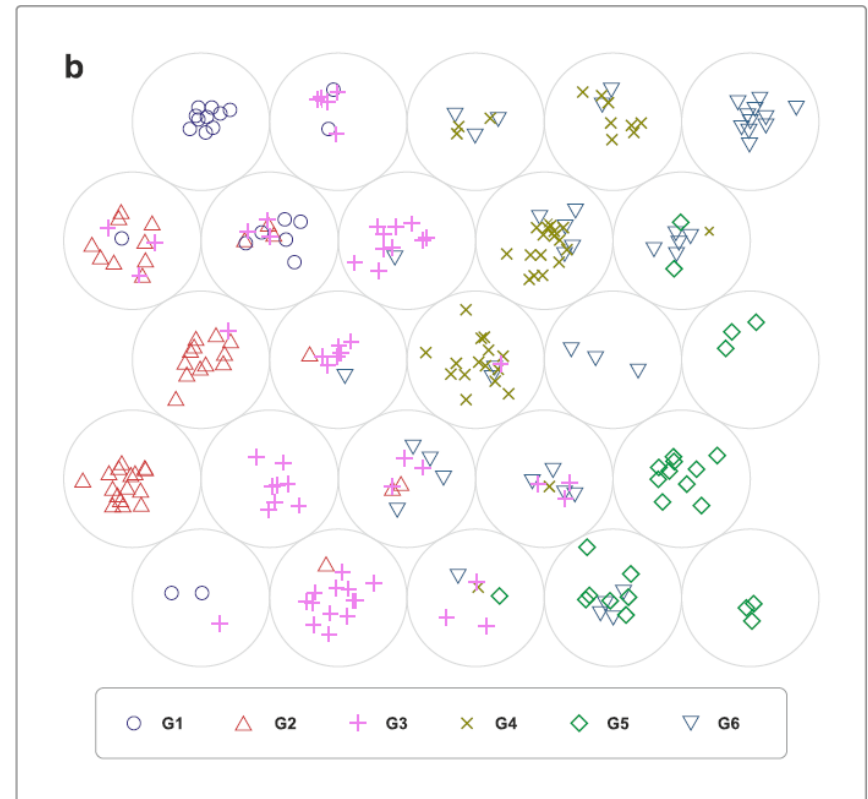
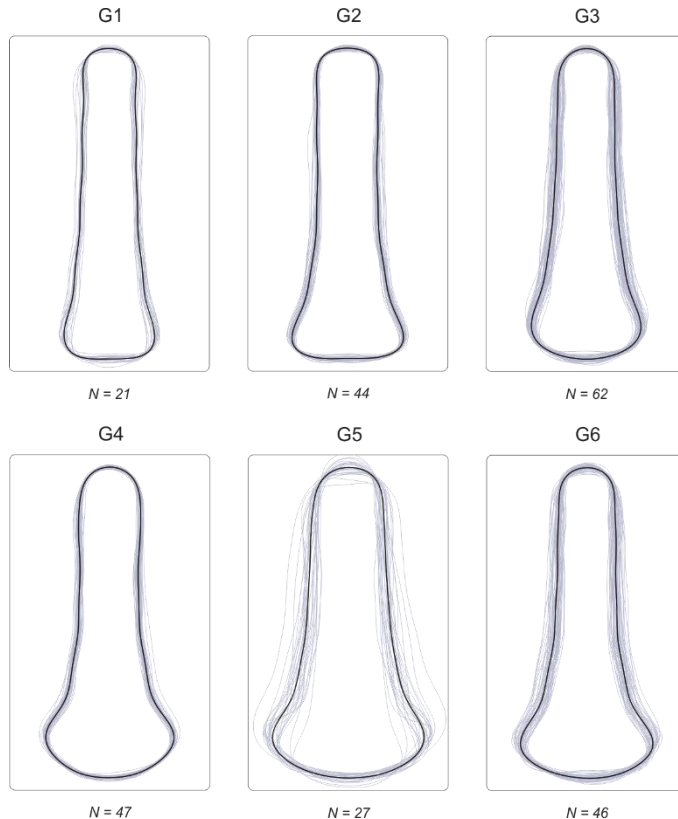
# Unsupervised models for typological classification

## Validation of the new classification

### 1) Visualisation

### 2) SOM

### 3) Spatial validation – Kernel density et Multinomial Scan Statistic (SatScan) - ( $p < 0.001$ )





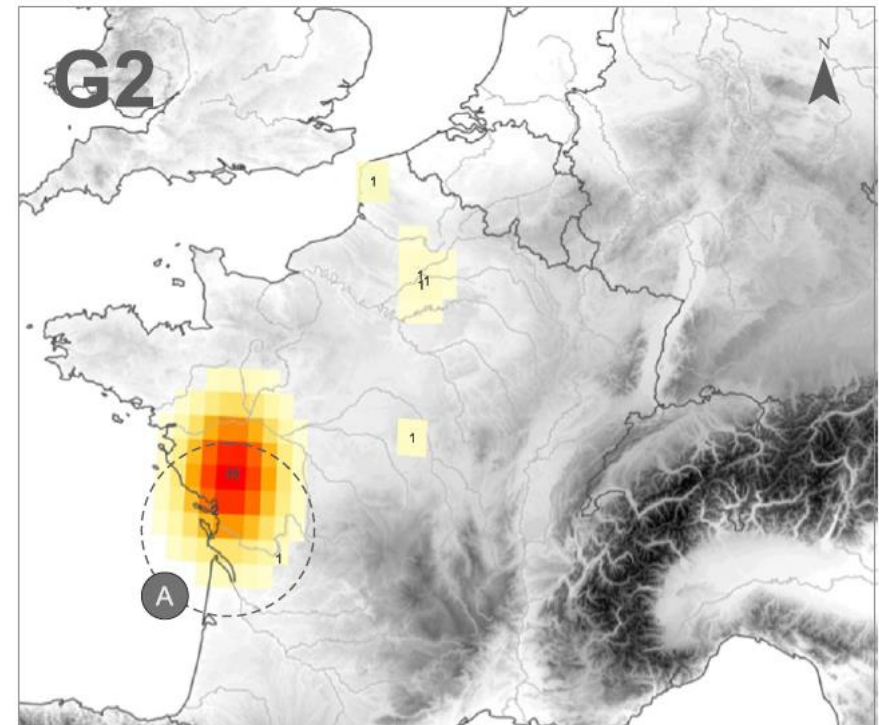
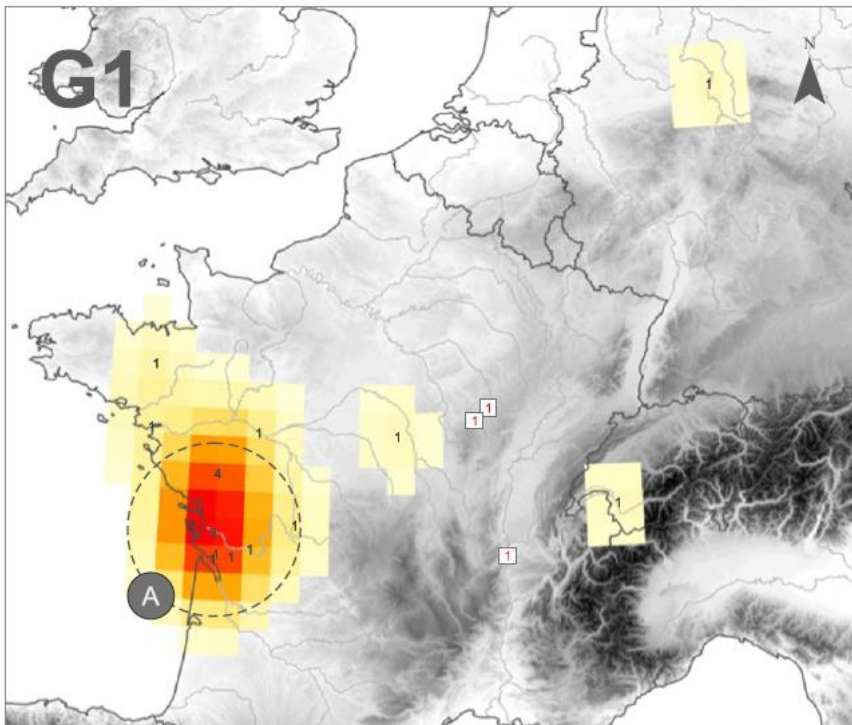
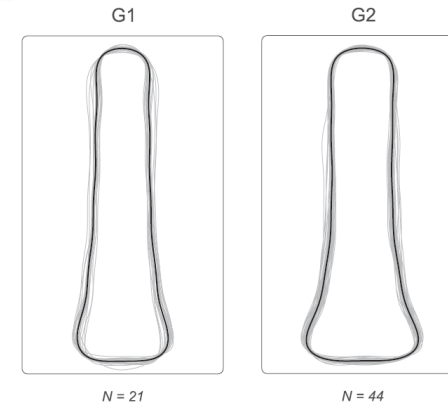
# Unsupervised models for typological classification

## Validation of the new classification

1) Visualisation

2) SOM

3) Spatial validation – Kernel density et Multinomial  
Scan Statistic (SatScan) - ( $p < 0.001$ )



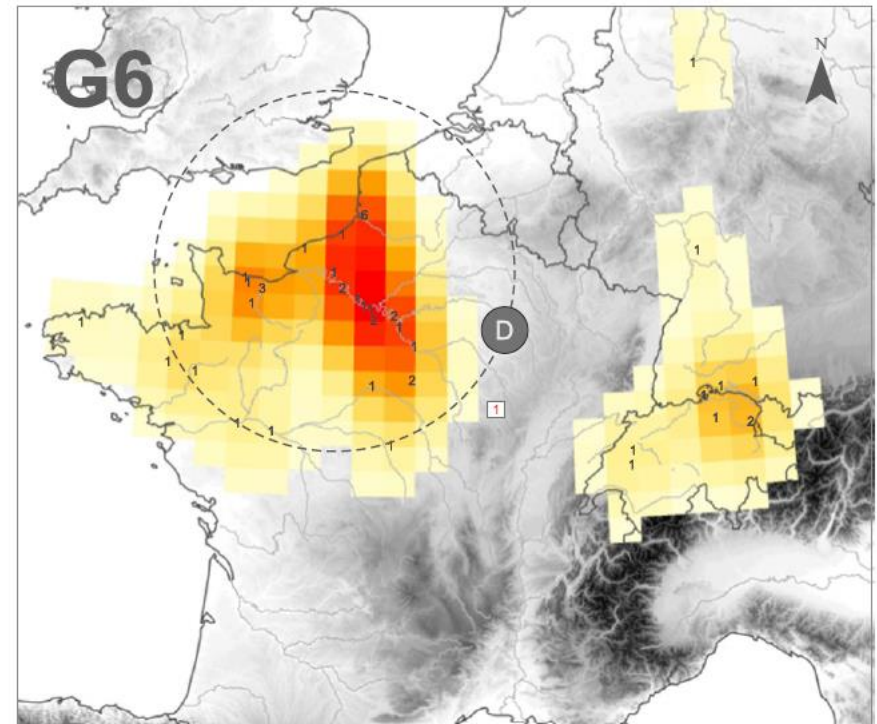
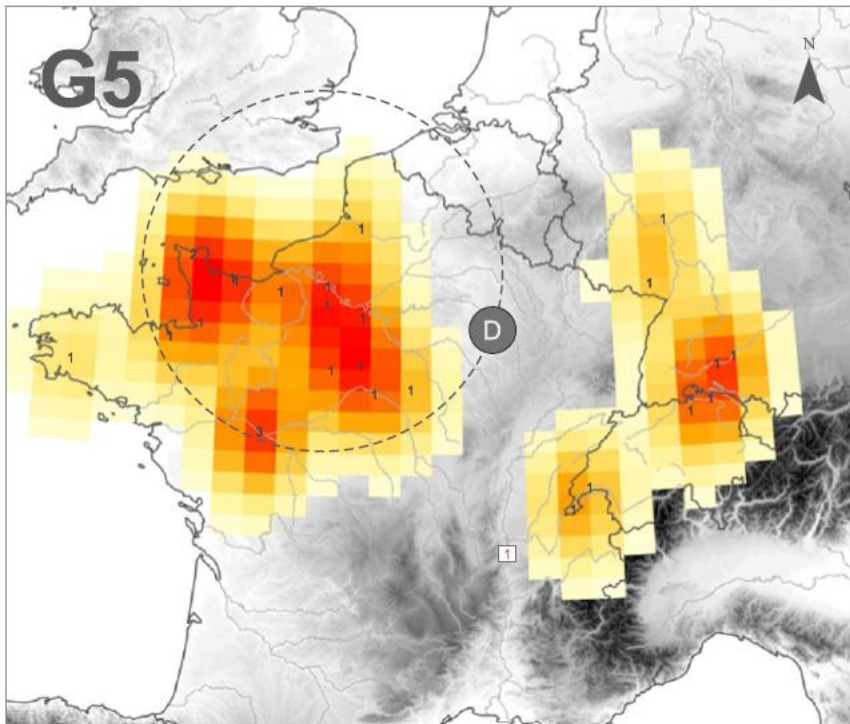
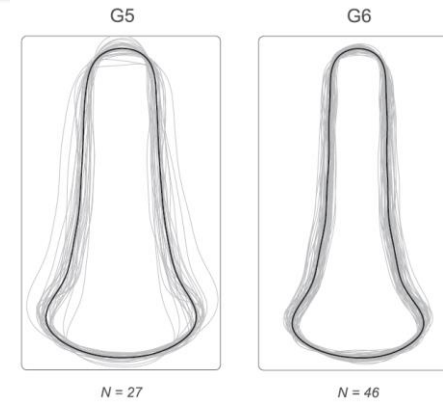
# Unsupervised models for typological classification

## Validation of the new classification

### 1) Visualisation

### 2) SOM

### 3) Spatial validation – Kernel density et Multinomial Scan Statistic (SatScan) - ( $p < 0.001$ )



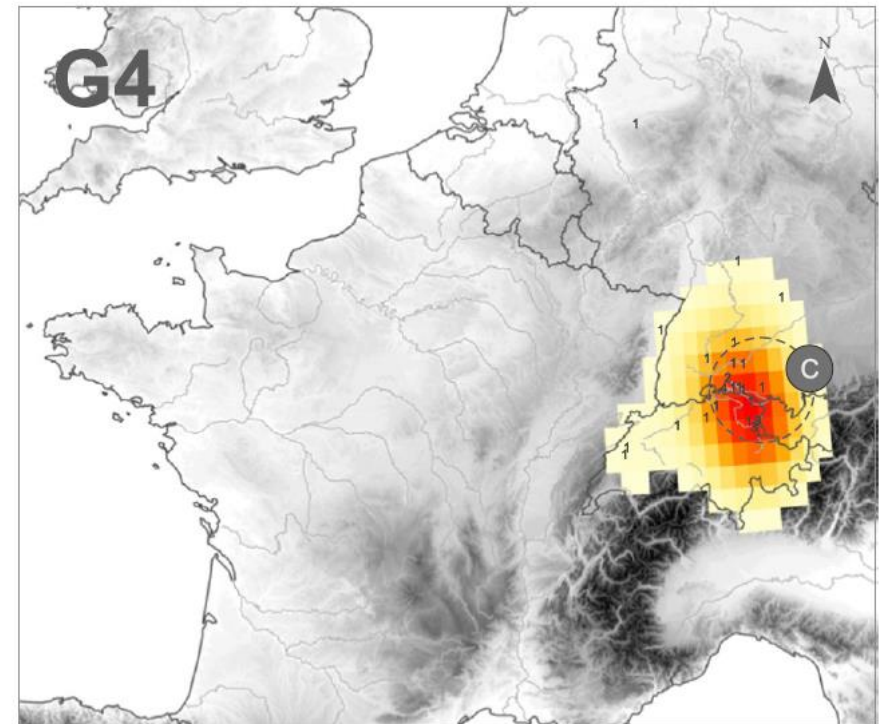
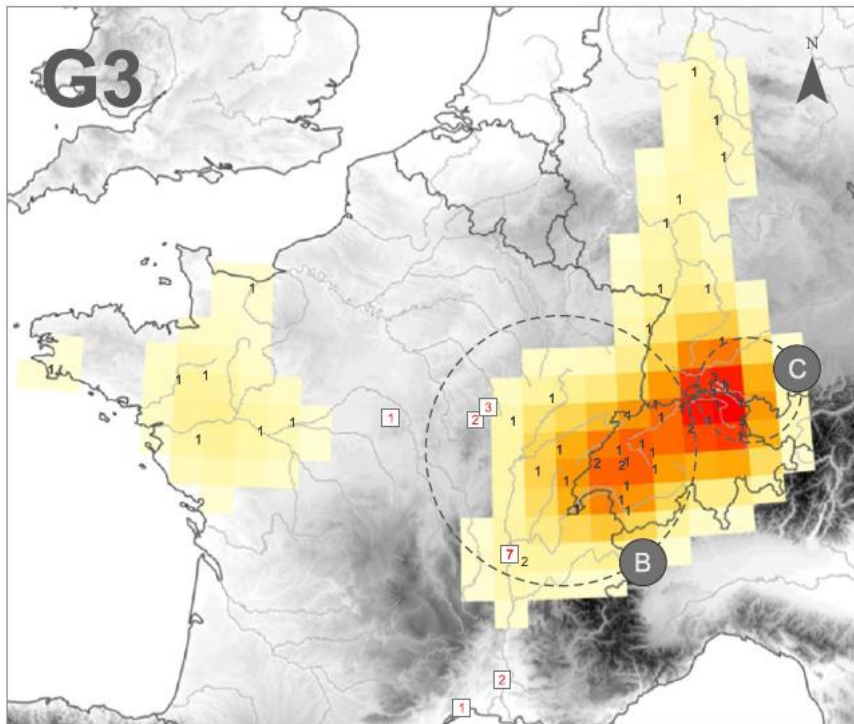
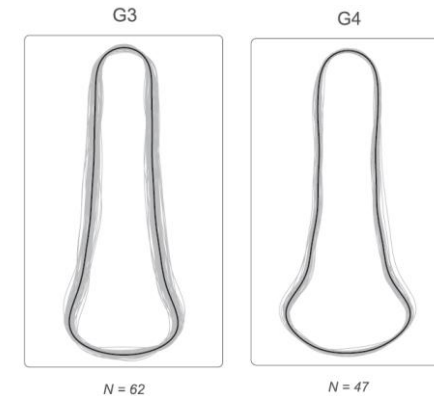
# Unsupervised models for typological classification

## Validation of the new classification

1) Visualisation

2) SOM

3) Spatial validation – Kernel density et Multinomial  
Scan Statistic (SatScan) - ( $p < 0.001$ )



## **Analyse discriminante (DA)**

### Analyse de fonction discriminante, DA

#### **Buts**

- rechercher un critère de distinction entre 2 ou plusieurs groupes d'objets assez comparables
- prédire l'appartenance d'un individu à une classe (groupe)

#### **Les conditions**

- Les deux groupes sont définis à l'avance.
- Un individu ne peut appartenir qu'à un groupe.
- Les deux groupes doivent avoir une intersection non nulle.

## **Comment discriminer?**

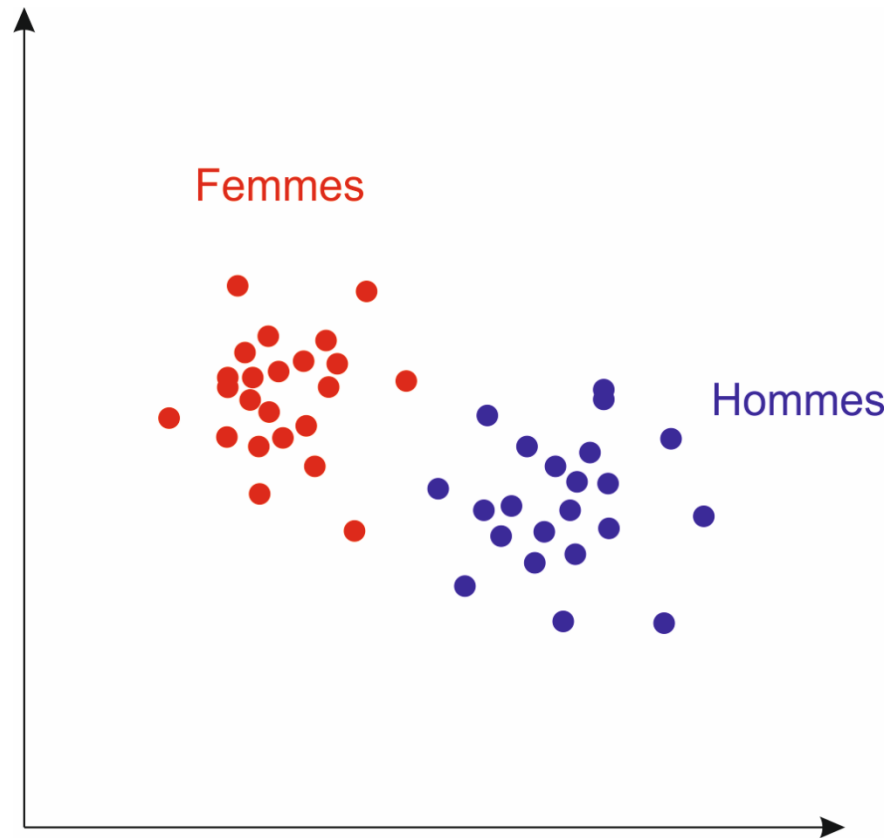
- Trouver des fonctions discriminantes qui sont des vecteurs dans les directions de **séparations optimales** des groupes.
- C'est à dire déterminer un critère de classement qui **maximise la différence** entre les groupes tout en **minimisant le risque de mauvais classement**.



### Exemple intuitif

On a deux groupes (H, F)

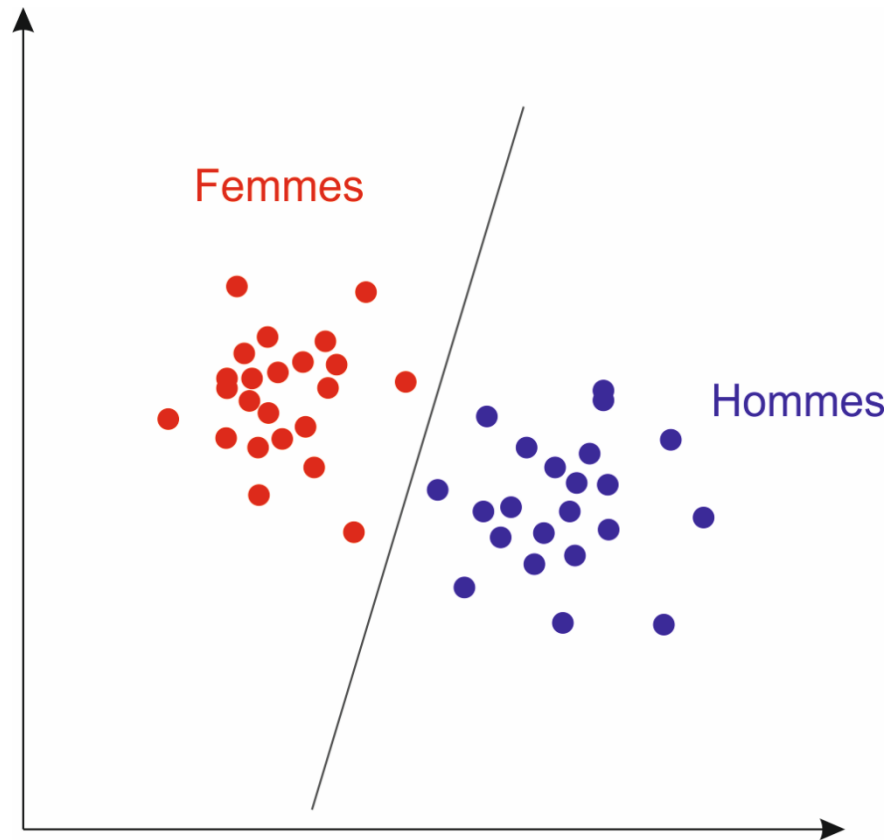
Quel est la ligne qui nous sépare ses deux groups?



### Exemple intuitif

On a deux groupes (H, F)

Quel est la ligne qui nous sépare ses deux groupes?

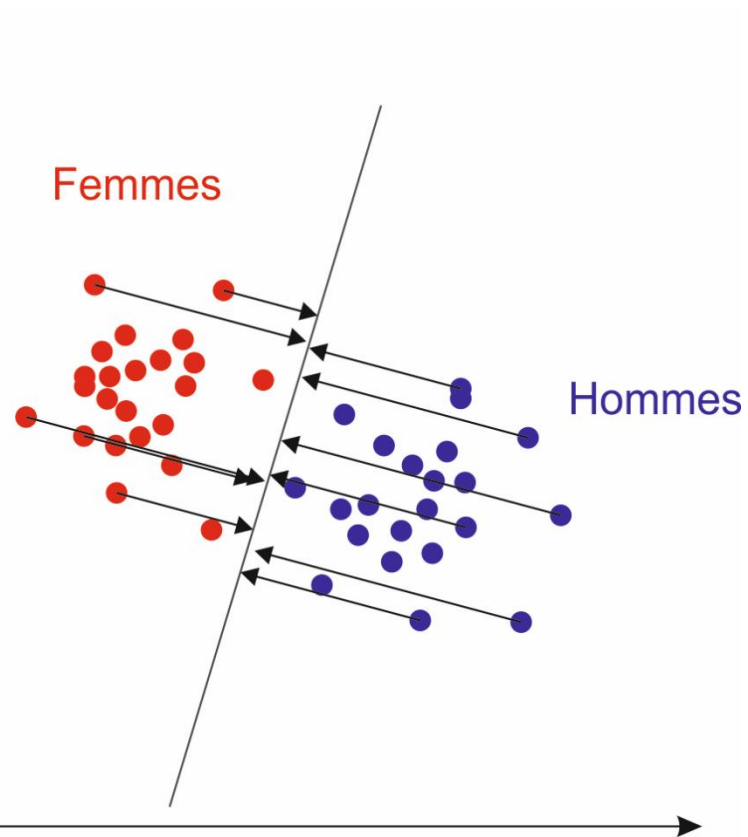




### Exemple intuitif

On a deux groupes (H, F)

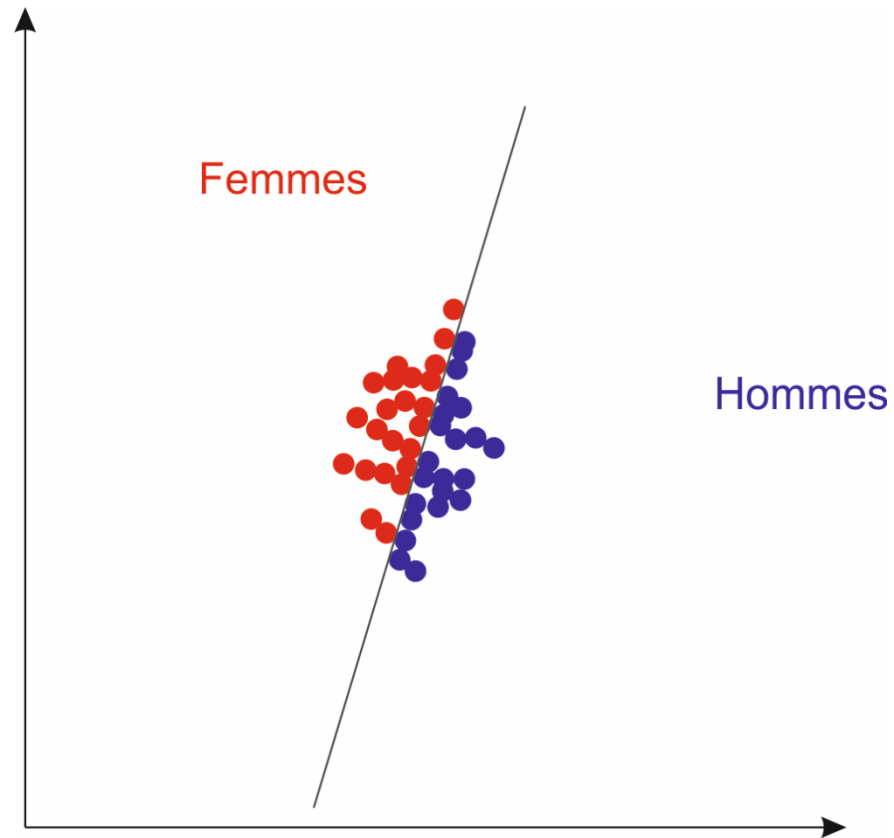
Quel est la ligne qui nous sépare ses deux groupes?



### Exemple intuitif

On a deux groupes (H, F)

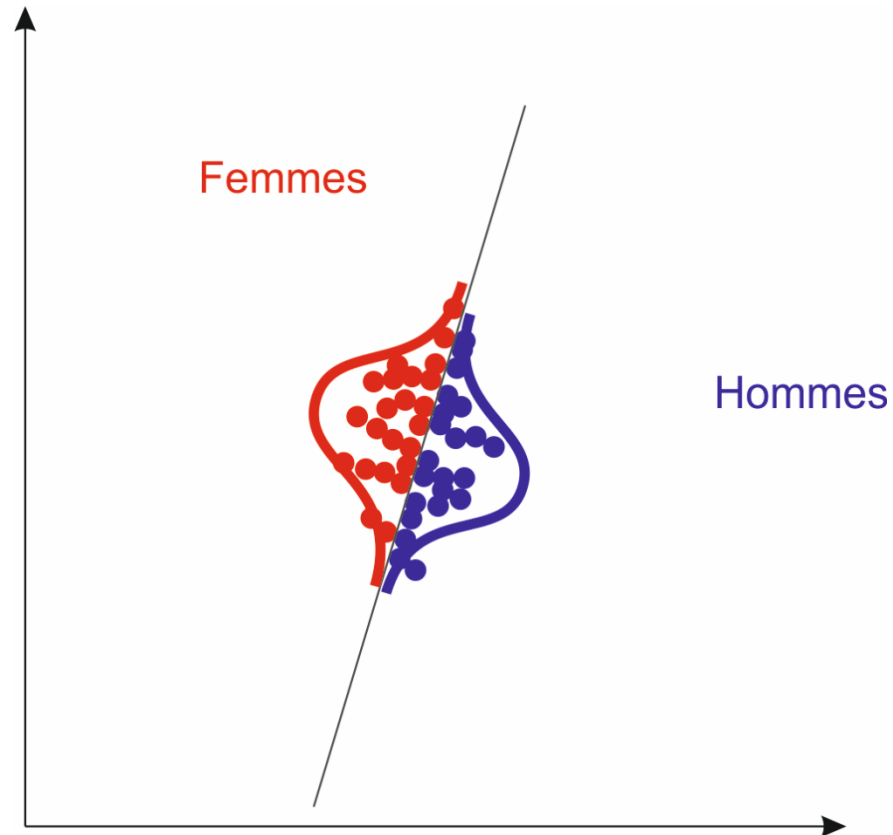
Quel est la ligne qui nous sépare ses deux groupes?



### Exemple intuitif

On a deux groupes (H, F)

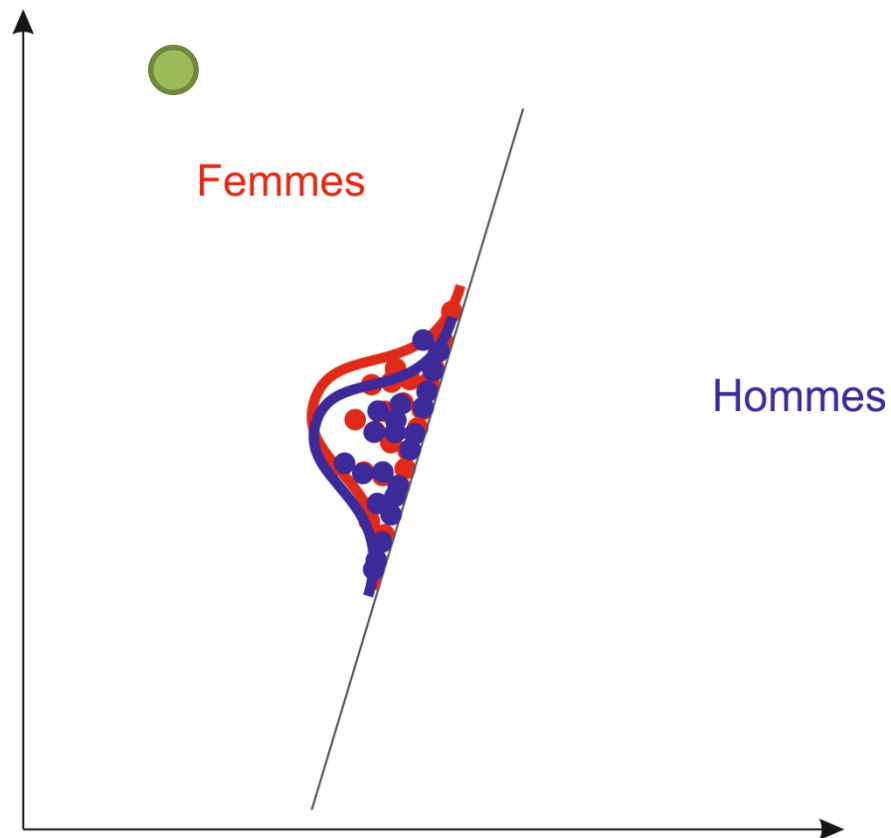
Quel est la ligne qui nous sépare ses deux groupes?



### Exemple intuitif

On a deux groupes (H, F)

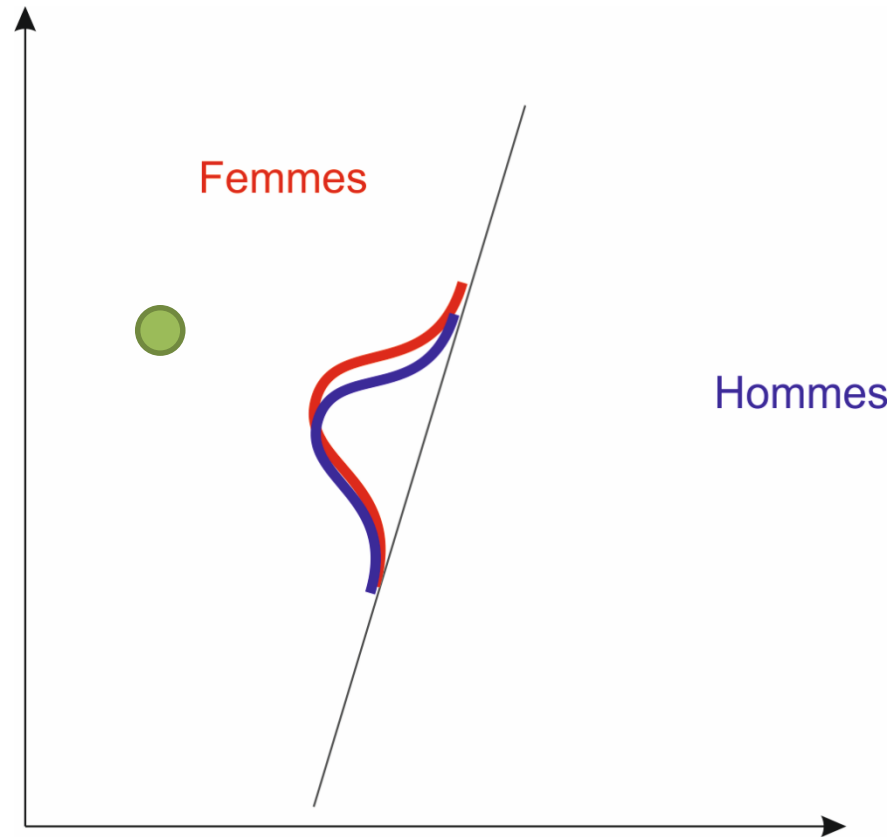
Quel est la ligne qui nous sépare ses deux groupes?



## Exemple intuitif

On a deux groupes (H, F)

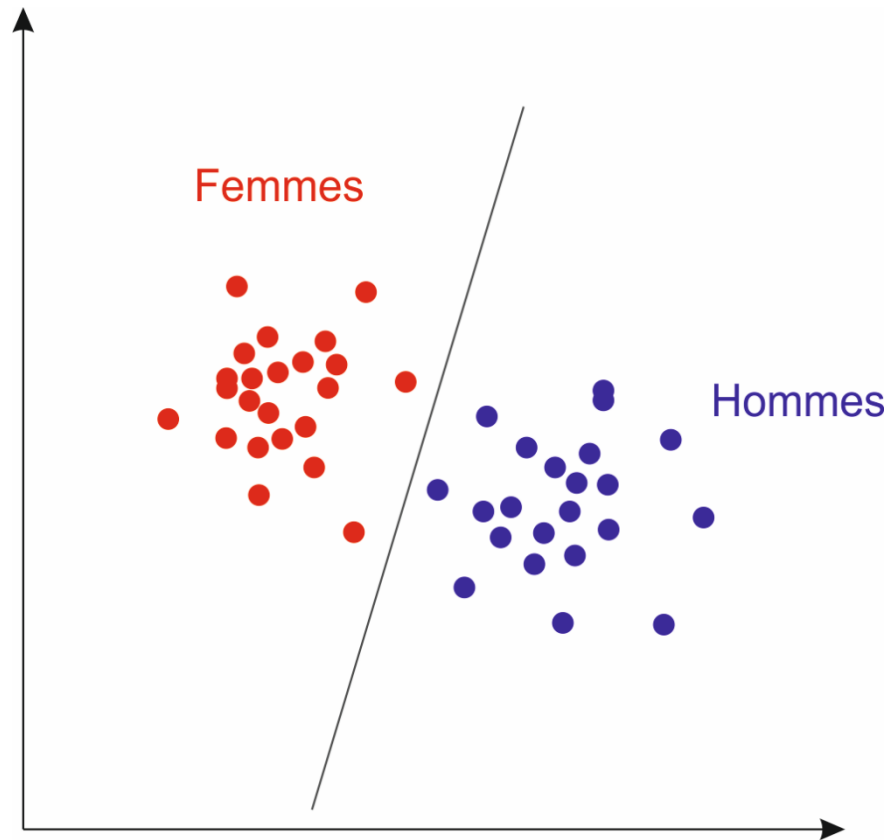
Quel est la ligne qui nous sépare ses deux groupes?



### Exemple intuitif

On a deux groupes (H, F)

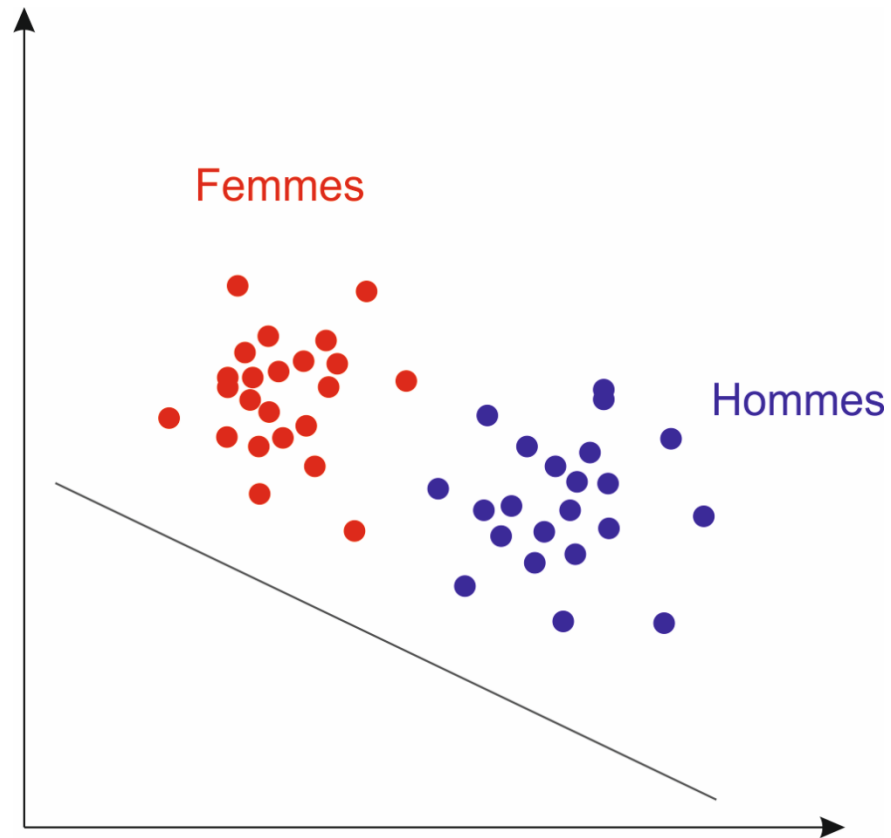
Quel est la ligne qui nous sépare ses deux groupes?



### Exemple intuitif

On a deux groupes (H, F)

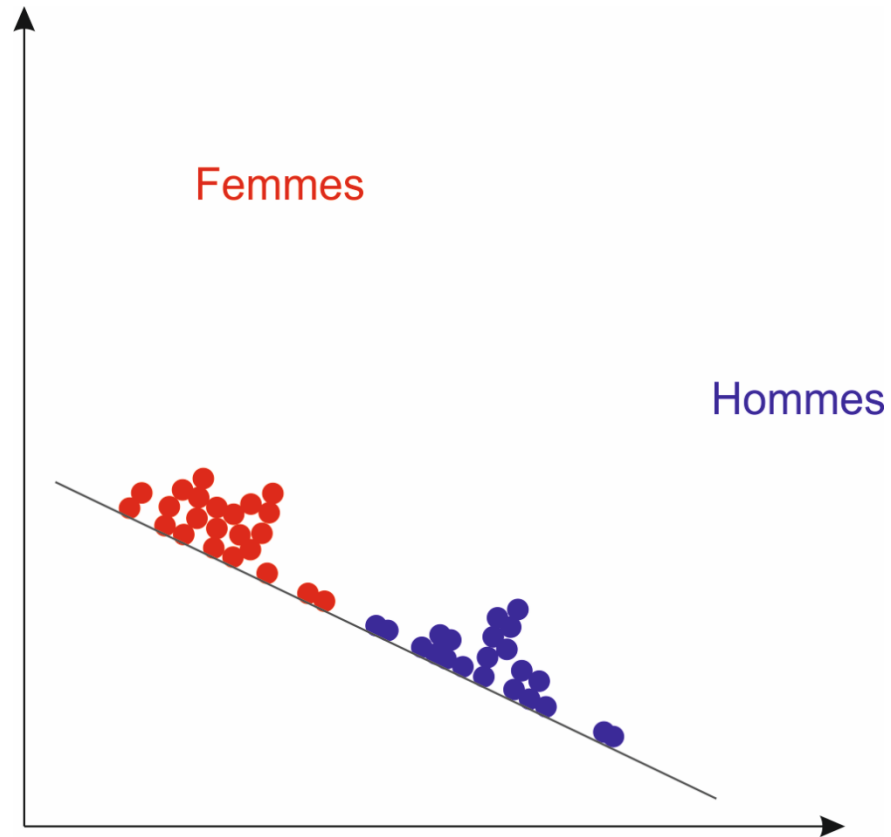
Quel est la ligne qui nous sépare ses deux groupes?



### Exemple intuitif

On a deux groupes (H, F)

Quel est la ligne qui nous sépare ses deux groupes?

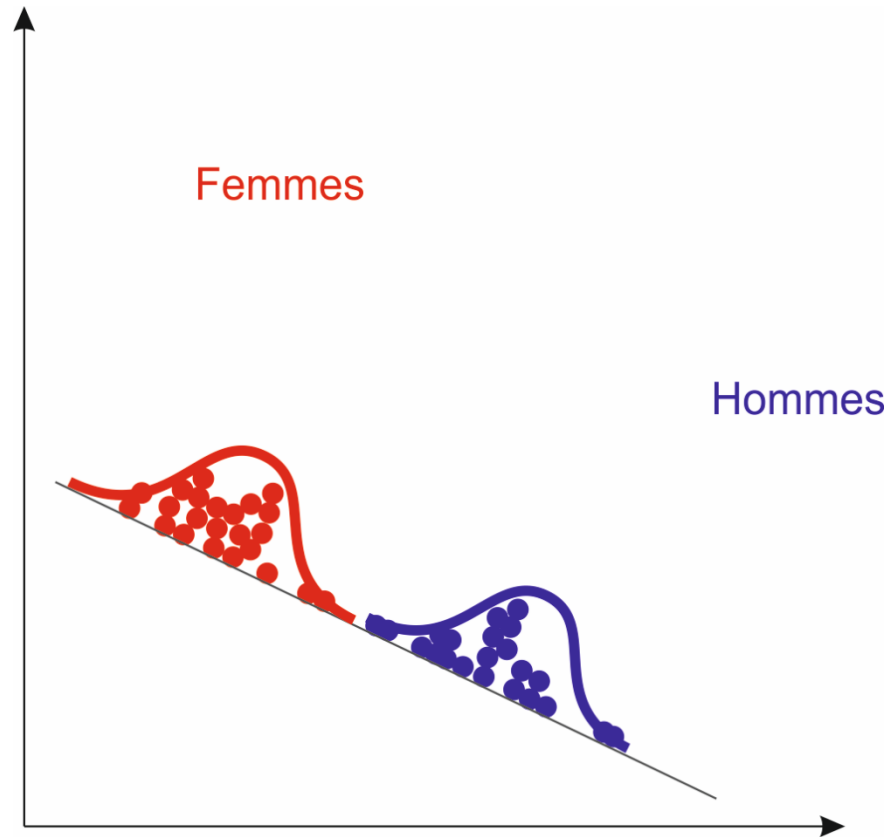




## Exemple intuitif

On a deux groupes (H, F)

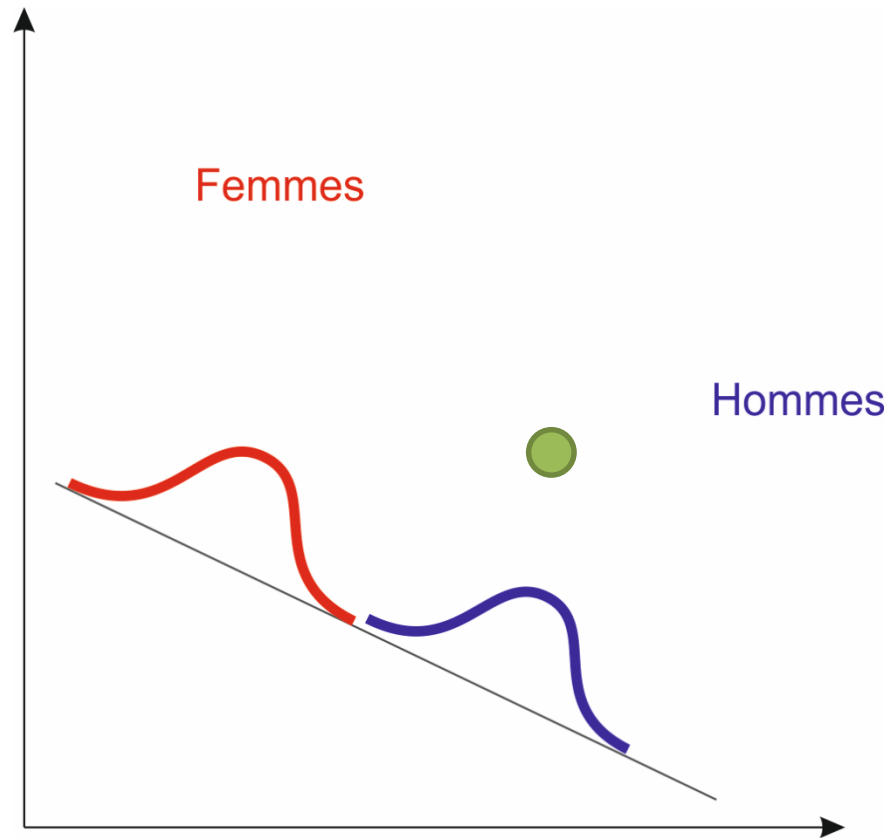
Quel est la ligne qui nous sépare ses deux groups?



## Exemple intuitif

On a deux groupes (H, F)

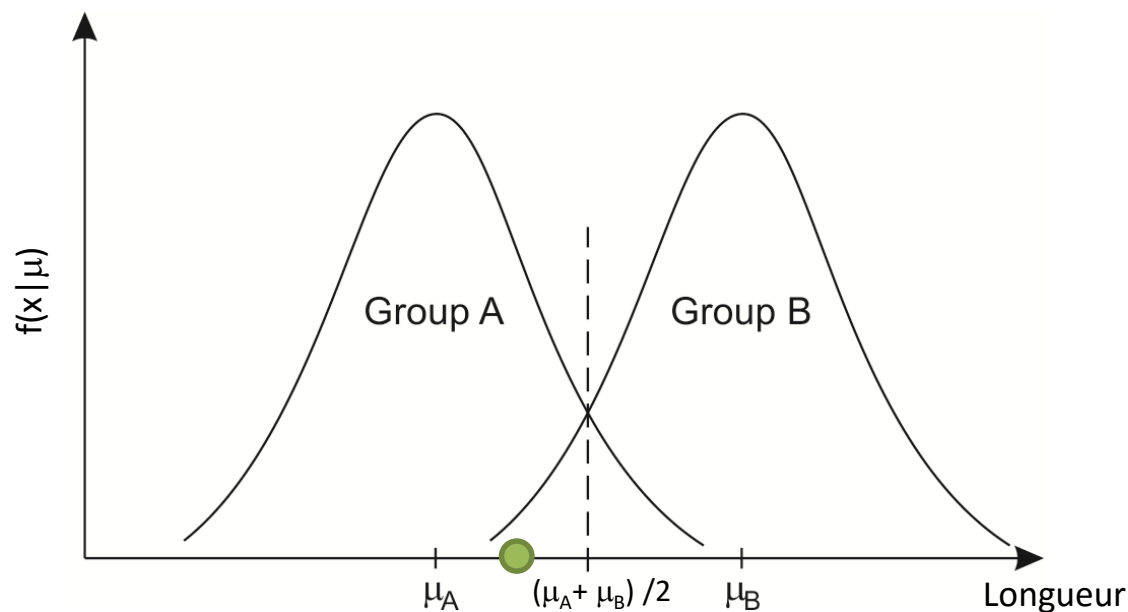
Quel est la ligne qui nous sépare ses deux groupes?



## Analyse à une variable, deux groupes

### Exemple

- on a juste une variable = longueur et on veut **connaitre l'appartenance d'un individu**
- deux groupes (normalement distribuées), **même variance**, mais différentes moyennes :  $\mu_A, \mu_B$



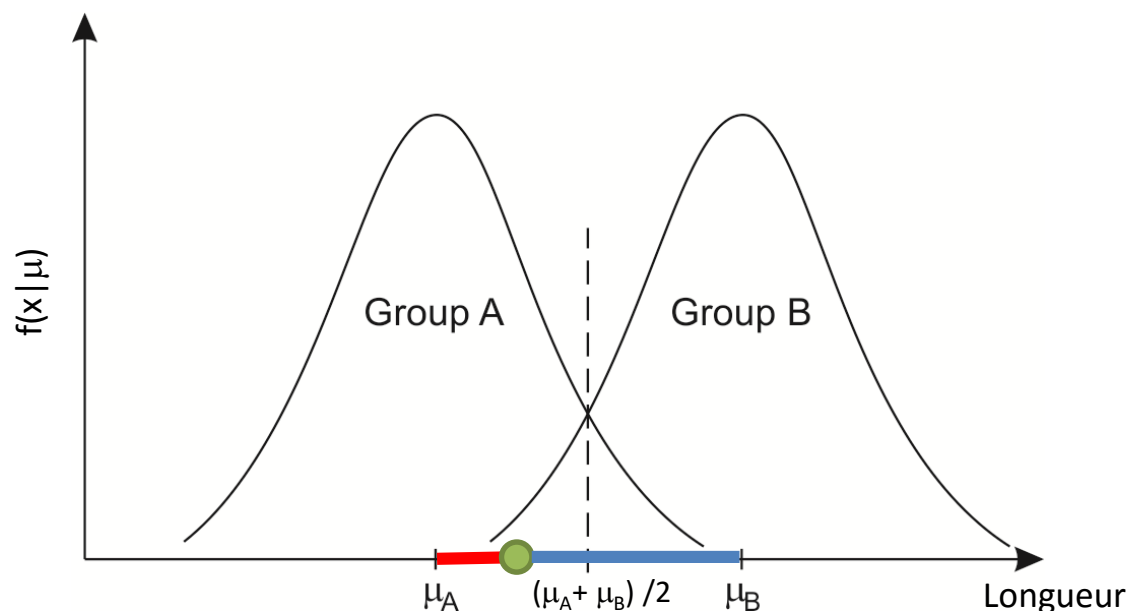
### Individu appartient a groupe A:

- 1) si la longueur  $x$  est plus près de  $\mu_A$  que de  $\mu_B$
- 2) si  $x < (\mu_A + \mu_B) / 2$
- 3) ou  $f(x | \mu_A) > f(x | \mu_B)$

## Analyse à une variable, deux groupes

### Exemple

- on a juste une variable = longueur et on veut **connaitre l'appartenance d'un individu**
- deux groupes (normalement distribuées), **même variance**, mais différentes moyennes :  $\mu_A, \mu_B$



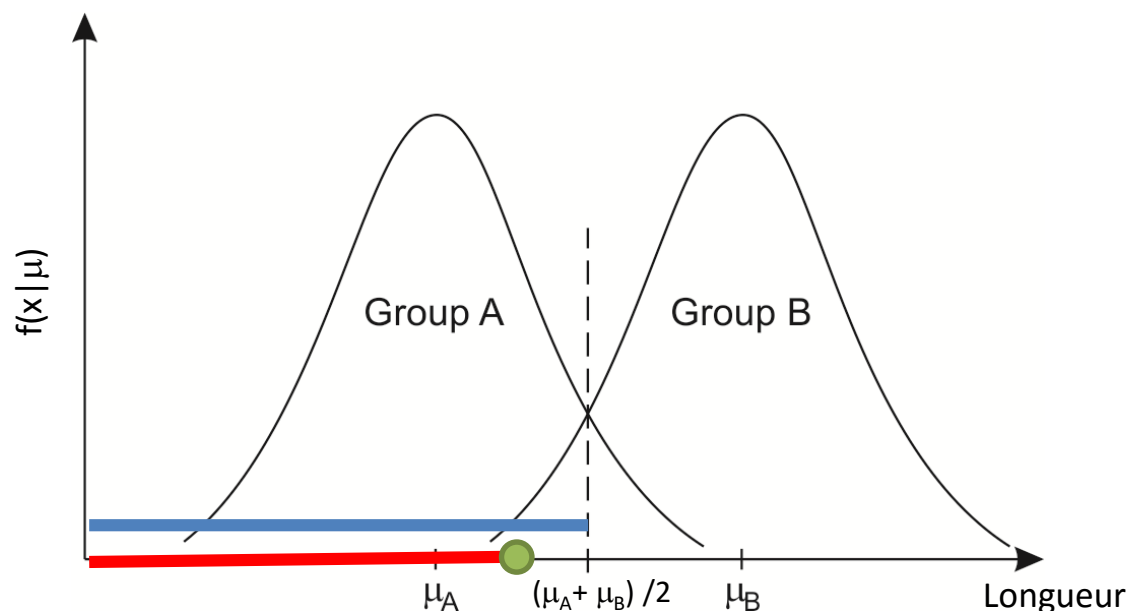
### Individu appartient a groupe A:

- 1) si la longueur  $x$  est **plus près de  $\mu_A$**  que de  $\mu_B$
- 2) si  $x < (\mu_A + \mu_B) / 2$
- 3) ou  $f(x|\mu_A) > f(x|\mu_B)$

## Analyse à une variable, deux groupes

### Exemple

- on a juste une variable = longueur et on veut **connaitre l'appartenance d'un individu**
- deux groupes (normalement distribuées), **même variance**, mais différentes moyennes :  $\mu_A, \mu_B$



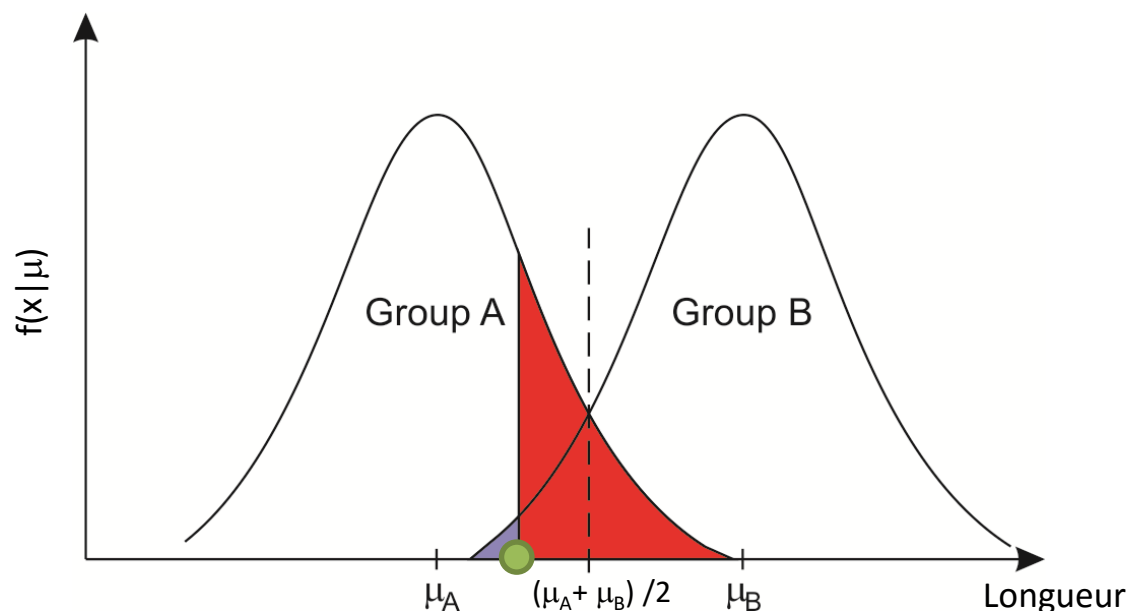
### Individu appartient a groupe A:

- 1) si la longueur  $x$  est plus près de  $\mu_A$  que de  $\mu_B$
- 2) si  $x < (\mu_A + \mu_B) / 2$
- 3) ou  $f(x | \mu_A) > f(x | \mu_B)$

## Analyse à une variable, deux groupes

### Exemple

- on a juste une variable = longueur et on veut **connaitre l'appartenance d'un individu**
- deux groupes (normalement distribuées), **même variance**, mais différentes moyennes :  $\mu_A, \mu_B$



### Individu appartient a groupe A:

- 1) si la longueur  $x$  est plus près de  $\mu_A$  que de  $\mu_B$
- 2) si  $x < (\mu_A + \mu_B) / 2$
- 3) ou  $f(x | \mu_A) > f(x | \mu_B)$

## Analyse à plusieurs variables (LDA)

- Dans la pratique on ne travaille presque jamais avec une seule variable. Il faut rechercher une ou plusieurs fonctions discriminantes

$$D_i = b_0 + \sum_{k=1}^p b_k X_k$$

- ...tels que les groupes diffèrent autant que possible sur D

- $D_i$  Fonctions discriminantes linéaire = **combinaisons linéaires** des prédicteurs ( $X_k$ )
- $X_k$  Prédicteurs / les variables indépendants
- $b_0, b_k$  Les paramètres inconnues

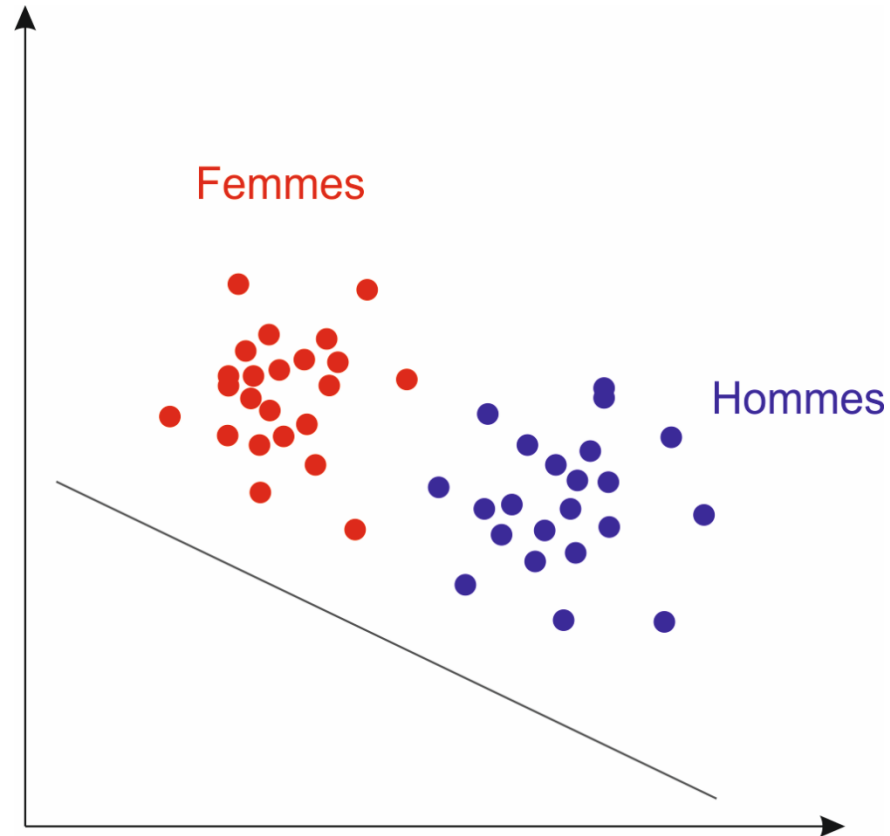
- Comment obtenir les valeurs b?
  - variation totale **T** décomposée
    - en variation entre les groupes **B** (B = between)
    - variation à l'intérieur des groupes **W** (W = within)
  - Les valeurs b sont calculées de sorte que le rapport variation entre les groupes / variation à l'intérieur des groupes est maximum!!

$$D_i = b_0 + \sum_{k=1}^p b_k X_k$$

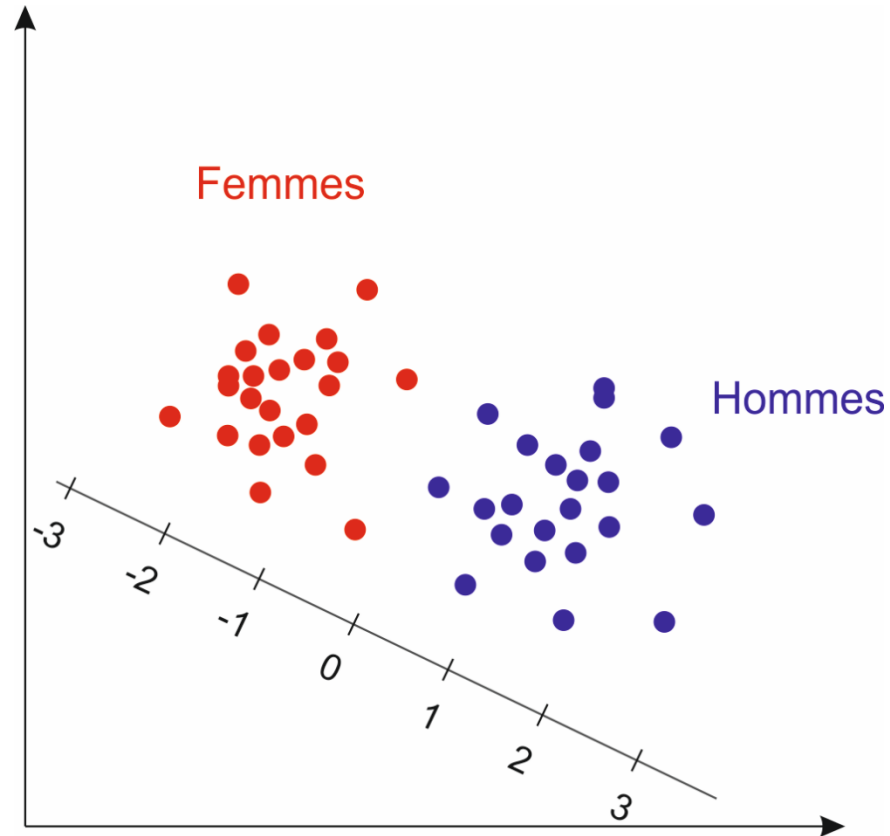
- **Une fois calculé, on obtiens**
  - **Fonctions discriminantes linéaire ( $D_i$ ) = les axes discriminants**
    - La première fonction discriminante D1 distingue le premier groupe des groupes 2,3,..N.
    - La deuxième fonction D2 distingue le second groupe des groupes 3, 4...,N.
    - etc.
  - On peut transformer les valeurs originales en **scores discriminants** qui représentent la position de chaque individu sur les axes discriminants.



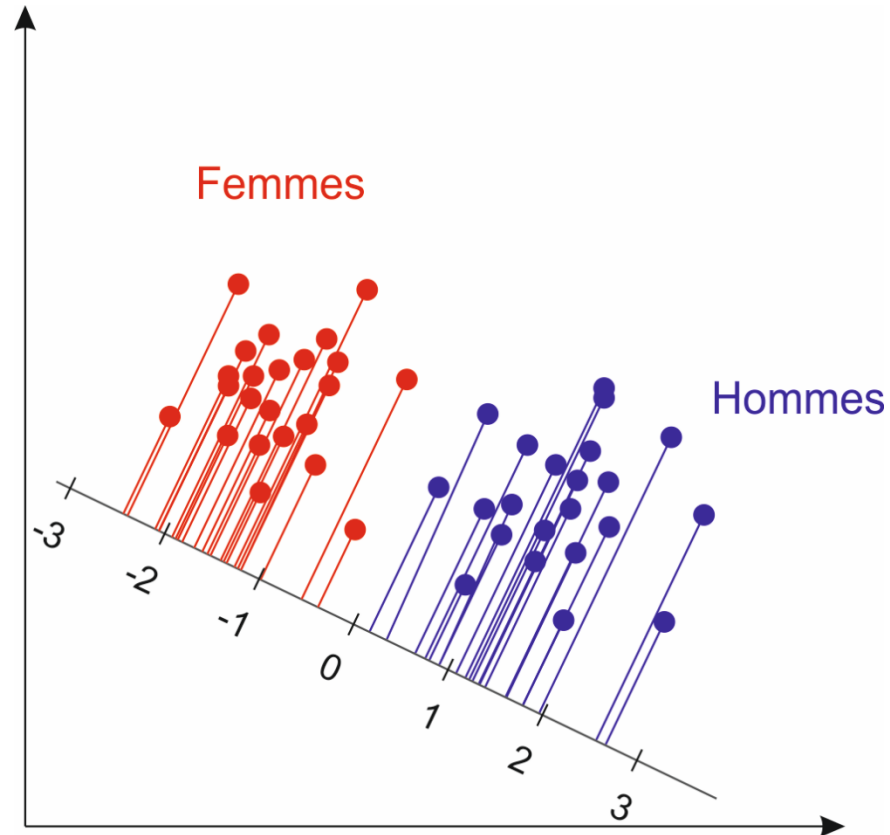
- Les scores discriminants



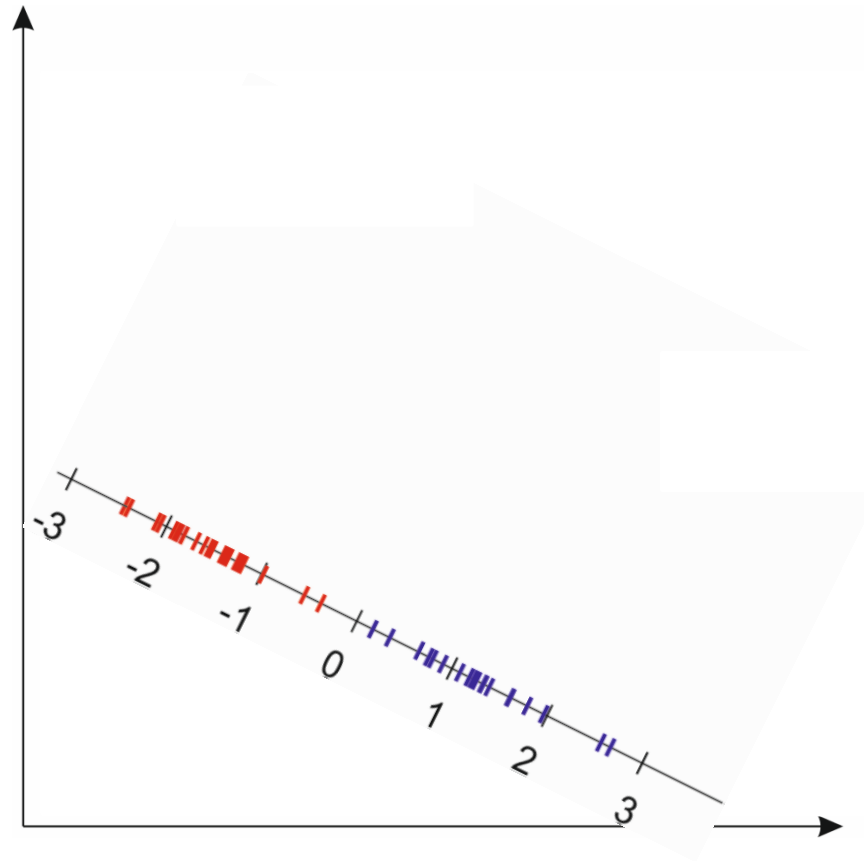
- Les scores discriminants



- Les scores discriminants

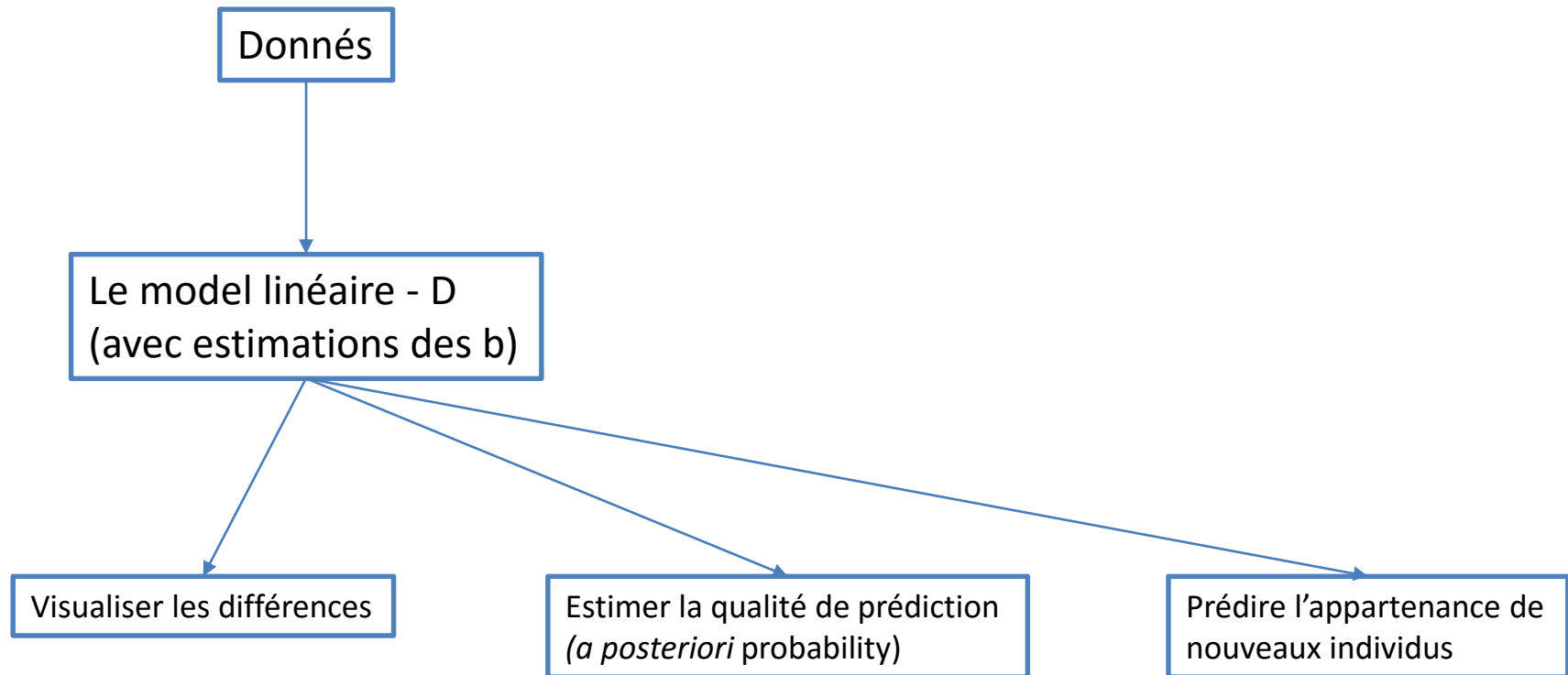


- Les scores discriminants

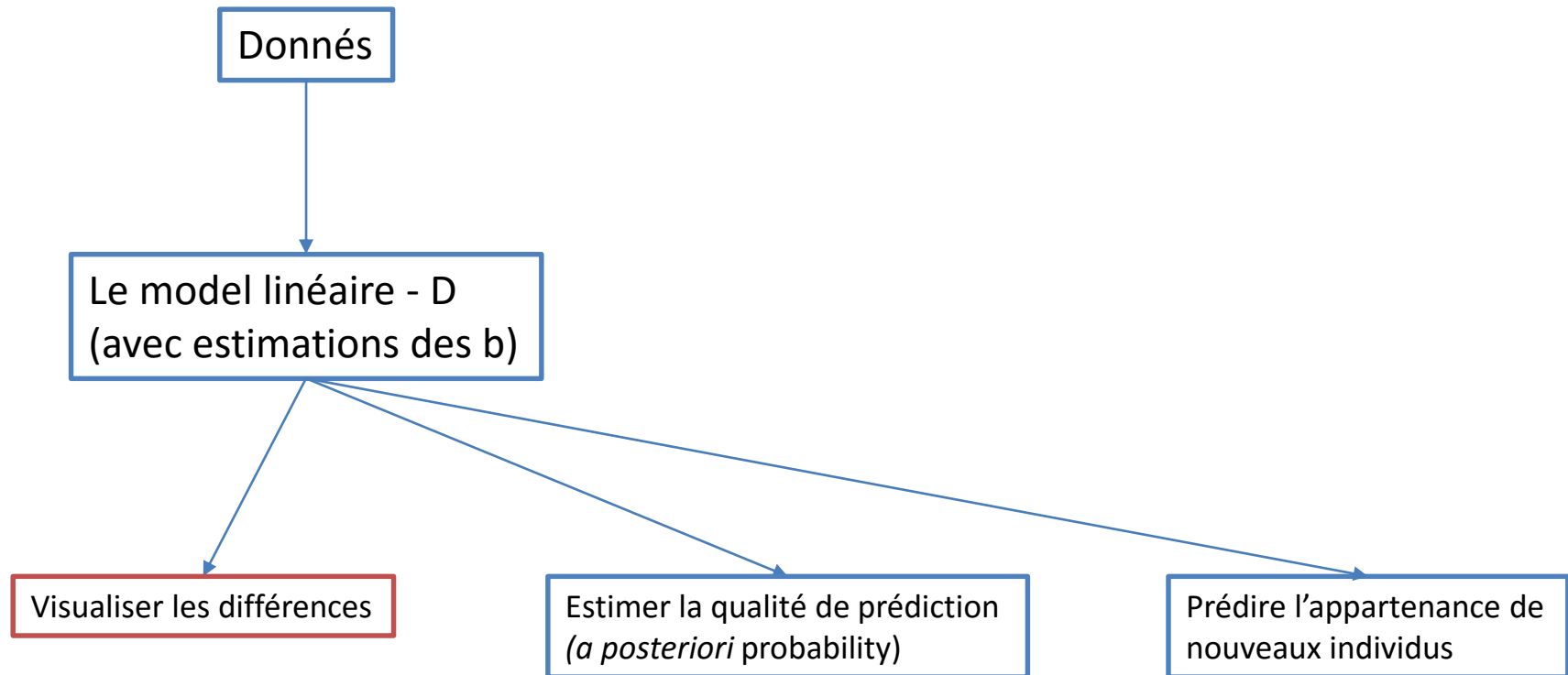


- Que-est ce qu'on peut faire avec LDA?

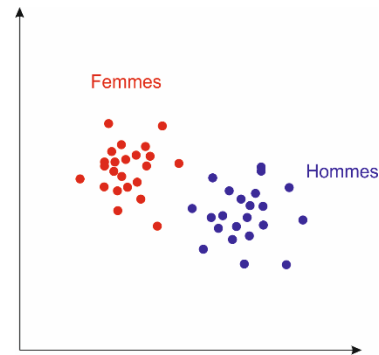
### - Que-est ce qu'on peut faire avec LDA?



### - Que-est ce qu'on peut faire avec LDA?



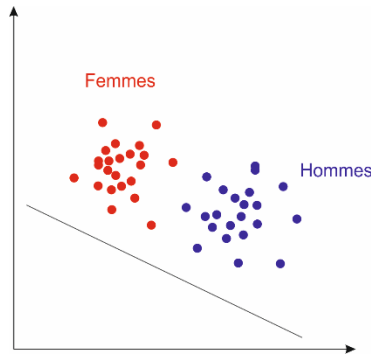
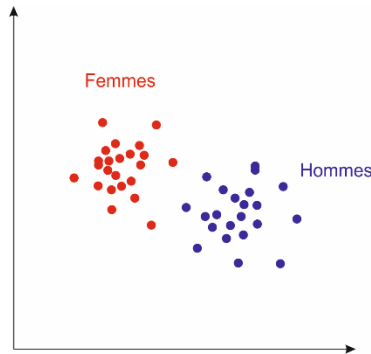
Donnés



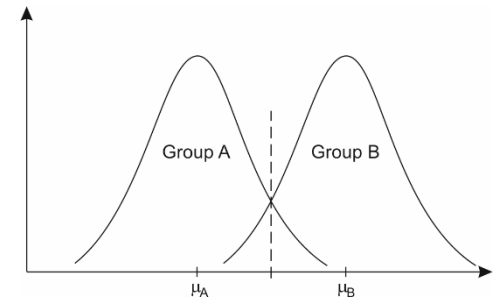


Donnés

Le model linéaire - D  
(avec estimations des b)



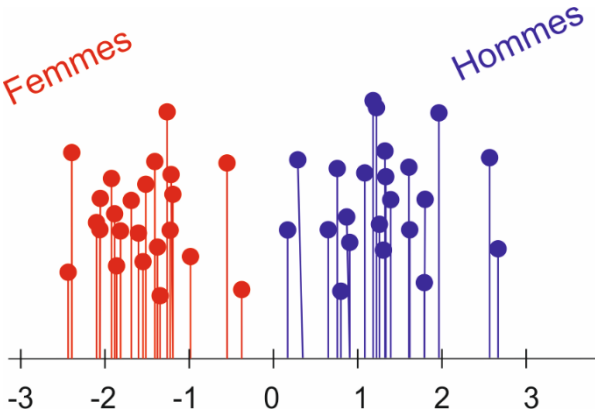
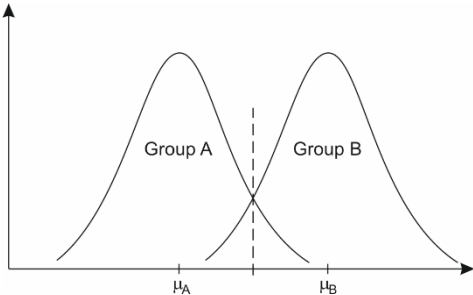
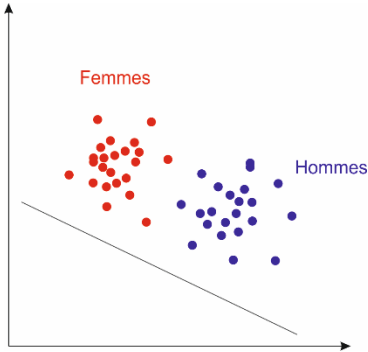
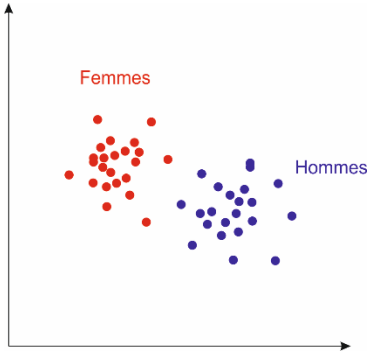
$$D_i = b_0 + \sum_{k=1}^p b_k X_k$$



Donnés

Le model linéaire - D  
(avec estimations des b)

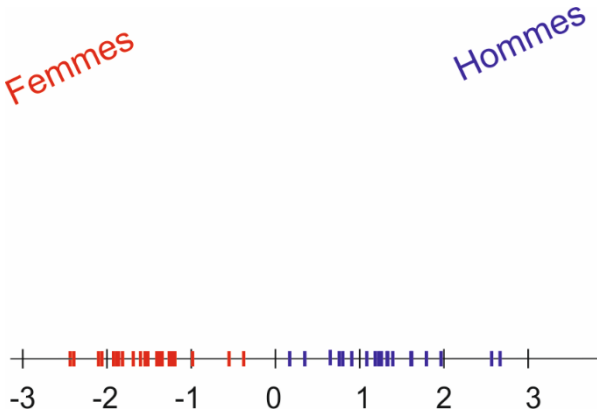
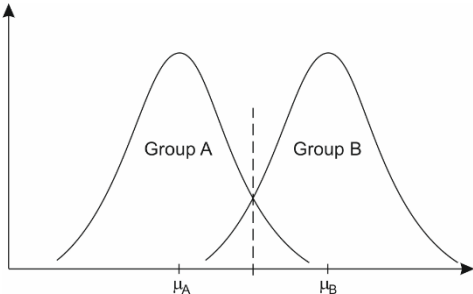
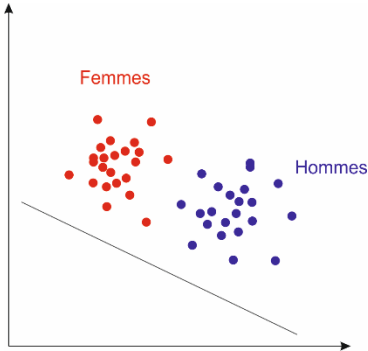
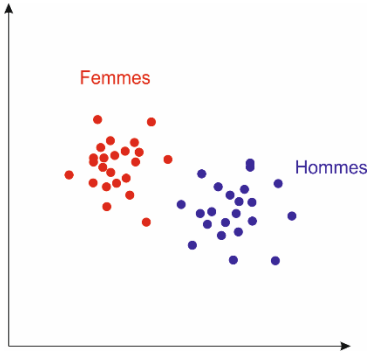
Visualiser les différences des groups



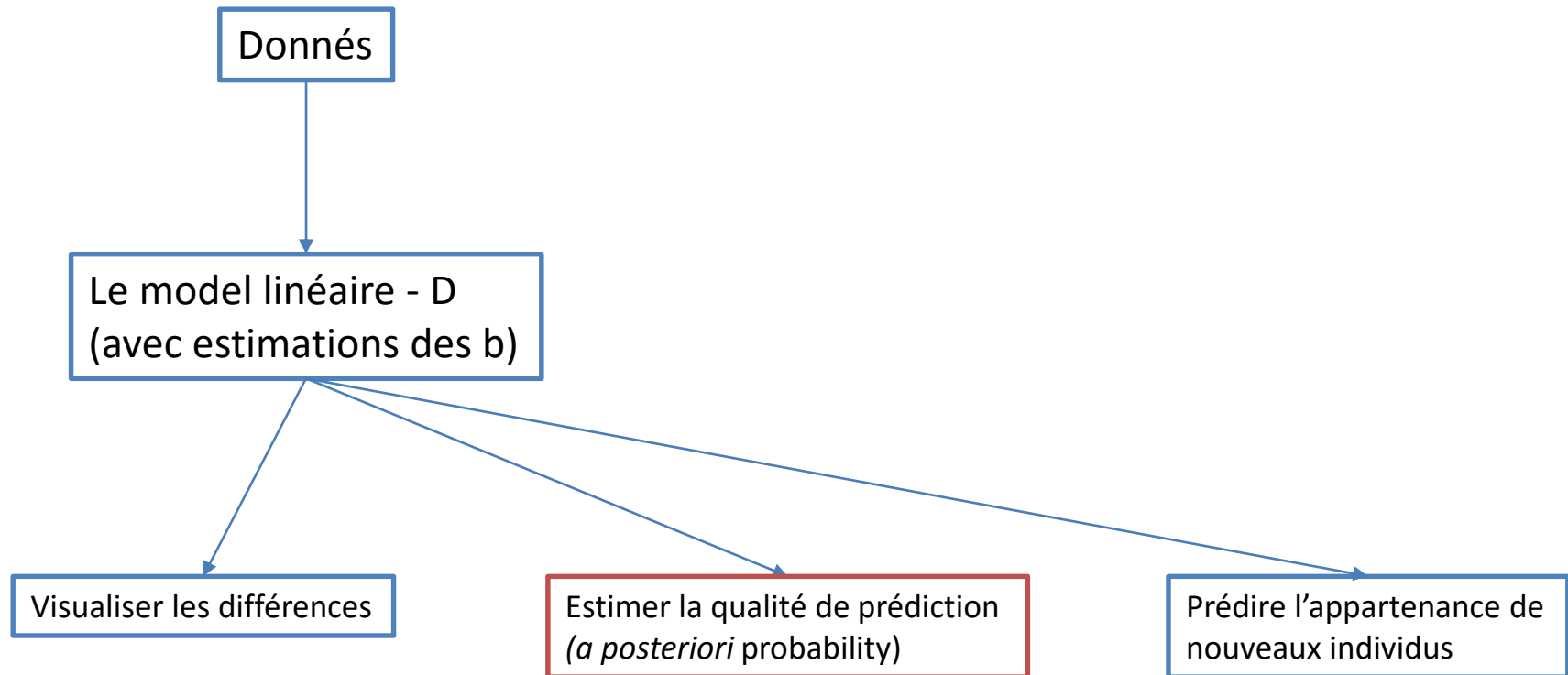
Donnés

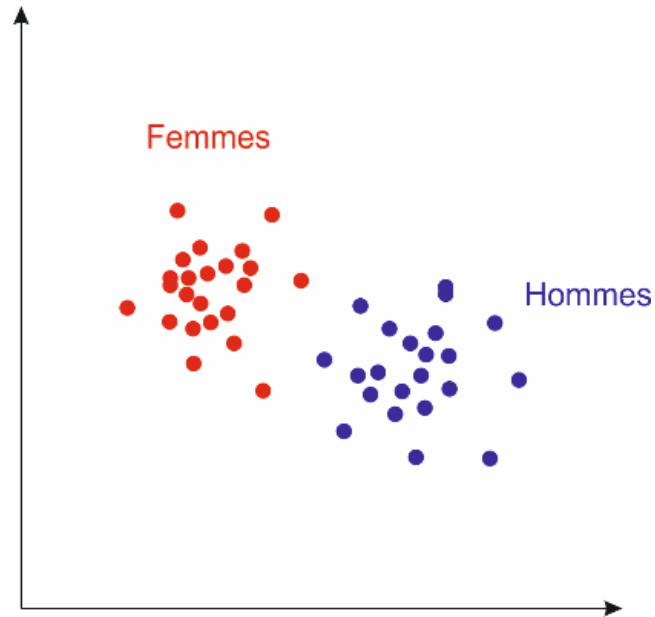
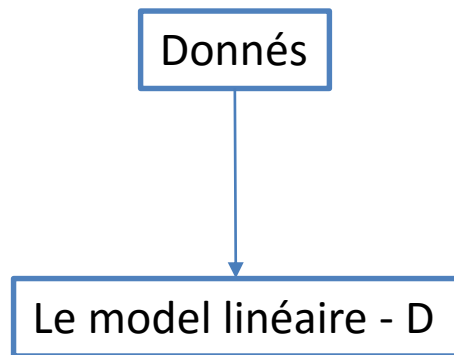
Le model linéaire - D  
(avec estimations des b)

Visualiser les différences des groups



### - Que-est ce qu'on peut faire avec LDA?





Estimer la qualité de prédiction  
(*a posteriori* probability)

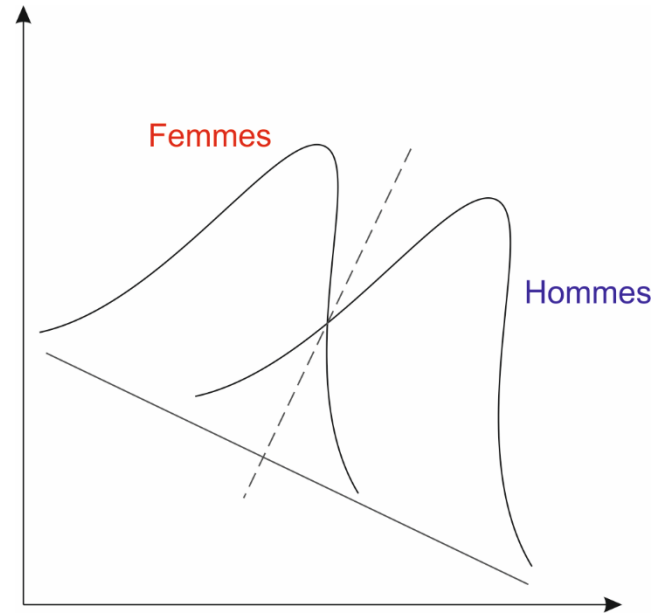
- Appliquer le model (comme prédicteur) sur les donnés
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

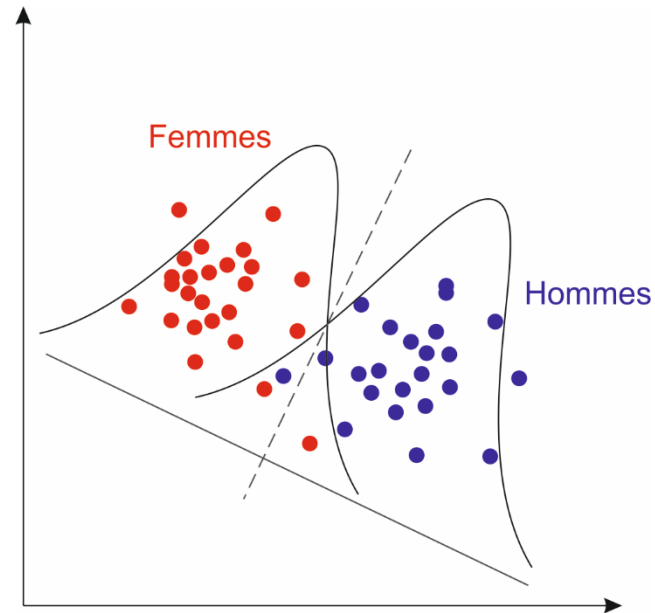
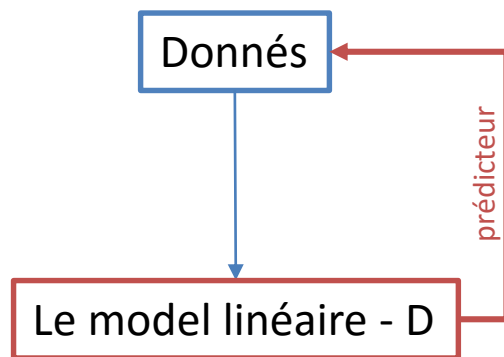
Donnés

Le model linéaire - D

Estimer la qualité de prédiction  
(*a posteriori* probability)

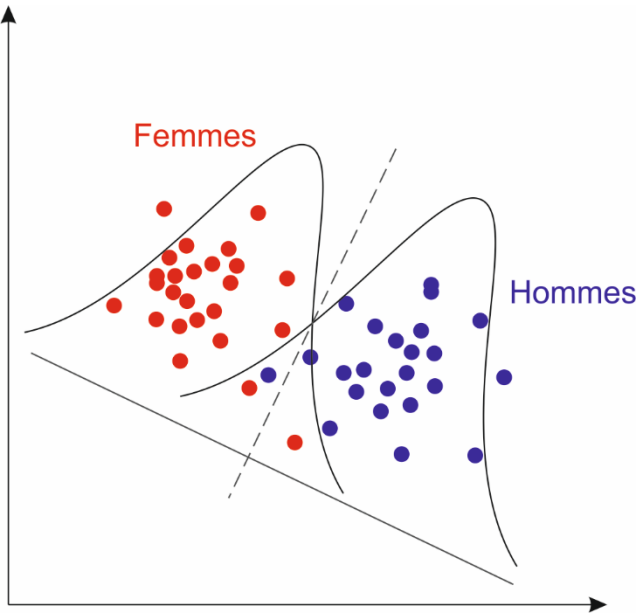
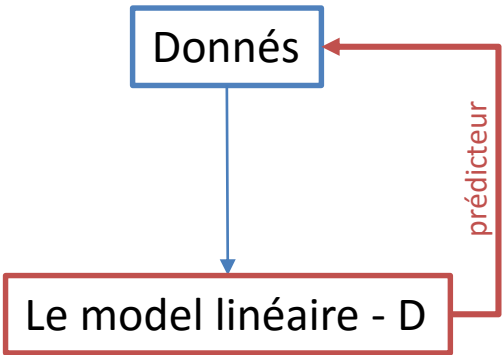
- Appliquer le model (comme prédicteur) sur les donnés
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions





Estimer la qualité de prédiction  
(*a posteriori* probability)

- Appliquer le model (comme prédicteur) sur les donnés
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

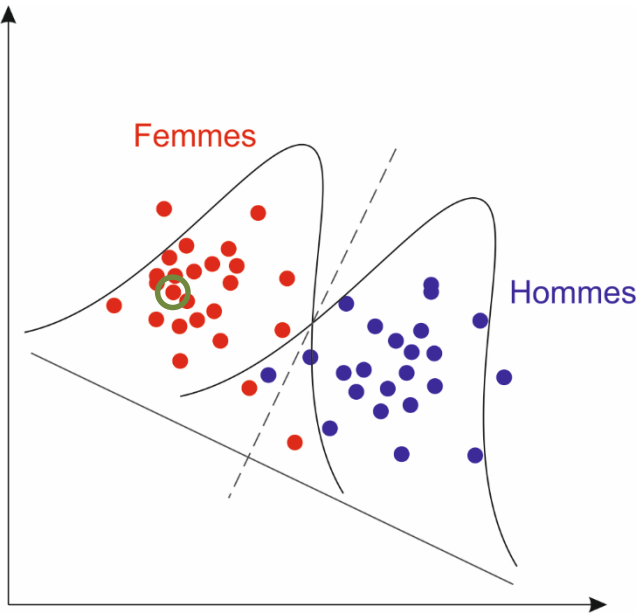
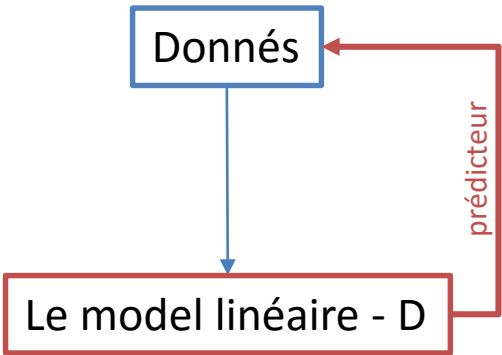


Estimer la qualité de prédiction  
(*a posteriori* probability)

- Appliquer le model (comme prédicteur) sur les données
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

	Original	Prédiction de model		
		Femme	Homme	Predicted
Ind1	Femme			
Ind2	Homme			
Ind3	Homme			
...				

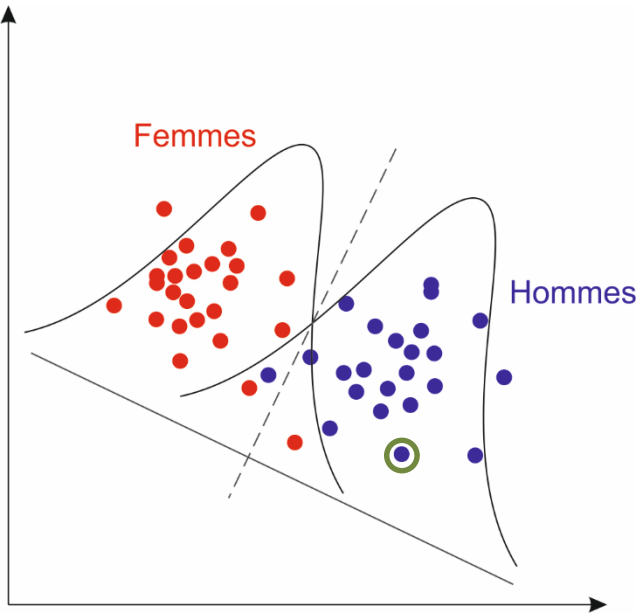
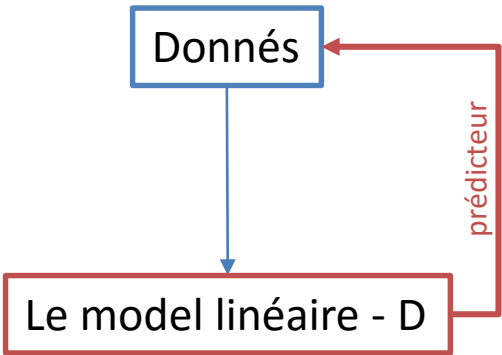




Estimer la qualité de prédiction  
(*a posteriori* probability)

- Appliquer le model (comme prédicteur) sur les données
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

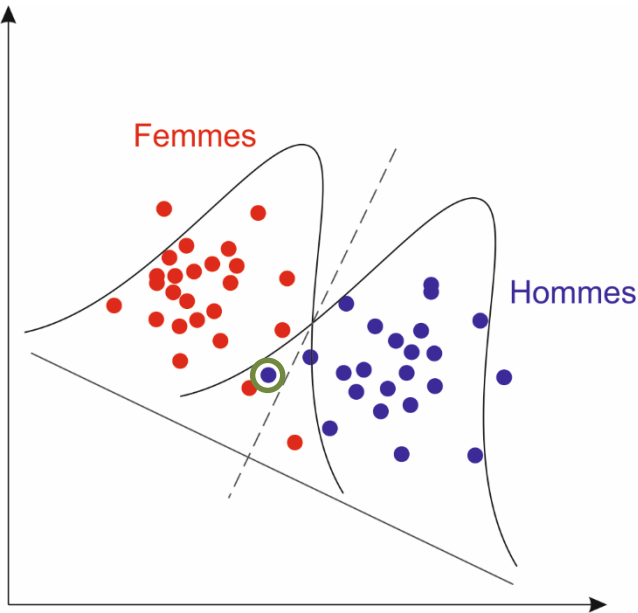
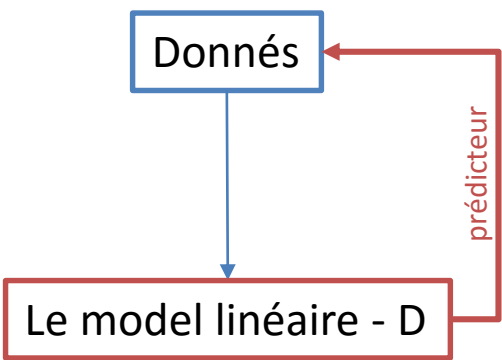
	Original	Prédiction de model		
		Femme	Homme	Predicted
Ind1	Femme	100	0	Femme
Ind2	Homme			
Ind3	Homme			
...				



Estimer la qualité de prédiction  
(*a posteriori* probability)

- Appliquer le model (comme prédicteur) sur les données
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

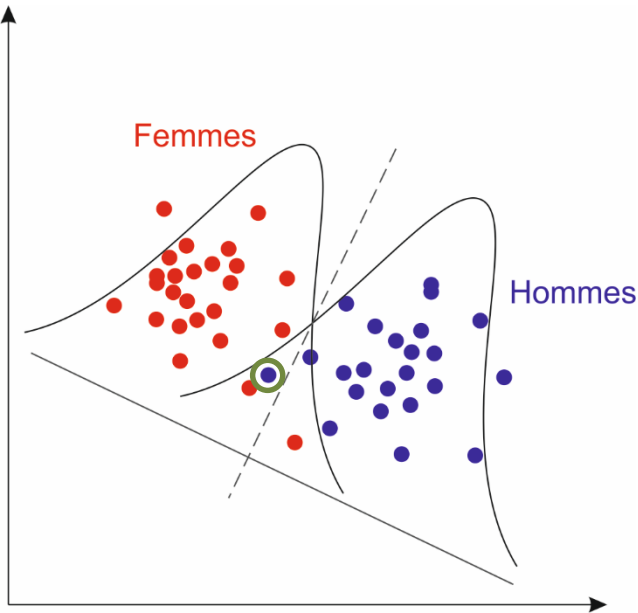
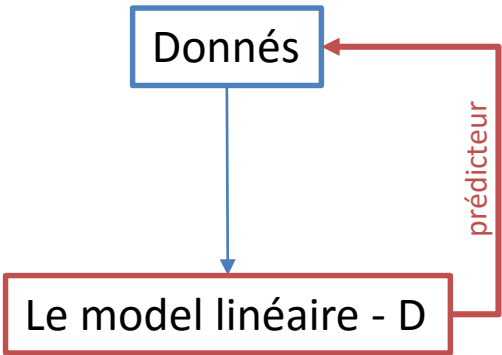
	Original	Prédiction de model		
		Femme	Homme	Predicted
Ind1	Femme	100	0	Femme
Ind2	Homme	0	100	Homme
Ind3	Homme			
...				



Estimer la qualité de prédiction  
(*a posteriori* probability)

- Appliquer le model (comme prédicteur) sur les données
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

	Original	Prédiction de model		
		Femme	Homme	Predicted
Ind1	Femme	100	0	Femme
Ind2	Homme	0	100	Homme
Ind3	Homme	60	40	Femme
...				

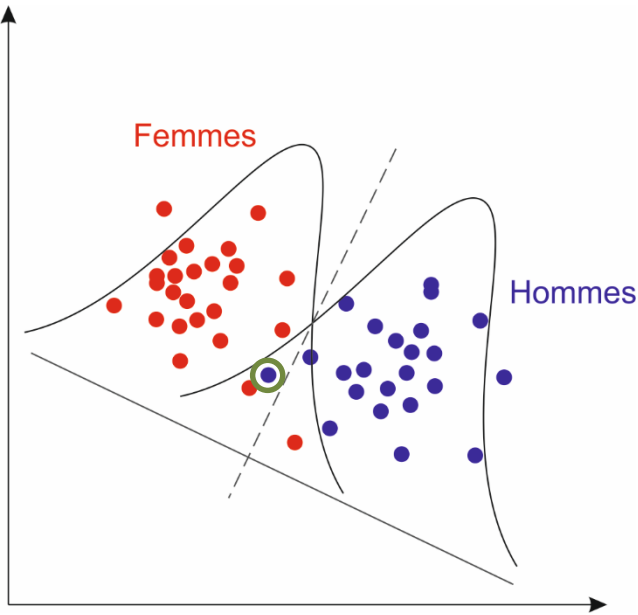
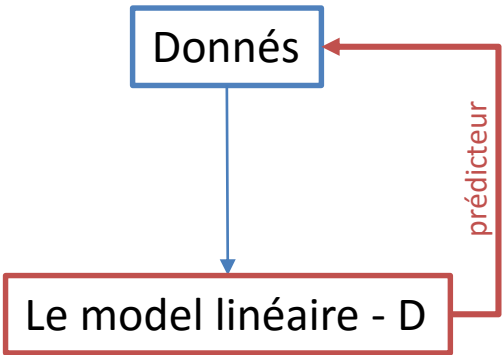


Estimer la qualité de prédiction  
(*a posteriori* probability)

- Appliquer le model (comme prédicteur) sur les donnés
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

	Original	Prédiction de model		
		Femme	Homme	Predicted
Ind1	Femme	100	0	Femme
Ind2	Homme	0	100	Homme
Ind3	Homme	60	40	Femme
...				

		Prédiction de model	
		Femmes	Hommes
Original	Femme	49	1
	Homme	3	47



Estimer la qualité de prédiction  
(*a posteriori* probability)

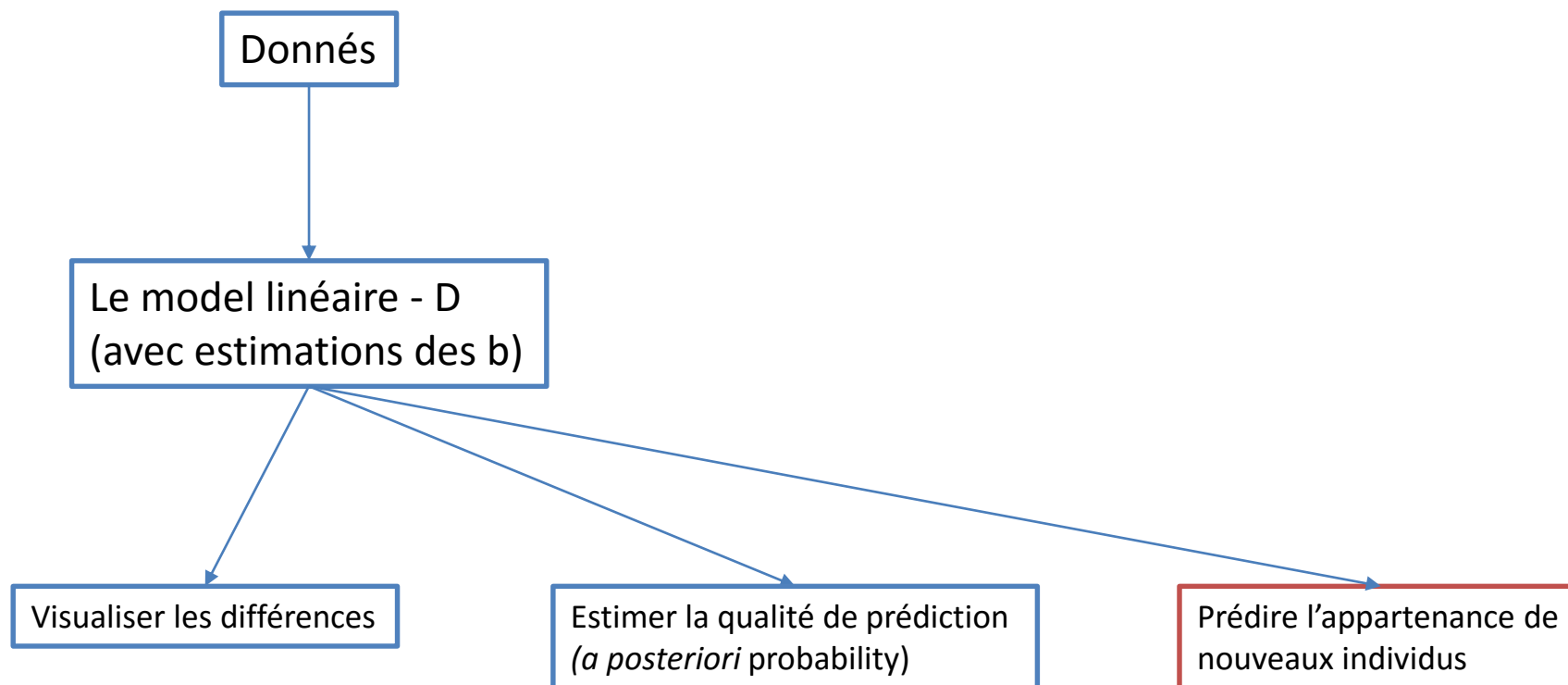
- Appliquer le model (comme prédicteur) sur les donnés
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

	Original	Prédiction de model		
		Femme	Homme	Predicted
Ind1	Femme	100	0	Femme
Ind2	Homme	0	100	Homme
Ind3	Homme	60	40	Femme
...				

		Prédiction de model	
		Femmes	Hommes
Original	Femme	49	1
	Homme	3	47

96% bien classés...

### - Que-est ce qu'on peut faire avec LDA?



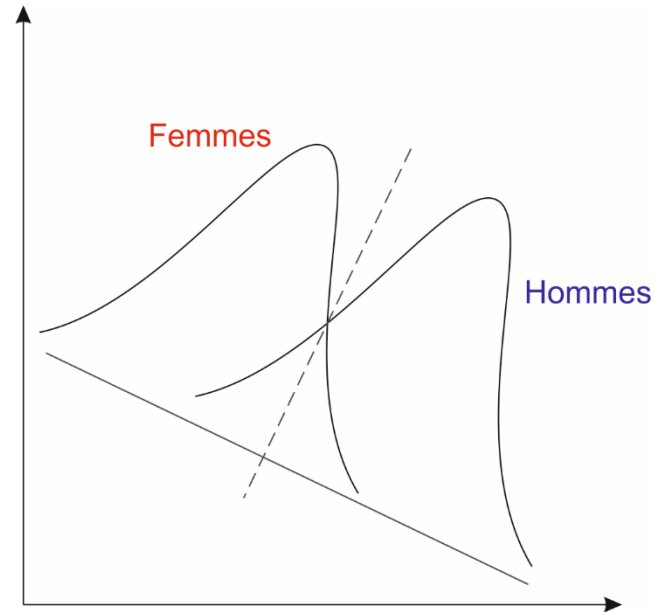
Donnés

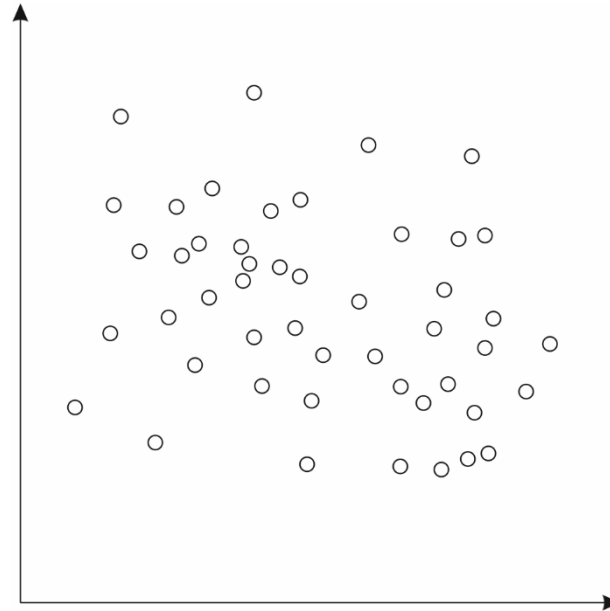
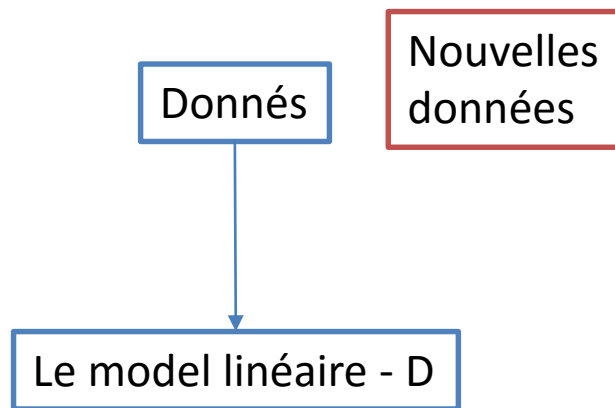


Le model linéaire - D

Prédire l'appartenance  
des nouveaux individus

- Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance
- Appliquer le model (comme prédicteur) sur les nouveau individus
- Regarder dans quels groupes les individus seront classées par le model

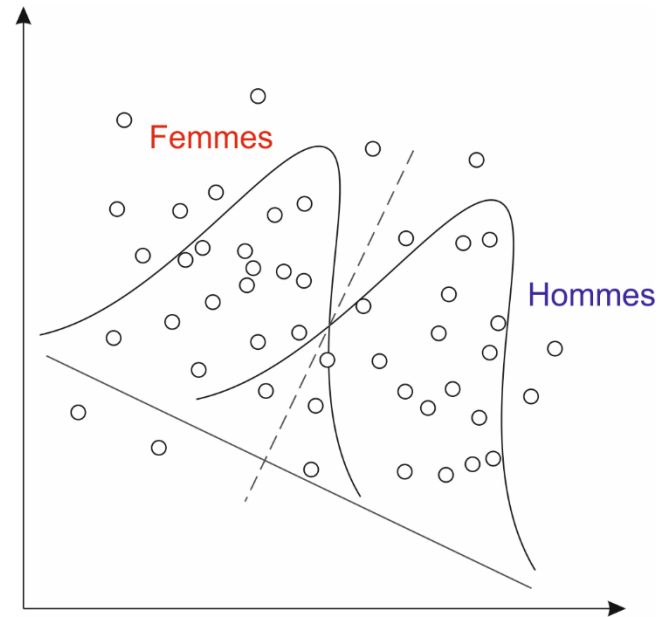
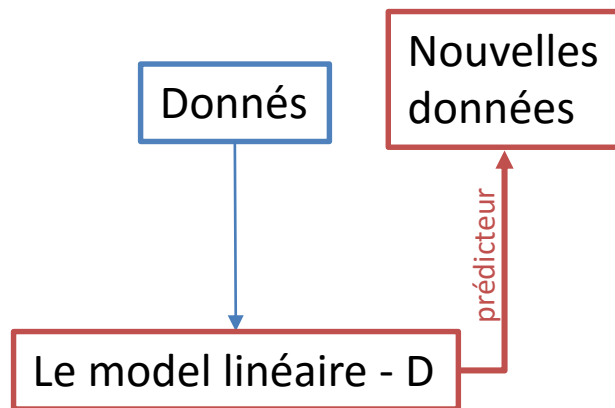




Prédire l'appartenance  
des nouveaux individus

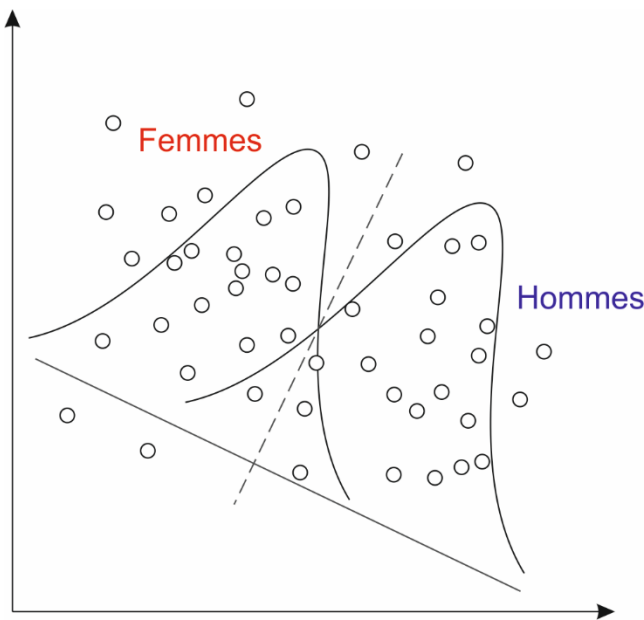
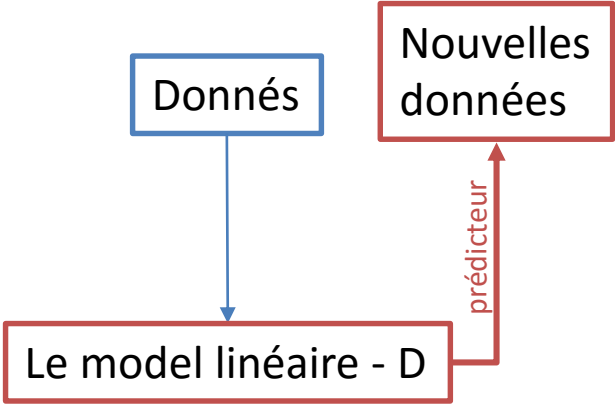
- **Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance**
- Appliquer le model (comme prédicteur) sur les nouveau individus
- Regarder dans quels groupes les individus seront classées par le model





Prédire l'appartenance  
des nouveaux individus

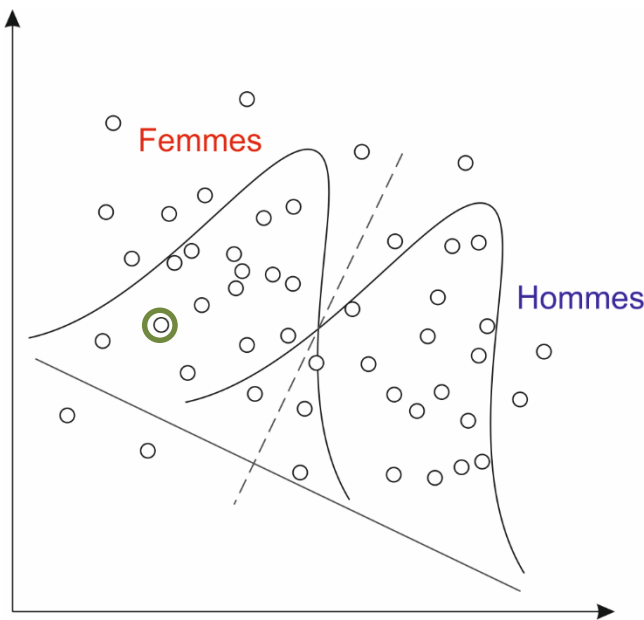
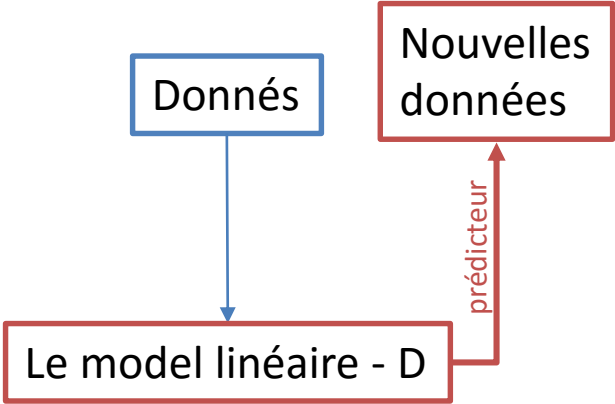
- Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance
- **Appliquer le model (comme prédicteur) sur les nouveau individus**
- Regarder dans quels groupes les individus seront classés par le model



Prédire l'appartenance des nouveaux individus

- Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance
- Appliquer le model (comme prédicteur) sur les nouveau individus
- **Regarder dans quels groupes les individus seront classés par le model**

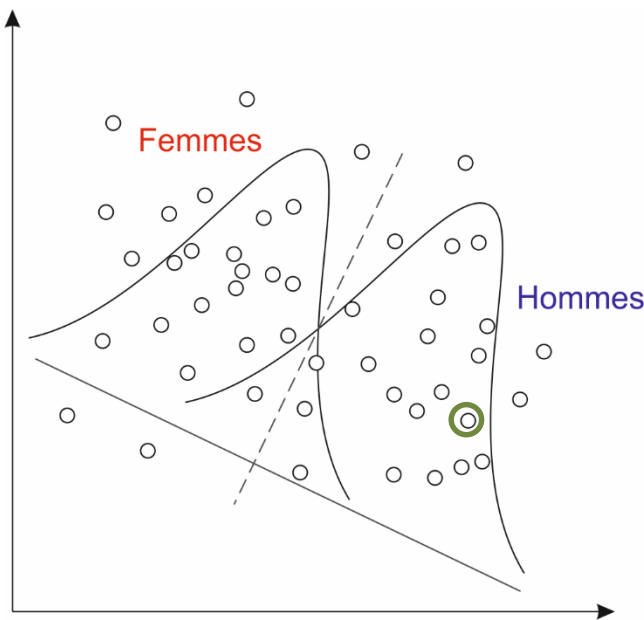
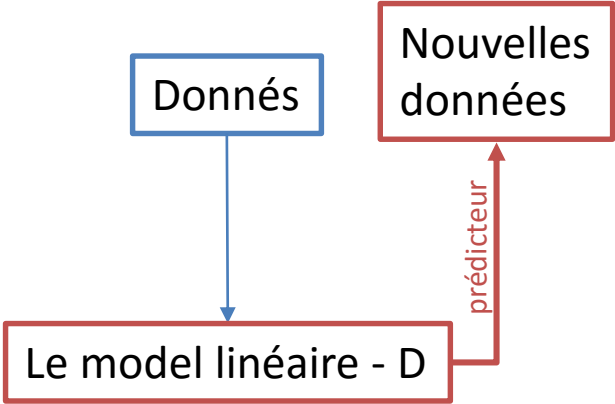
	Original	Predicted		
		Femme	Homme	Predicted
Ind1	-			
Ind2	-			
Ind3	-			
...				



Prédire l'appartenance des nouveaux individus

- Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance
- Appliquer le model (comme prédicteur) sur les nouveau individus
- **Regarder dans quels groupes les individus seront classés par le model**

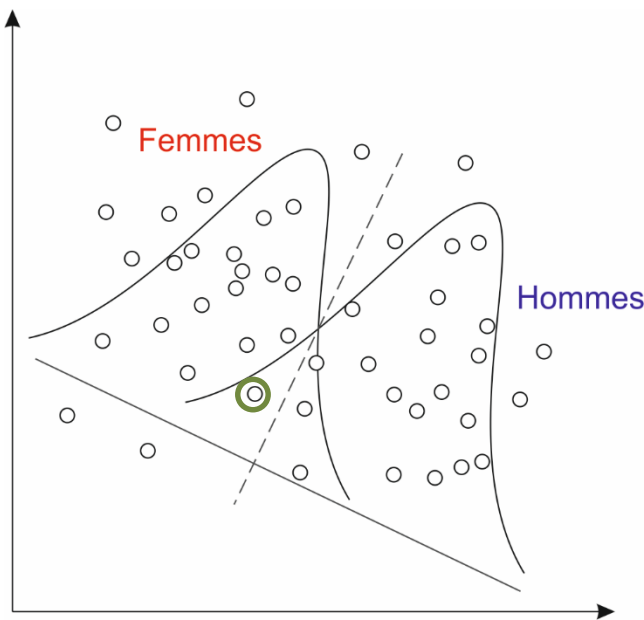
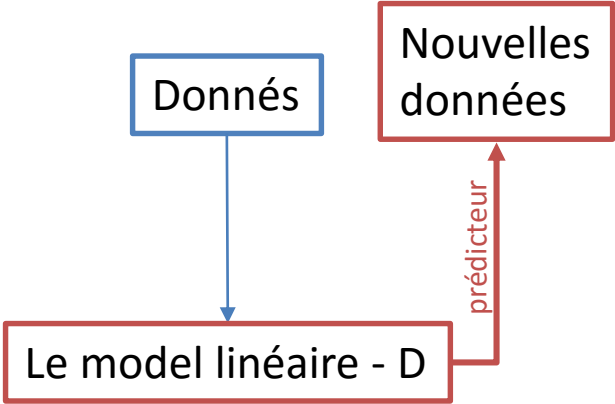
	Original	Predicted		
		Femme	Homme	Predicted
Ind1	-	100	0	Femme
Ind2	-			
Ind3	-			
...				



Prédire l'appartenance des nouveaux individus

- Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance
- Appliquer le model (comme prédicteur) sur les nouveau individus
- **Regarder dans quels groupes les individus seront classés par le model**

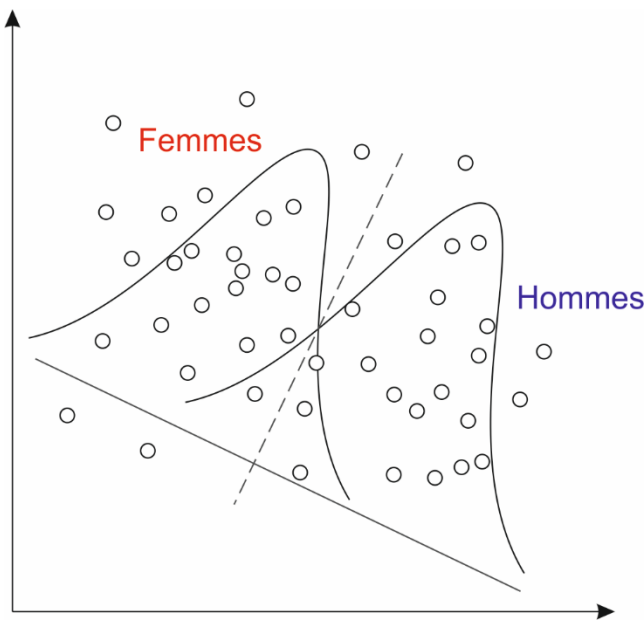
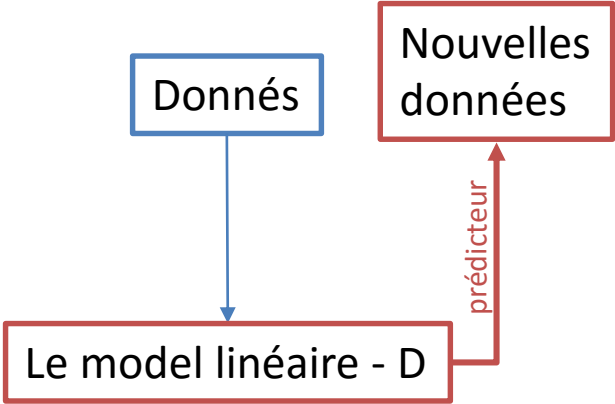
	Original	Predicted		
		Femme	Homme	Predicted
Ind1	-	100	0	Femme
Ind2	-	0	100	Homme
Ind3	-			
...				



Prédire l'appartenance des nouveaux individus

- Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance
- Appliquer le model (comme prédicteur) sur les nouveau individus
- **Regarder dans quels groupes les individus seront classés par le model**

	Original	Predicted		
		Femme	Homme	Predicted
Ind1	-	100	0	Femme
Ind2	-	0	100	Homme
Ind3	-	60	40	Femme
...				



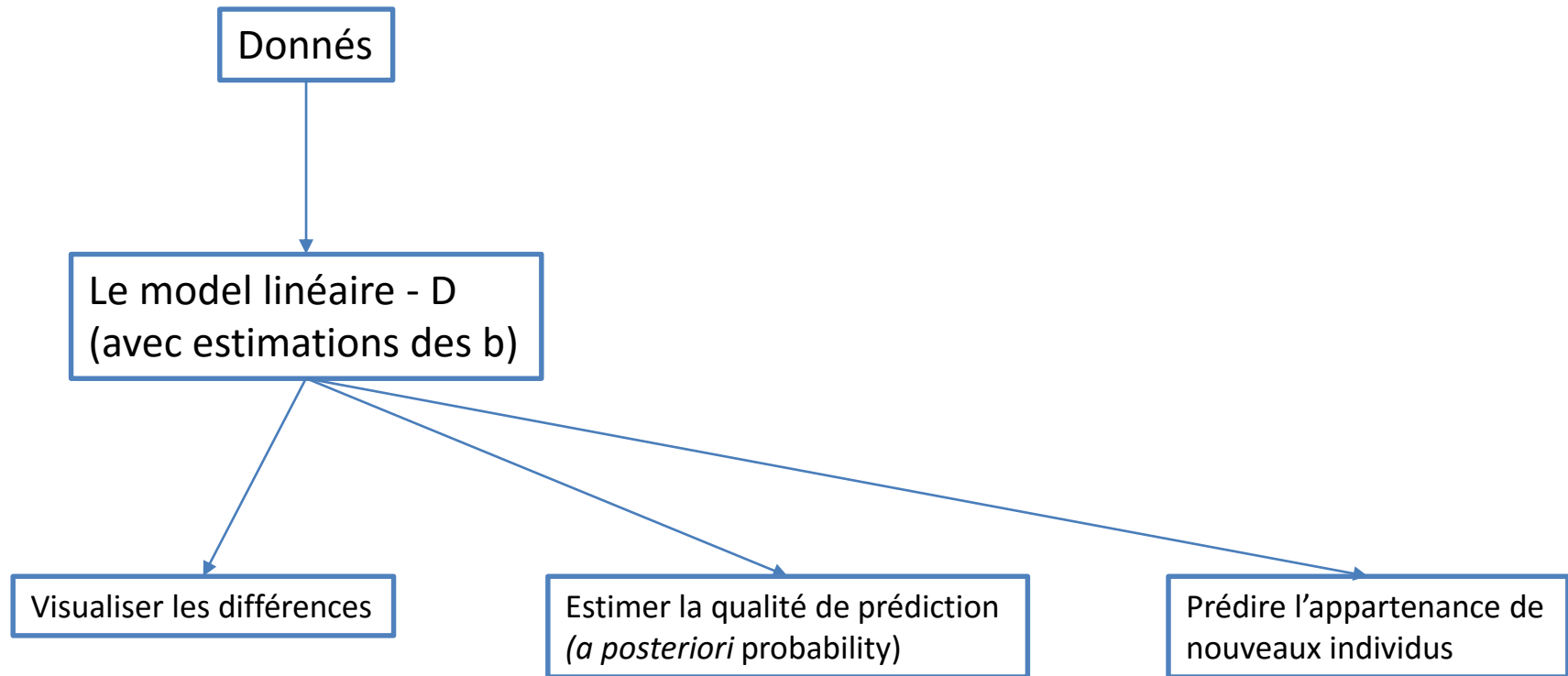
Prédire l'appartenance des nouveaux individus

- Injecter de nouveaux individus pour lesquels on ne connaît pas l'appartenance
- Appliquer le model (comme prédicteur) sur les nouveau individus
- **Regarder dans quels groupes les individus seront classés par le model**

	Original	Predicted		
		Femme	Homme	Predicted
Ind1	-	100	0	Femme
Ind2	-	0	100	Homme
Ind3	-	60	40	Femme
...				

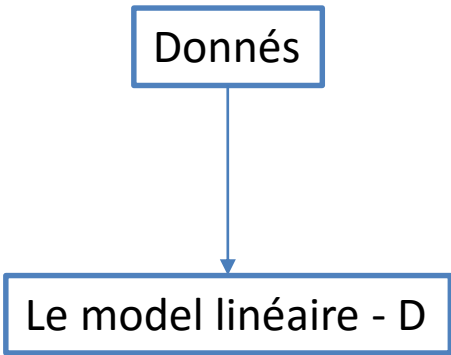
R

DA sur IRIS



On veut discriminer les trois espèces  
d'iris....





Visualiser les différences

```
# charger les données
data(iris)
iris
```

On veut discriminer les trois espèces d'iris....

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa

- On veut discriminer les espèces d'après les mesures
- Variable dépendante?
  - Variables indépendantes?
  - Nombre de fonctions discriminantes?

Donnés

Le model linéaire - D

Visualiser les différences

# charger les données

data(iris)

iris

# calculer le model linéaire

library(MASS)

iris.lda <- lda(data=iris, Species ~ Sepal.Length + Sepal.Width + Petal.Length  
+ Petal.Width)

iris.lda

Tilde

Call:

lda(Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width,  
data = iris)

Prior probabilities of groups:

setosa	versicolor	virginica
0.3333333	0.3333333	0.3333333

Group means:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	5.006	3.428	1.462	0.246
versicolor	5.936	2.770	4.260	1.326
virginica	6.588	2.974	5.552	2.026

Coefficients of linear discriminants:

	LD1	LD2
Sepal.Length	0.8293776	0.02410215
Sepal.Width	1.5344731	2.16452123
Petal.Length	-2.2012117	-0.93192121
Petal.Width	-2.8104603	2.83918785

$$D_i = b_0 + \sum_{k=1}^p b_k X_k$$

Proportion of trace:

LD1	LD2
0.9912	0.0088

Forte discrimination sur LD1

Donnés



Le model linéaire - D

# charger les données

```
data(iris)
```

```
iris
```

# calculer le model linéaire

```
library(MASS)
```

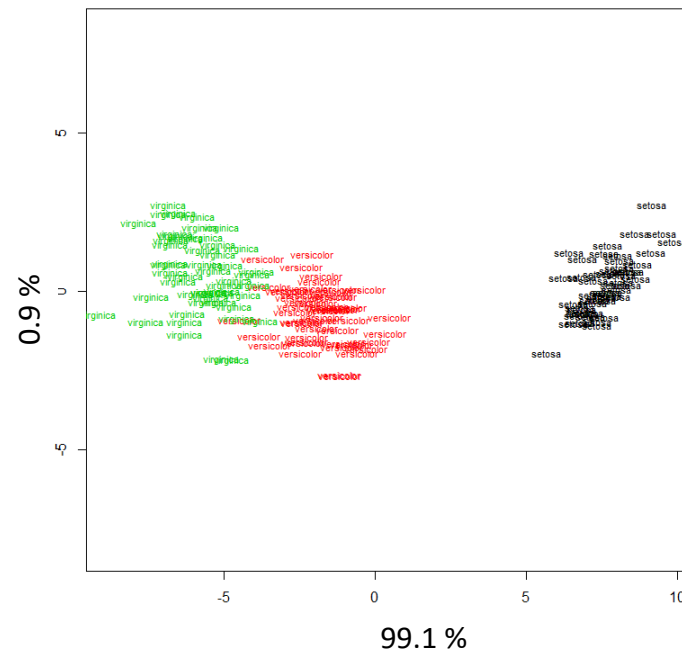
```
iris.Lda <- lda(data=iris, Species ~ Sepal.Length + Sepal.Width + Petal.Length  
+ Petal.Width)
```

```
iris.Lda
```

# visualiser les résultats

```
plot (iris.Lda, col=unclass(iris$Species))
```

Visualiser les différences



Donnés



Le model linéaire - D

Visualiser les différences

```
# charger les données
```

```
data(iris)
```

```
iris
```

```
# calculer le model linéaire
```

```
library(MASS)
```

```
iris.lda <- lda(data=iris, Species ~ Sepal.Length + Sepal.Width + Petal.Length  
+ Petal.Width)
```

```
iris.lda
```

```
# visualiser les résultats
```

```
plot (iris.lda, col=unclass(iris$Species))
```

```
abline(h=0, lty=2); abline(v=0, lty=2)
```

```
# obtenir les scores discriminantes
```

```
LD.scores <- predict(iris.lda)$x
```

```
# projeter les scores sur le graph
```

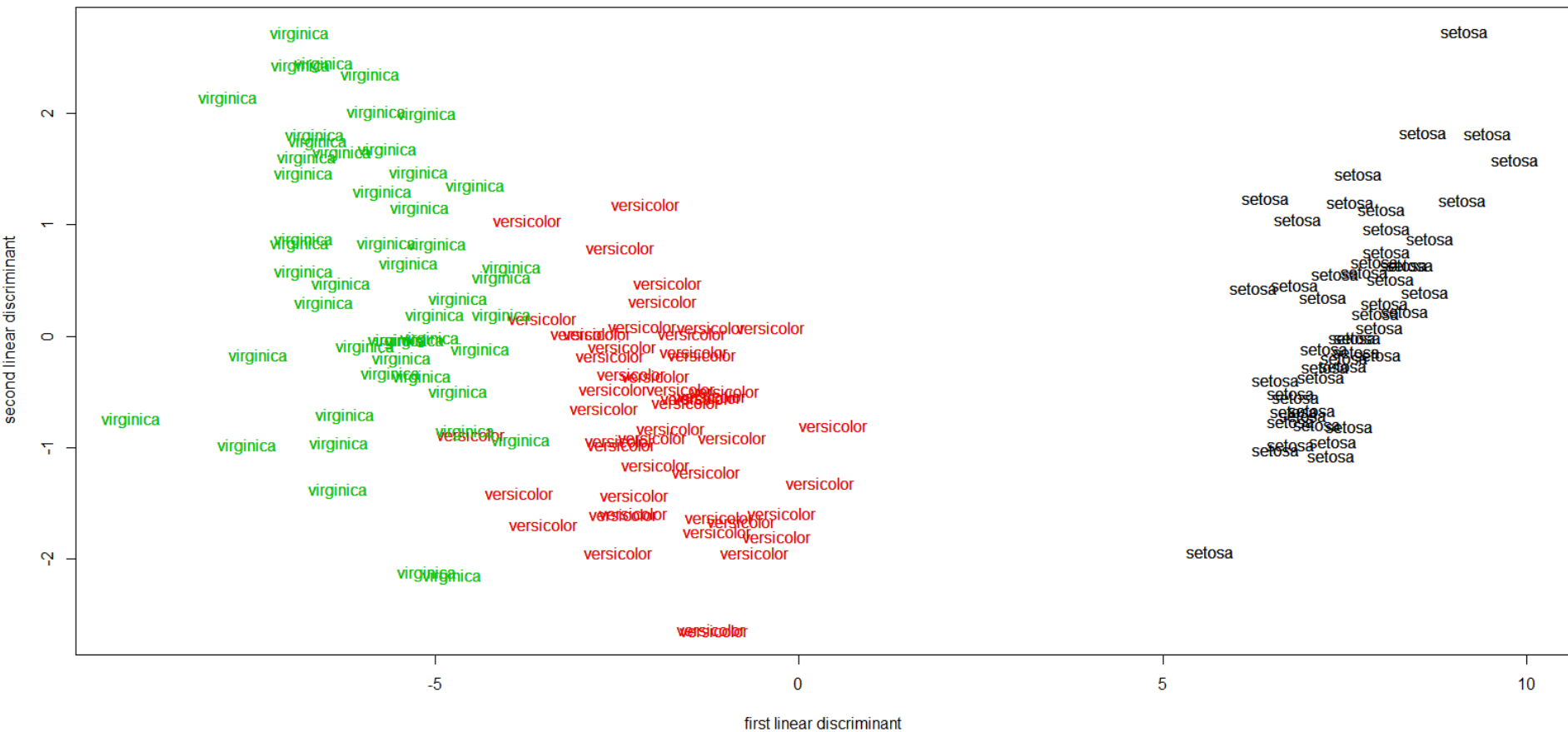
```
plot(LD.scores[,1], LD.scores[,2], xlab="first linear discriminant",
```

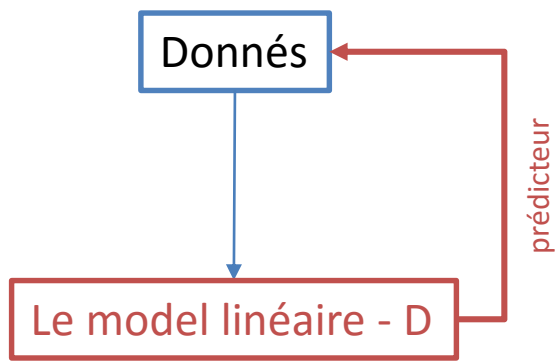
```
ylab="second linear discriminant", type="n")
```

```
text(LD.scores[,1], LD.scores[,2], labels=iris$Species,
```

```
col=unclass(iris$Species))
```

```
abline(h=0, lty=2); abline(v=0, lty=2)
```





Estimer la qualité de la prédiction  
(*a posteriori* probability)

- Appliquer le model (comme prédicteur) sur les donnés
- Regarder dans quels groupes les individus seront classés par le model
- Comparer les classes originales avec les prédictions

```
# utiliser le model (iris.lda) comme prédicteur sur les donnés (iris)
iris.pred <- predict(iris.lda, iris)

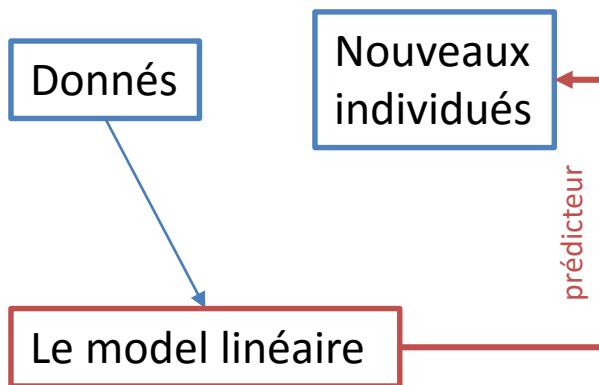
iris.pred$posterior # les posteriori probabilités pour chaque individu
                    # de tomber dans un des trois groupes
iris.pred$class     # le nouveau classement basé sur la plus grande
                    # probabilité
iris.pred$x         # les scores de la prédiction

# la qualité de la prédiction
cbind(Original=as.vector(iris$Species), Predicted=as.vector(iris.pred$class),
      BienClasse=iris$Species==iris.pred$class)
```

	Original	Predicted	BienClasse
[1,]	setosa	setosa	TRUE
[2,]	setosa	setosa	TRUE
[51,]	versicolor	versicolor	TRUE
[71,]	versicolor	virginica	FALSE
[134,]	virginica	versicolor	FALSE

```
table(Original=iris$Species, Predicted=iris.pred$class)
```

Original	Predicted		
	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	48	2
virginica	0	1	49



```
# créer les nouveau individus
new.data <- as.data.frame(sample(iris[,1:4]))
new.data[,1] <- sample(new.data[,1])
new.data[,2] <- sample(new.data[,2])
new.data[,3] <- sample(new.data[,3])
new.data[,4] <- sample(new.data[,4])
```

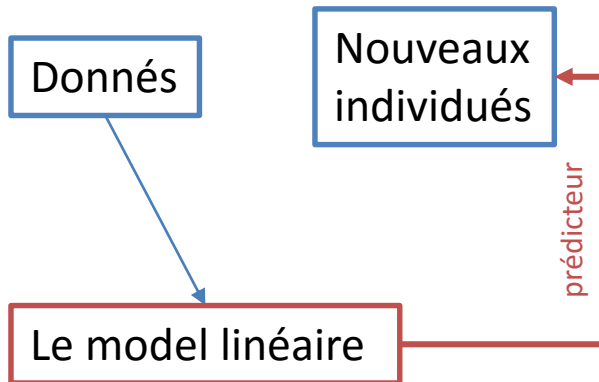
```
# utiliser le model (iris.lda) comme prédicteur sur les nouveau individus
(new data)
new.data.pred <- predict(iris.lda, new.data)
```

```
# projeter les individus originales sur le graph
plot(LD.scores[,1], LD.scores[,2], xlab="first linear discriminant",
     ylab="second linear discriminant", type="n")
abline(h=0, lty=2); abline(v=0, lty=2);
text(LD.scores[,1], LD.scores[,2], labels=iris$Species,
     col=unclass(iris$Species))
```

Prédire l'appartenance de nouveaux individus

- Appliquer le model (comme prédicteur) sur les nouveau individus
- Regarder dans quels groupes les individus seront classées par le model

# Traitement des données – Analyse discriminante



# créer les nouveau individus

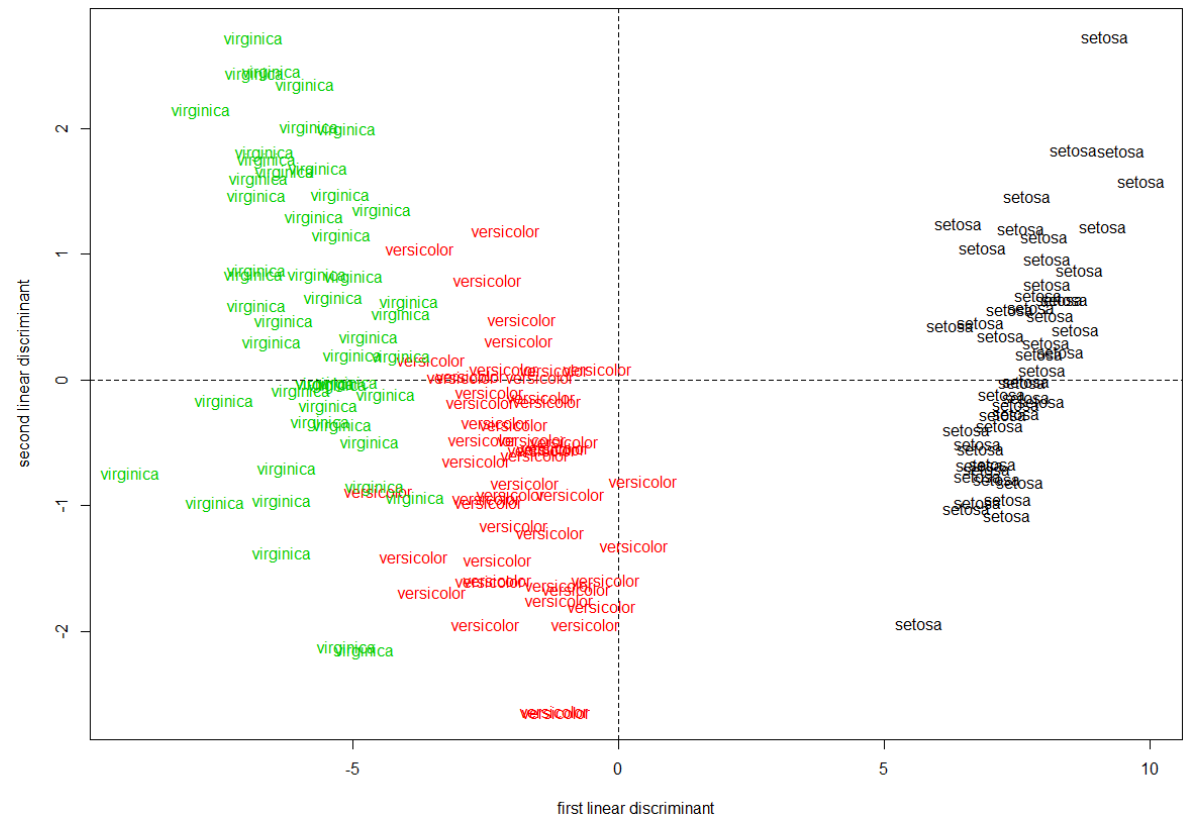
```
new.data <- as.data.frame(sample(iris[,1:4]))  
new.data[,1] <- sample(new.data[,1])  
new.data[,2] <- sample(new.data[,2])  
new.data[,3] <- sample(new.data[,3])  
new.data[,4] <- sample(new.data[,4])
```

# utiliser le model (iris.lda) comme prédicteur sur les nouveau individus (new data)

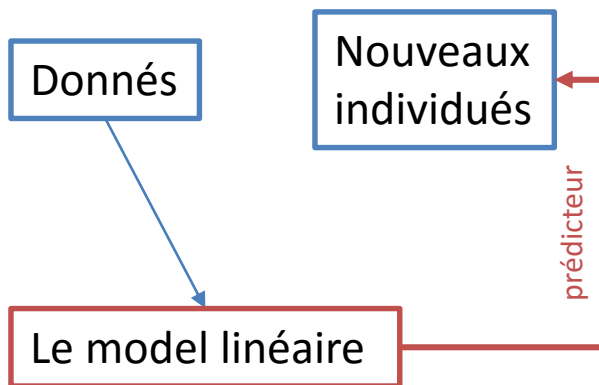
```
new.data.pred <- predict(iris.lda, new.data)
```

## Prédire l'appartenance de nouveaux individus

- Appliquer le model (comme prédicteur) sur les nouveau individus
- Regarder dans quels groupes les individus seront classées par le model







```
# créer les nouveau individus
new.data <- as.data.frame(sample(iris[,1:4]))
new.data[,1] <- sample(new.data[,1])
new.data[,2] <- sample(new.data[,2])
new.data[,3] <- sample(new.data[,3])
new.data[,4] <- sample(new.data[,4])
```

```
# utiliser le model (iris.lda) comme prédicteur sur les nouveau individus
(new data)
new.data.pred <- predict(iris.lda, new.data)
```

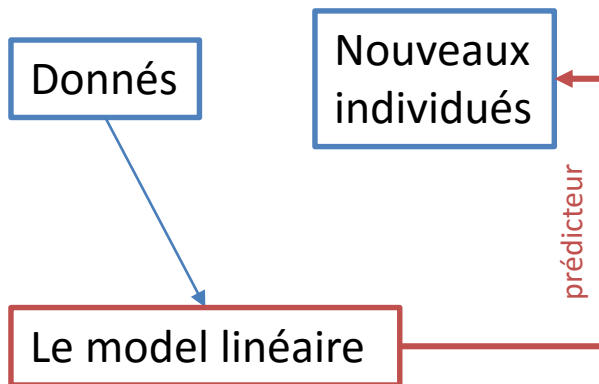
```
# projeter les individus originales sur le graph
plot(LD.scores[,1], LD.scores[,2], xlab="first linear discriminant",
     ylab="second linear discriminant", type="n")
abline(h=0, lty=2); abline(v=0, lty=2);
text(LD.scores[,1], LD.scores[,2], labels=iris$Species,
     col=unclass(iris$Species))
```

```
# injecter les nouveau individus dans le graph
points(new.data.pred$x, col=unlist(new.data.pred$class), pch=19)
```

Prédire l'appartenance de nouveaux individus

- Appliquer le model (comme prédicteur) sur les nouveau individus
- Regarder dans quels groupes les individus seront classées par le model

# Traitement des données – Analyse discriminante



# créer les nouveau individus

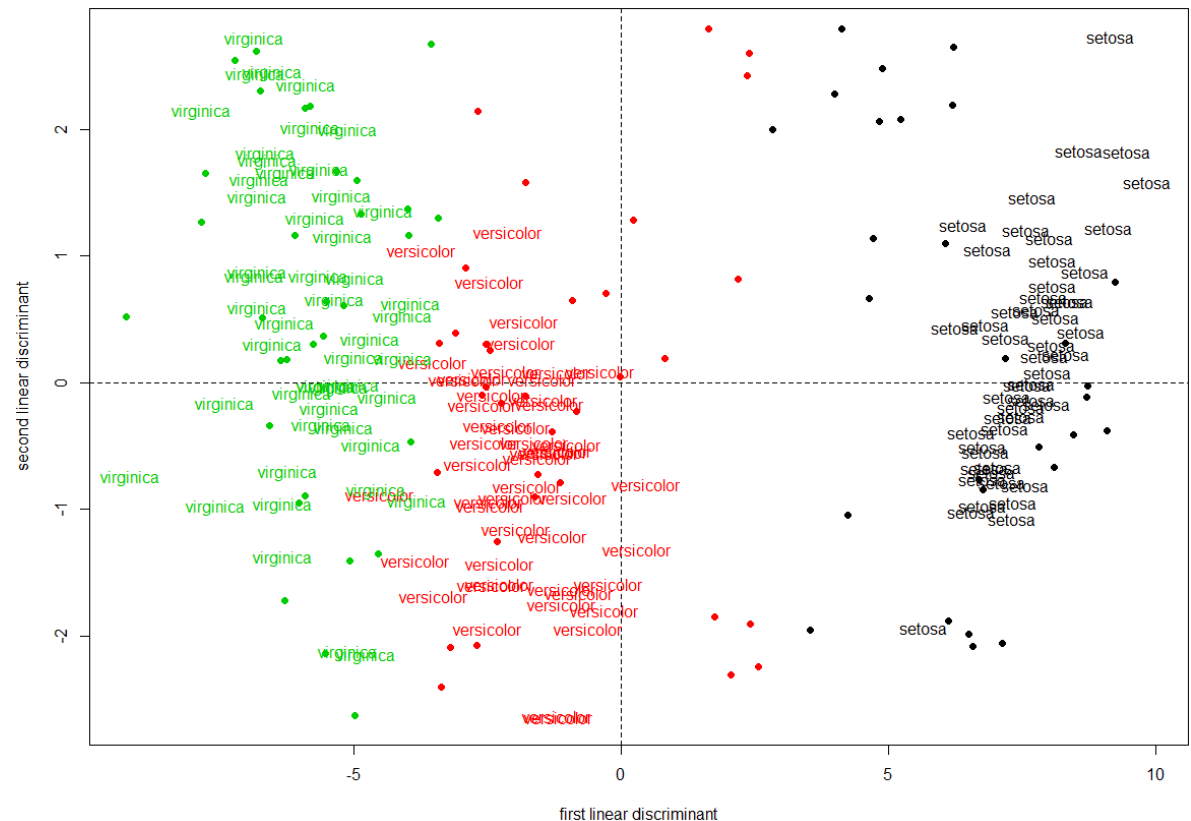
```
new.data <- as.data.frame(sample(iris[,1:4]))
new.data[,1] <- sample(new.data[,1])
new.data[,2] <- sample(new.data[,2])
new.data[,3] <- sample(new.data[,3])
new.data[,4] <- sample(new.data[,4])
```

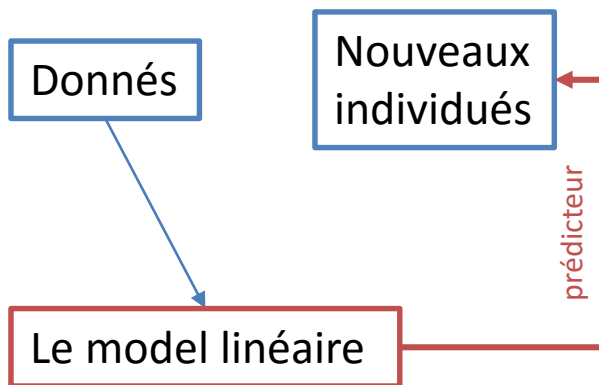
# utiliser le model (iris.lda) comme prédicteur sur les nouveau individus (new data)

```
new.data.pred <- predict(iris.lda, new.data)
```

## Prédire l'appartenance de nouveaux individus

- Appliquer le model (comme prédicteur) sur les nouveau individus
- Regarder dans quels groupes les individus seront classées par le model





```
# créer les nouveau individus
new.data <- as.data.frame(sample(iris[,1:4]))
new.data[,1] <- sample(new.data[,1])
new.data[,2] <- sample(new.data[,2])
new.data[,3] <- sample(new.data[,3])
new.data[,4] <- sample(new.data[,4])
```

```
# utiliser le model (iris.lda) comme prédicteur sur les nouveau individus
(new data)
new.data.pred <- predict(iris.lda, new.data)
```

```
# projeter les individus originales sur le graph
plot(LD.scores[,1], LD.scores[,2], xlab="first linear discriminant",
     ylab="second linear discriminant", type="n")
abline(h=0, lty=2); abline(v=0, lty=2);
text(LD.scores[,1], LD.scores[,2], labels=iris$Species,
     col=unclass(iris$Species))
```

```
# injecter les nouveau individus dans le graph
points(new.data.pred$x, col=unlist(new.data.pred$class), pch=19)
```

```
# on peut regarder le nouveau classement ...
table(new.data.pred$class)
```

setosa	versicolor	virginica
46	69	35

## Prédire l'appartenance de nouveaux individus

- Appliquer le model (comme prédicteur) sur les nouveau individus
- Regarder dans quels groupes les individus seront classées par le model