

# Détection automatique de faux billets



# Plan

## → Description des données

- ◆ Univariée et bivariée

- ◆ Multivariée

## → Classification

- ◆ Non-supervisée

- ◆ Supervisée

→ Programme de détection de faux billets

# Plan

## → Description des données

- ◆ Univariée et bivariée

- ◆ Multivariée

## → Classification

- ◆ Non-supervisée

- ◆ Supervisée

→ Programme de détection de faux billets

# Univariée

170 billets avec 7 variables

1 Qualitative

is\_genuine

100 vrais billets

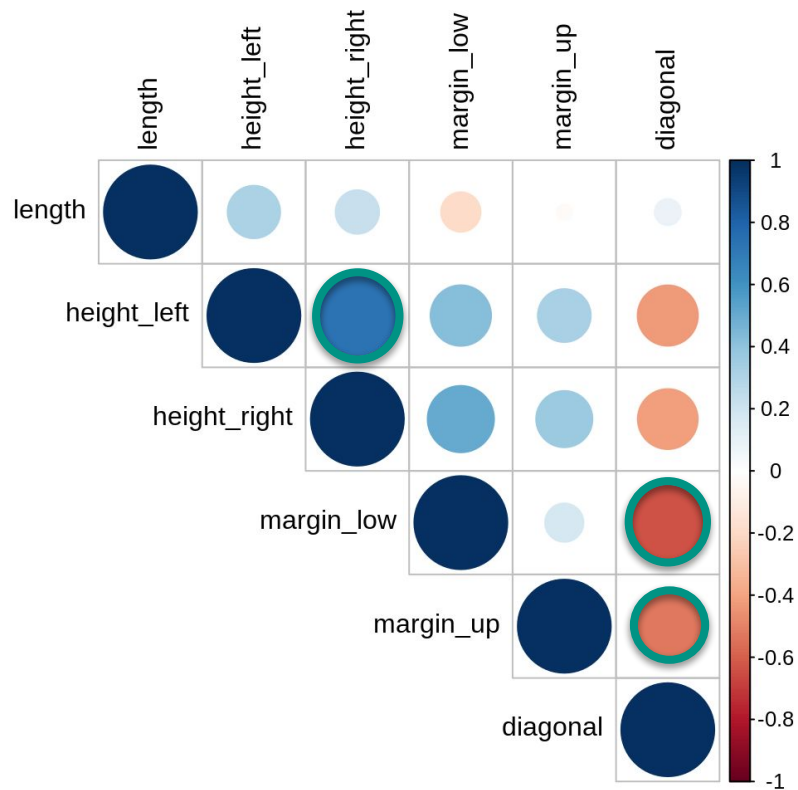
70 faux billets

6 Quantitatives

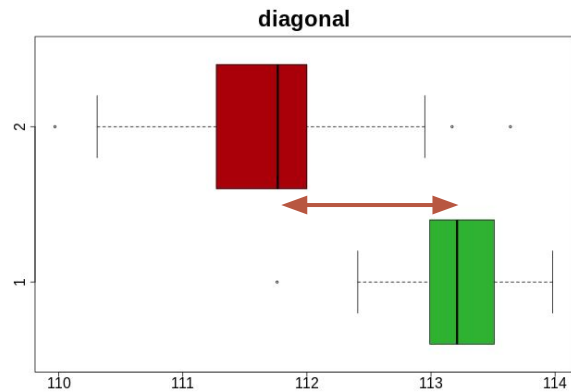
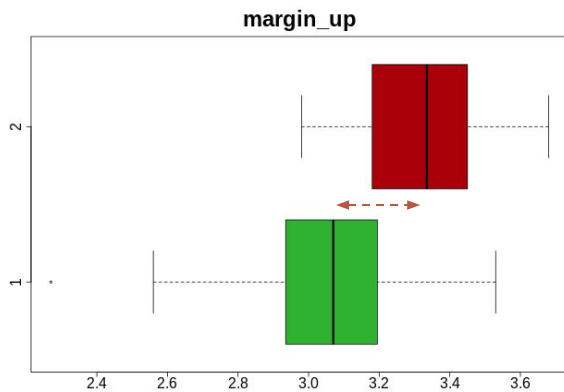
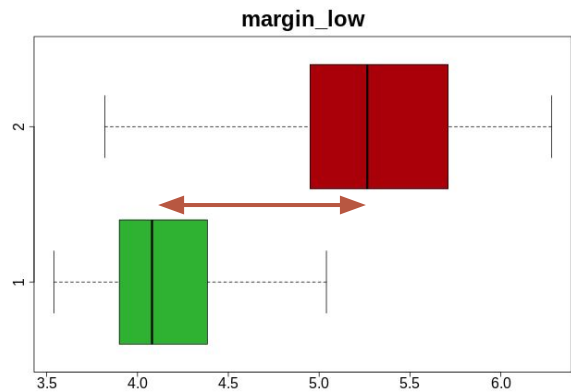
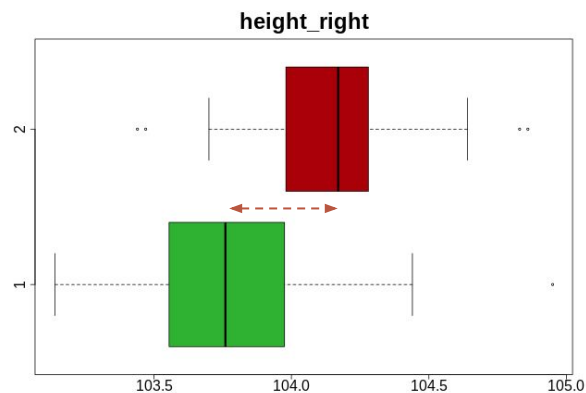
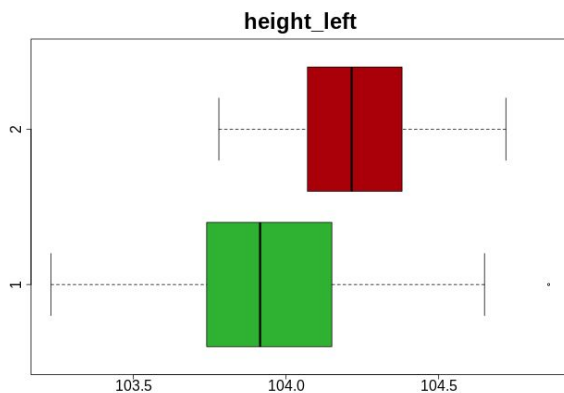
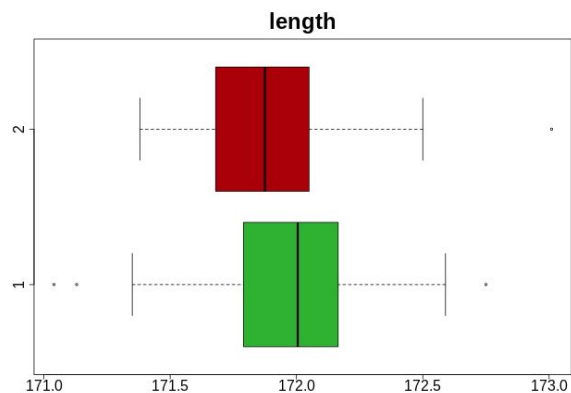


# Bivariée

Matrice de corrélation



# Bivariée Boîtes à moustaches



# Plan

## → Description des données

- ◆ Univariée et bivariée

- ◆ Multivariée

## → Classification

- ◆ Non-supervisée

- ◆ Supervisée

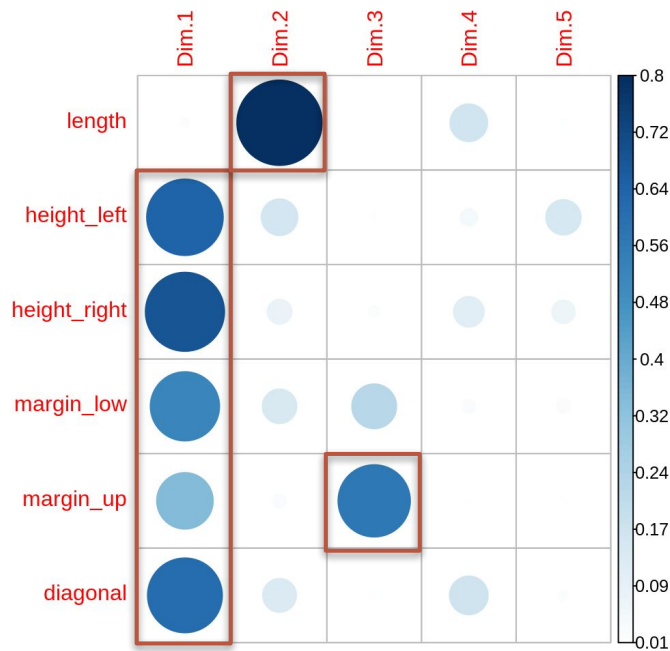
→ Programme de détection de faux billets

# Multivariée

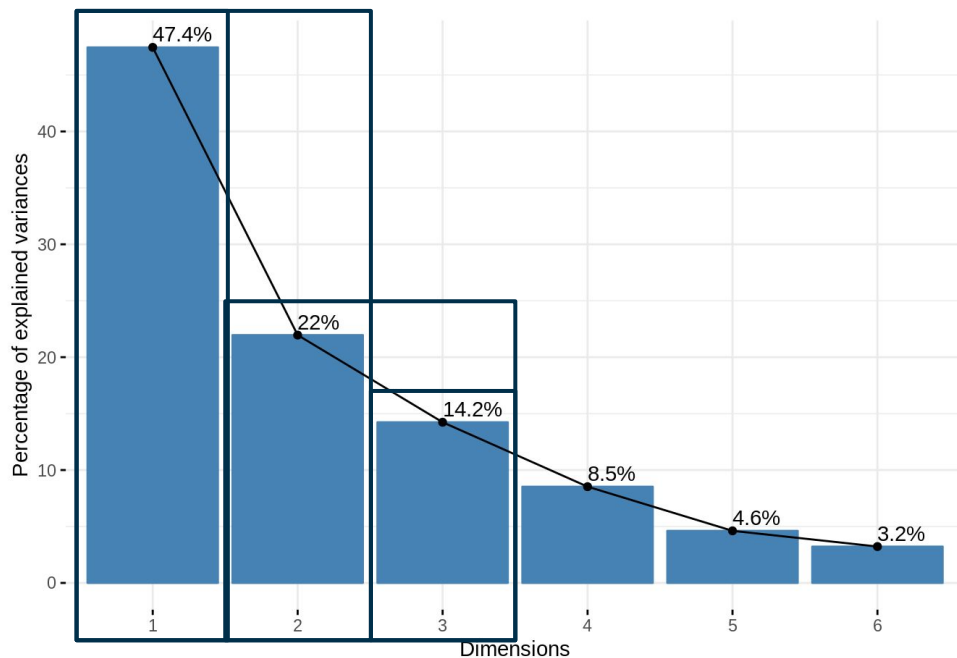
## Analyse en Composante Principale

→ *PCA()* de *factomineR*

Répartition des cos2



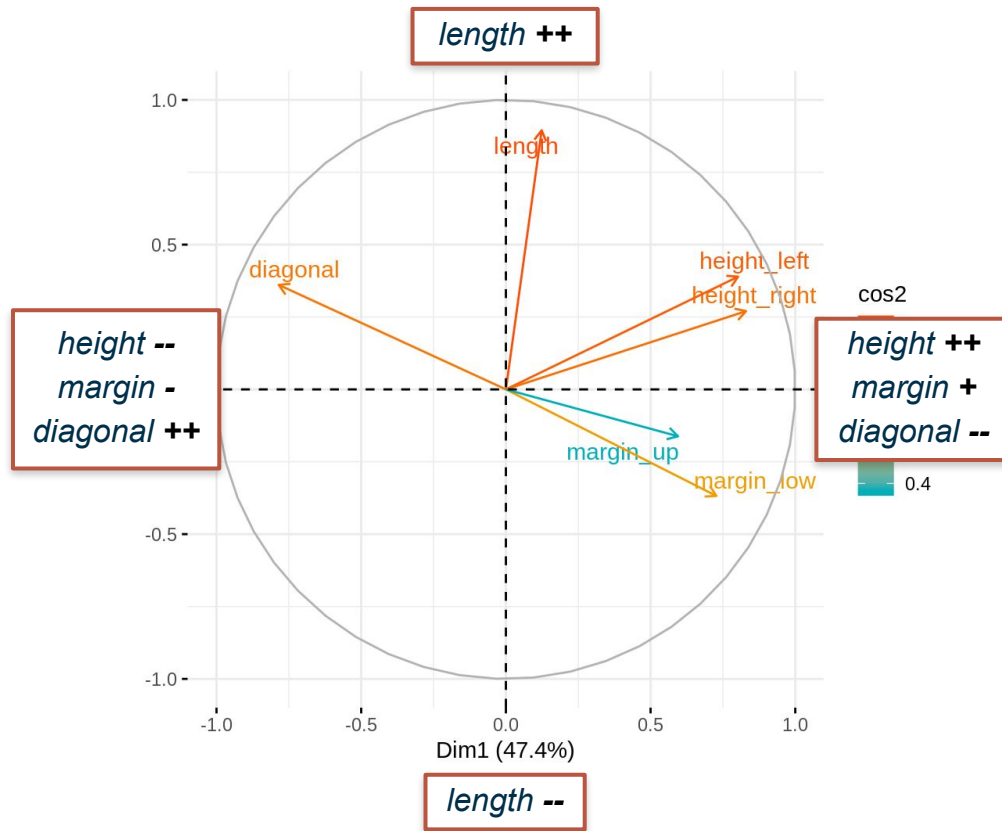
Éboulis des valeurs propres



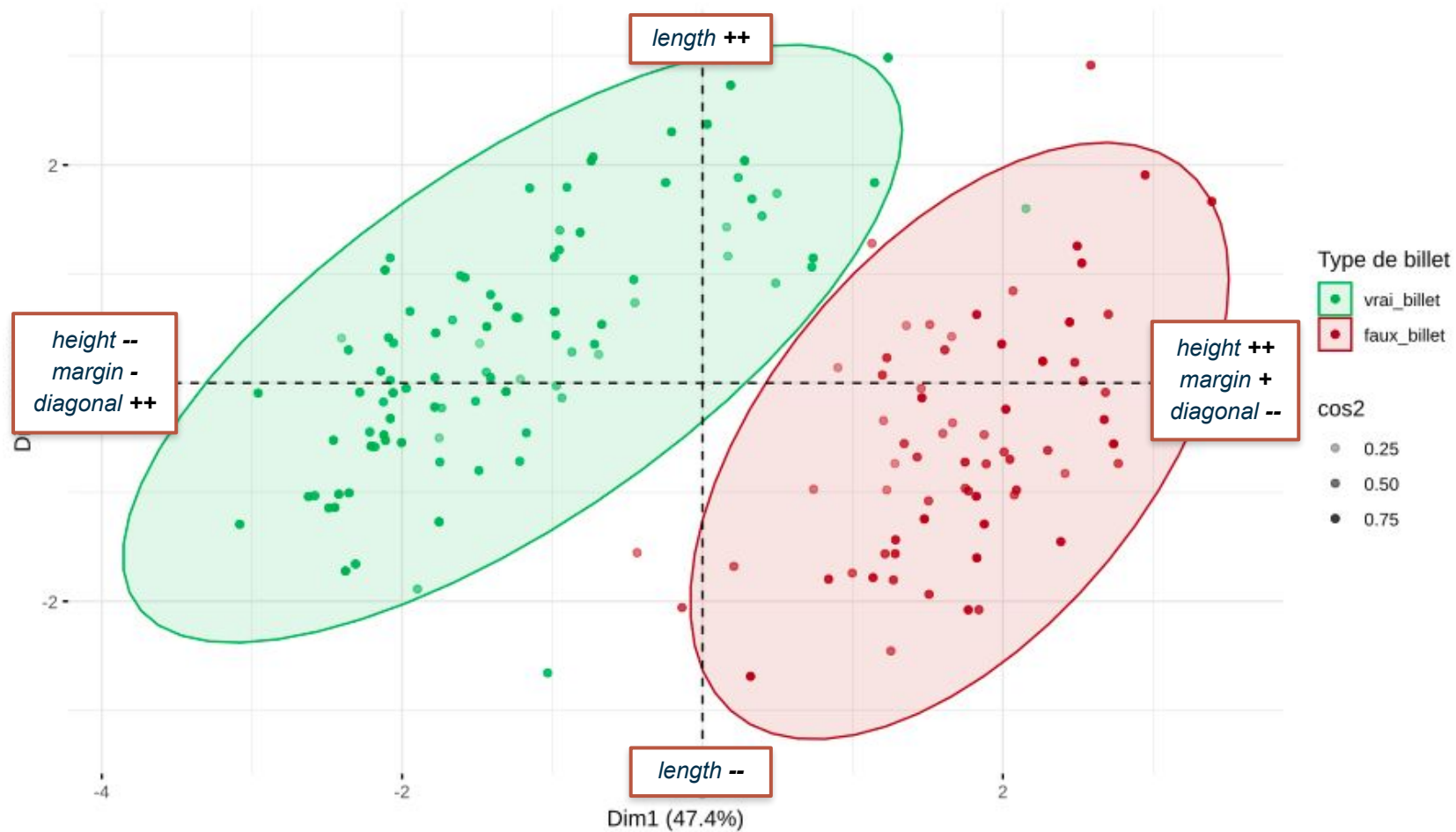


# Multivariée

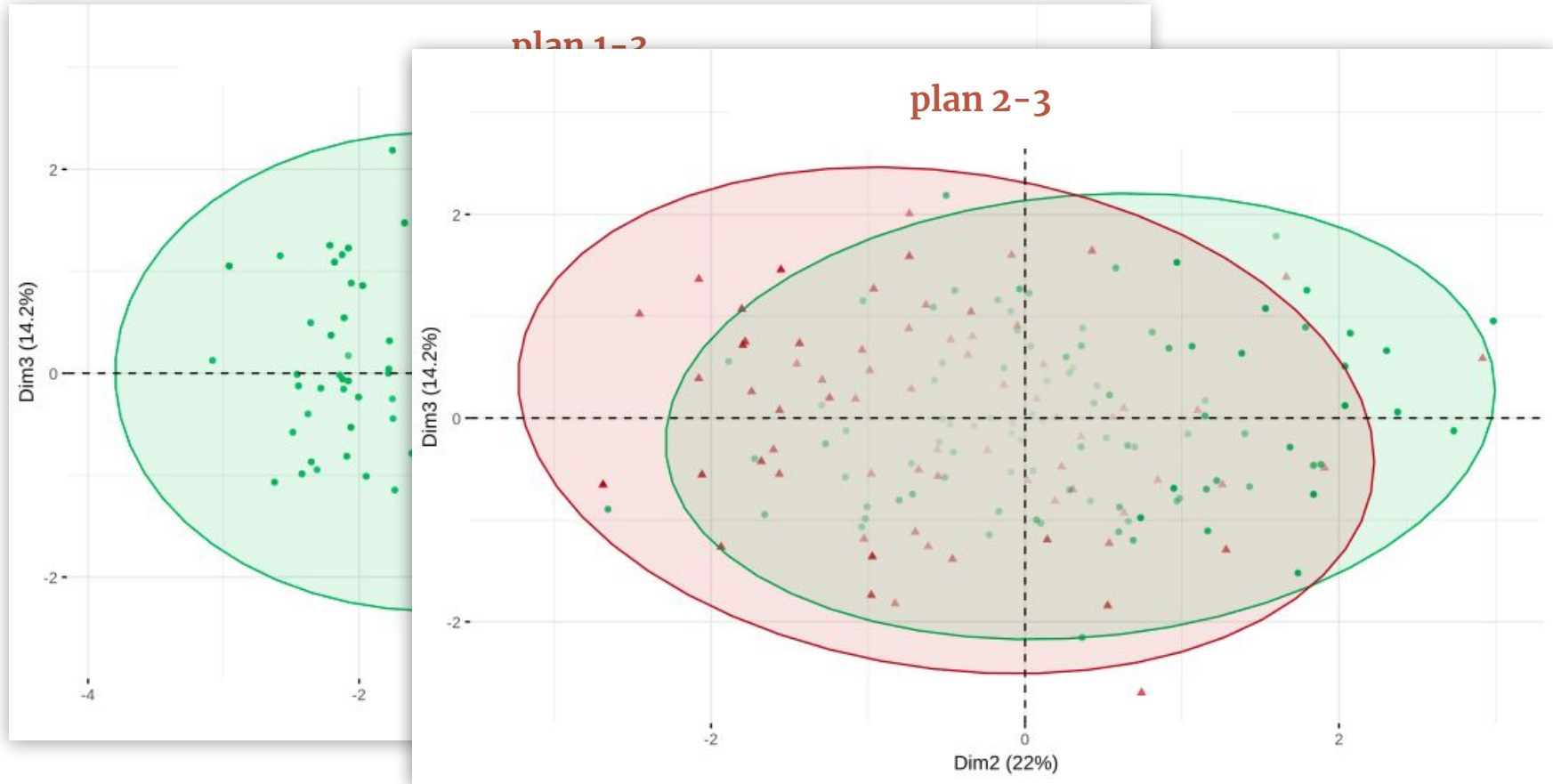
## Plan 1-2 Cercle de corrélation des variables



# Multivariée Plan 1-2 Nuage des individus



# Multivariée Plans suivants



# Plan

## → Description des données

- ◆ Univariée et bivariée

- ◆ Multivariée

## → Classification

- ◆ Non-supervisée

- ◆ Supervisée

→ Programme de détection de faux billets

# Non-supervisée **Clustering Ascendant Hiérarchique**

→ **Normalisation**

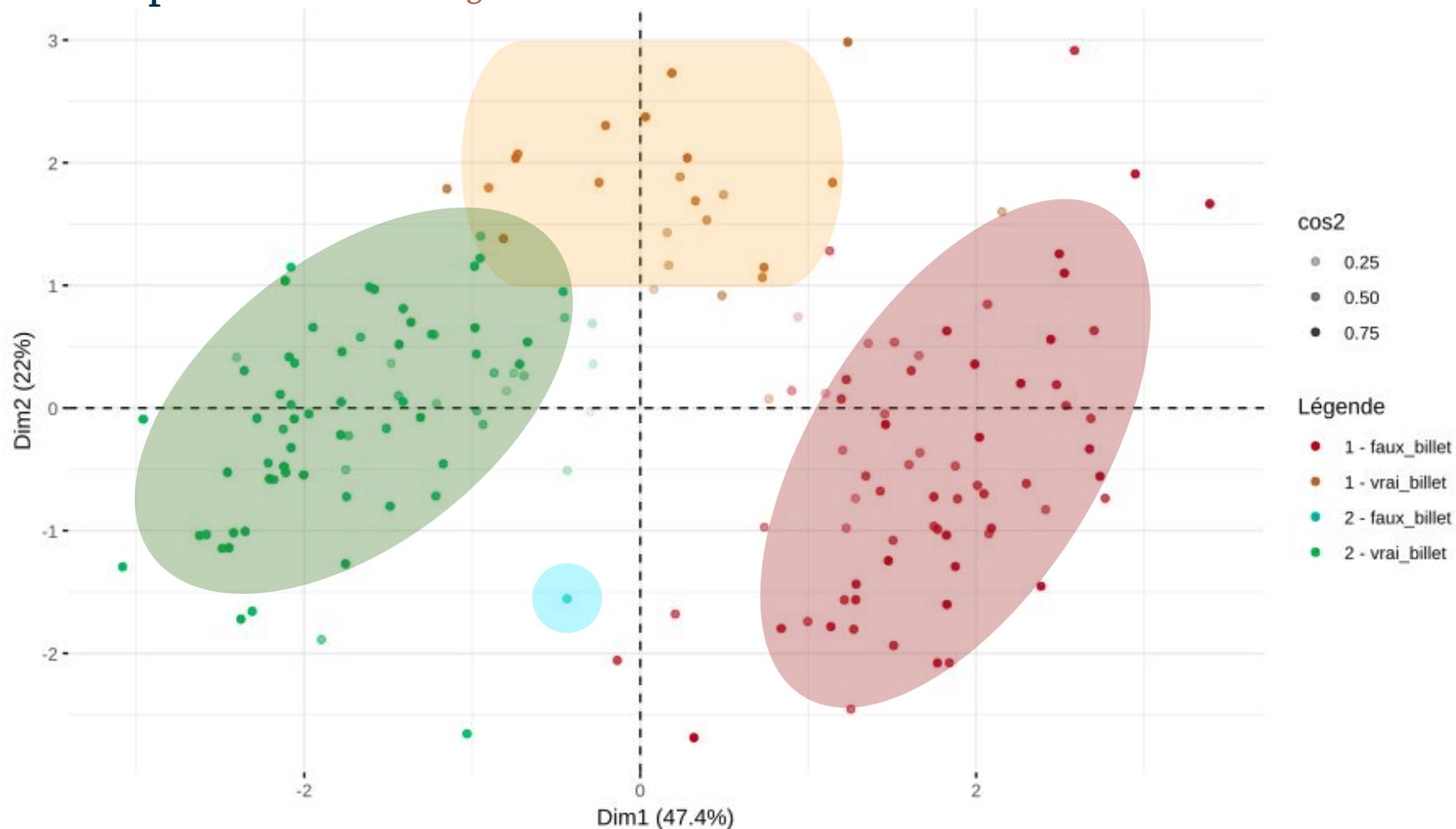
→ Méthode de **Ward** – fonction *hclust()*

Matrice de confusion

		Vrai billet	Faux billet
Cluster	1	24	69
	2	76	1

85% bien classés

# Non-supervisée CAH - Nuage des individus



# Non-supervisée **K-means**

→ Normalisation

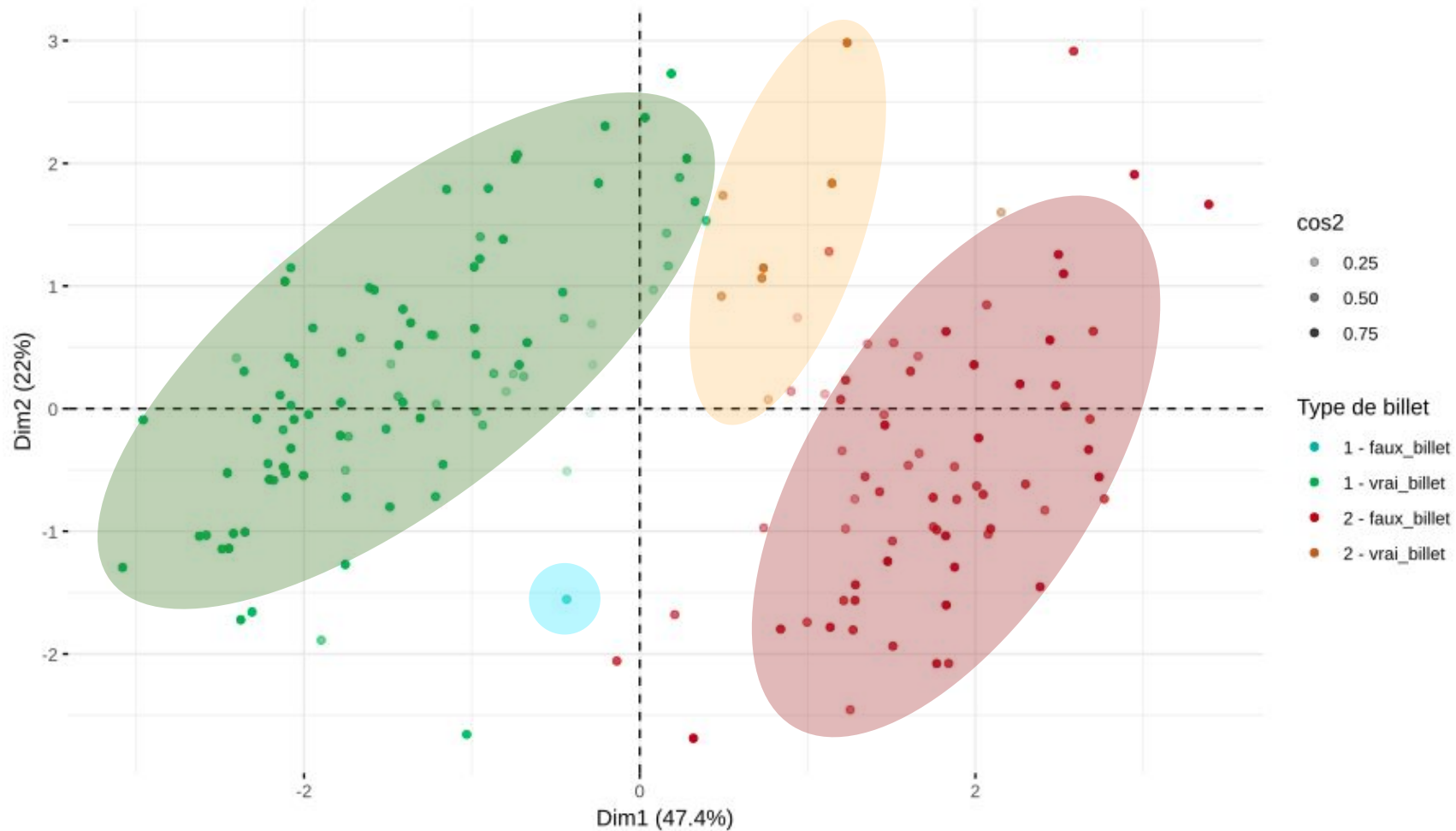
→ Fonction *kmeans* ()

Matrice de confusion

Cluster		
	Vrai billet	Faux billet
1	92	1
2	8	69

95% bien classés

# Non-supervisée K-means - Nuage des individus





# Non-supervisée **HCPC**

→ Combinaison d'ACP, CAH et K-means

→ Fonction *HCPC()* de *FactorMineR*

Matrice de confusion

Cluster		
	Vrai billet	Faux billet
1	92	1
2	8	69

→ Même résultats qu'avec la méthode des **K-means**

# Plan

## → Description des données

- ◆ Univariée et bivariée

- ◆ Multivariée

## → Classification

- ◆ Non-supervisée

- ◆ Supervisée

→ Programme de détection de faux billets

# Supervisée Régression logistique

## Division des données

→ Apprentissage : 75%

→ Test : 25% → `createDataPartition()` de `caret`

## Calculs

→ Approche stepwise (définition) → `glm()` et `stepAIC()`

## Modèle

Variables choisies	<i>Intercept</i>	<i>diagonal</i>	<i>margin_low</i>	<i>margin_up</i>
Coefficients	3247.43	-36.58	63.59	174.72

# Supervisée Régression logistique Évaluation du modèle

jeu de test → 42 billets

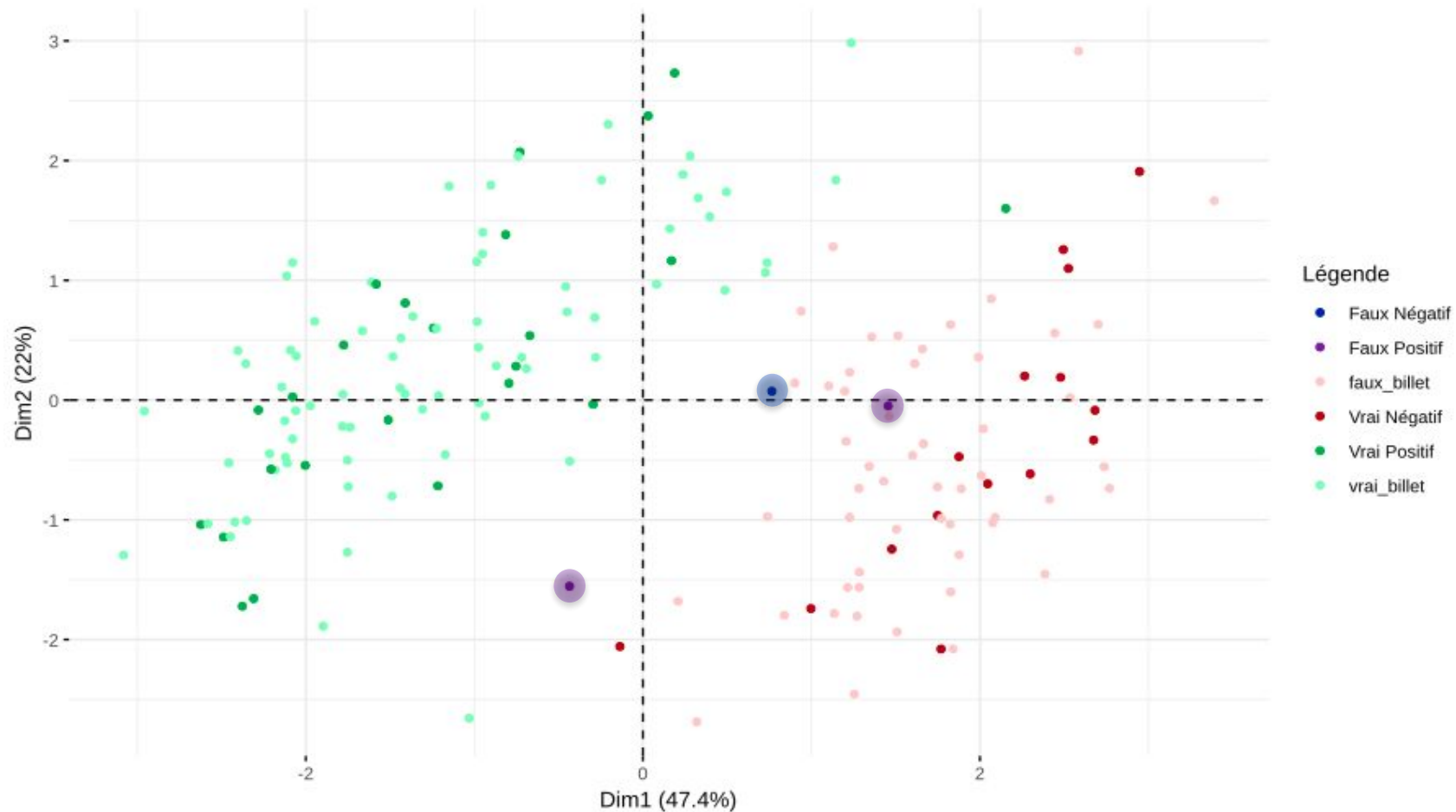
Matrice de confusion

		Réalité	
		Faux billet	Vrai billet
Cluster	Faux	15 vrai négatif	1 faux négatif
	Vrai	2 faux positif	25 vrai positif

Précision = 0.92

Rappel = 0.96

# Supervisée régression logistique - Nuage des individus



# Plan

## → Description des données

- ◆ Univariée et bivariée
- ◆ Multivariée

## → Classification

- ◆ Non-supervisée
- ◆ Supervisée

## → Programme de détection de faux billets