



INSTITUT DES  
**ACTUAIRES**

Projet de recherche de la Commission Santé

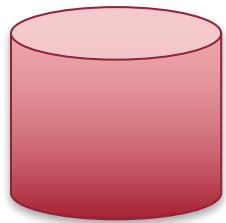
16 Novembre 2017

## 100% Actulaires

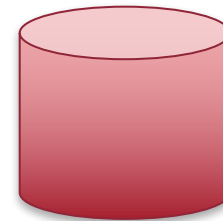
# Analyse du reste à charge des ménages sur la base de données EGB

- La commission Santé de l'Institut des Actuaires a souhaité étudier les restes à charge Santé des ménages et la dynamique des dépenses de Santé
- Après une première étude sur des données agrégées (DAMIR), l'IA a obtenu l'accès à l'EGB, échantillon représentatif des bénéficiaires de l'Assurance Maladie pour un étude plus détaillée

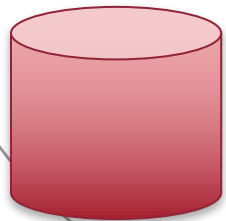
## Le SNDS est le regroupement de plusieurs bases de données



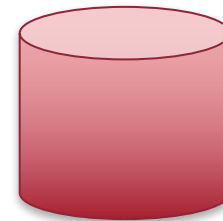
SNIRAM : données de l'Assurance Maladie



CepiDC: Causes médicales de décès



PMSI: données des séjours hospitaliers



Données relatives au handicap (CNSA)

# SNDS / EGB : qu'est-ce que c'est

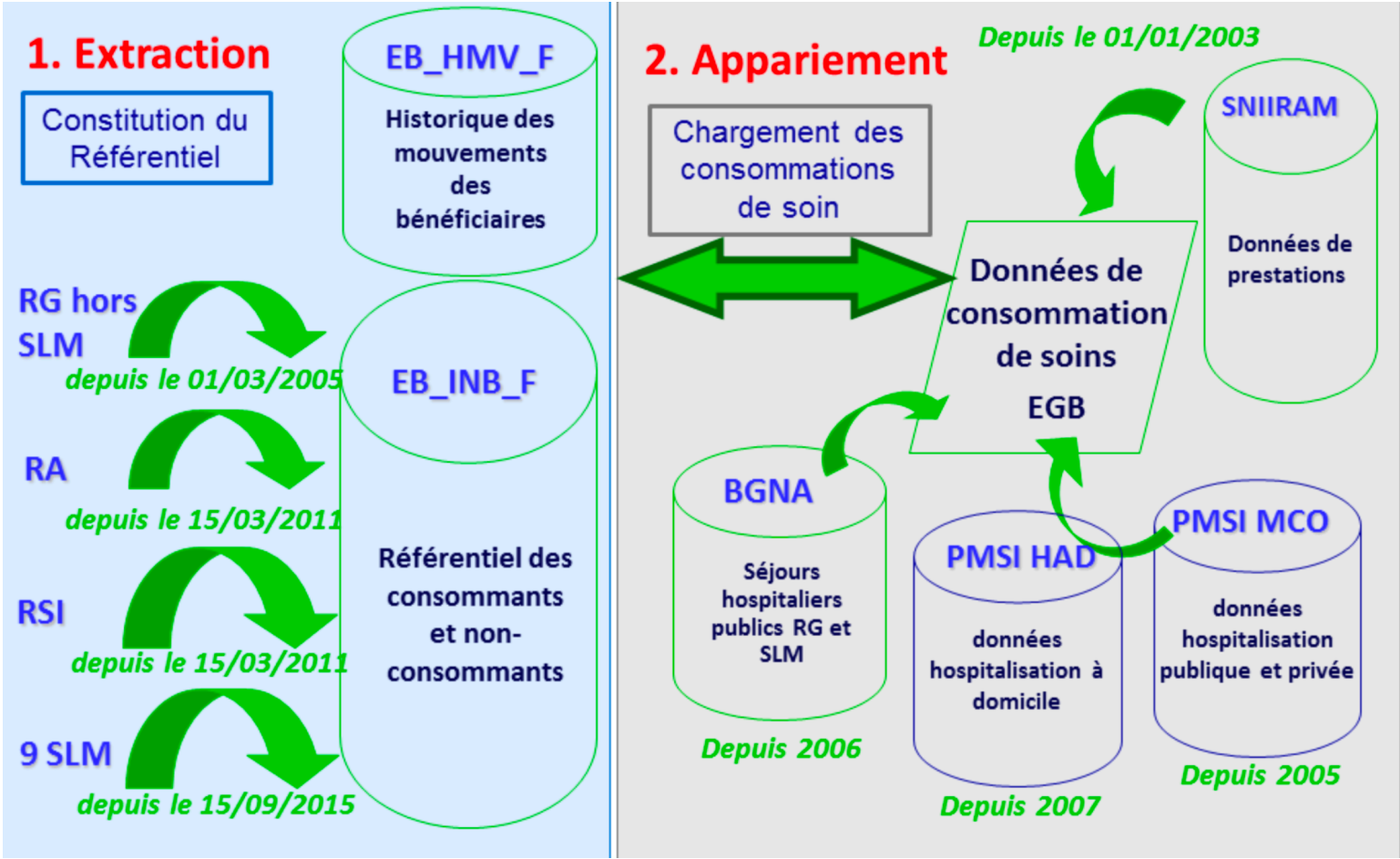
**L'EGB est un échantillon permanent anonymisé et représentatif du SNDS**

- EGB : Échantillon Généraliste des Bénéficiaires
- 1/97<sup>e</sup> de la population française
- 10 années de recul
- Caractéristiques sociodémographiques et médicales
- Prestations
- Suivi de la consommation de soins et taux de recours aux soins
- Bénéficiaires "consommants ou non"

# LE SNDS est une base de données très complexe

1. Données élémentaires de flux non retraitées
2. Données très détaillées (individuelles, PS, établissements et bénéficiaires)
3. Nomenclatures très complexes à gérer (plus d'un millier de types d'acte par exemple)
4. Evolution constante des SI, des pratiques de codage et des nomenclatures à connaître

# LE SNDS est une base de données très complexe



## 700 000 assurés entre 2013 et 2015

- Les informations administratives (âge, sexe)
- Les dépenses détaillés de ville par année
- Les dépenses détaillés d'hospitalisation
- Les pathologies éventuelles (30 ALD différentes) et leur date de début
- Des indicateurs pour approcher le niveau socio-économique
- La commune de résidence

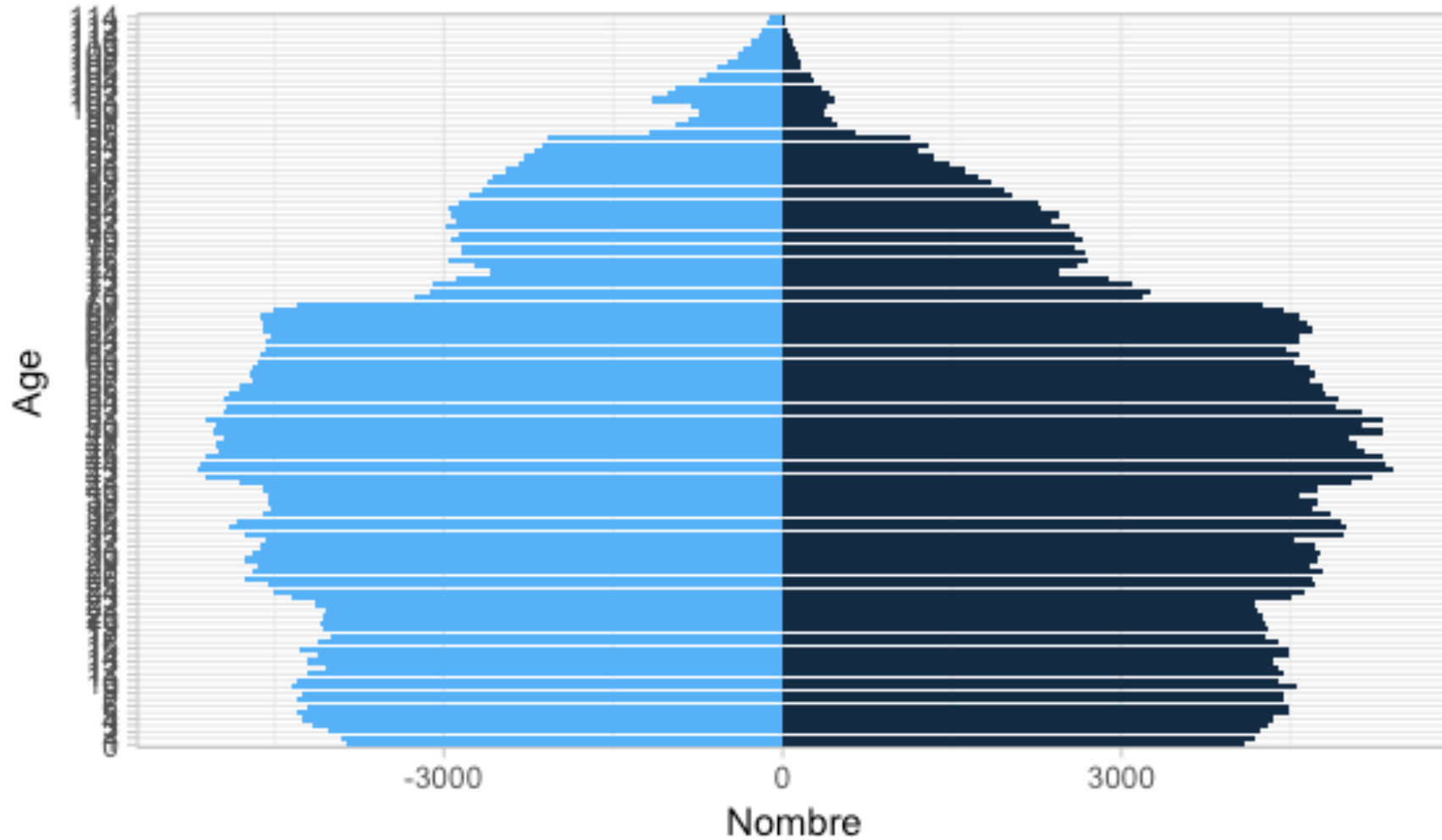


**Les dépenses qui ne sont pas rattachées à un bénéficiaire ne sont donc pas pris en compte**

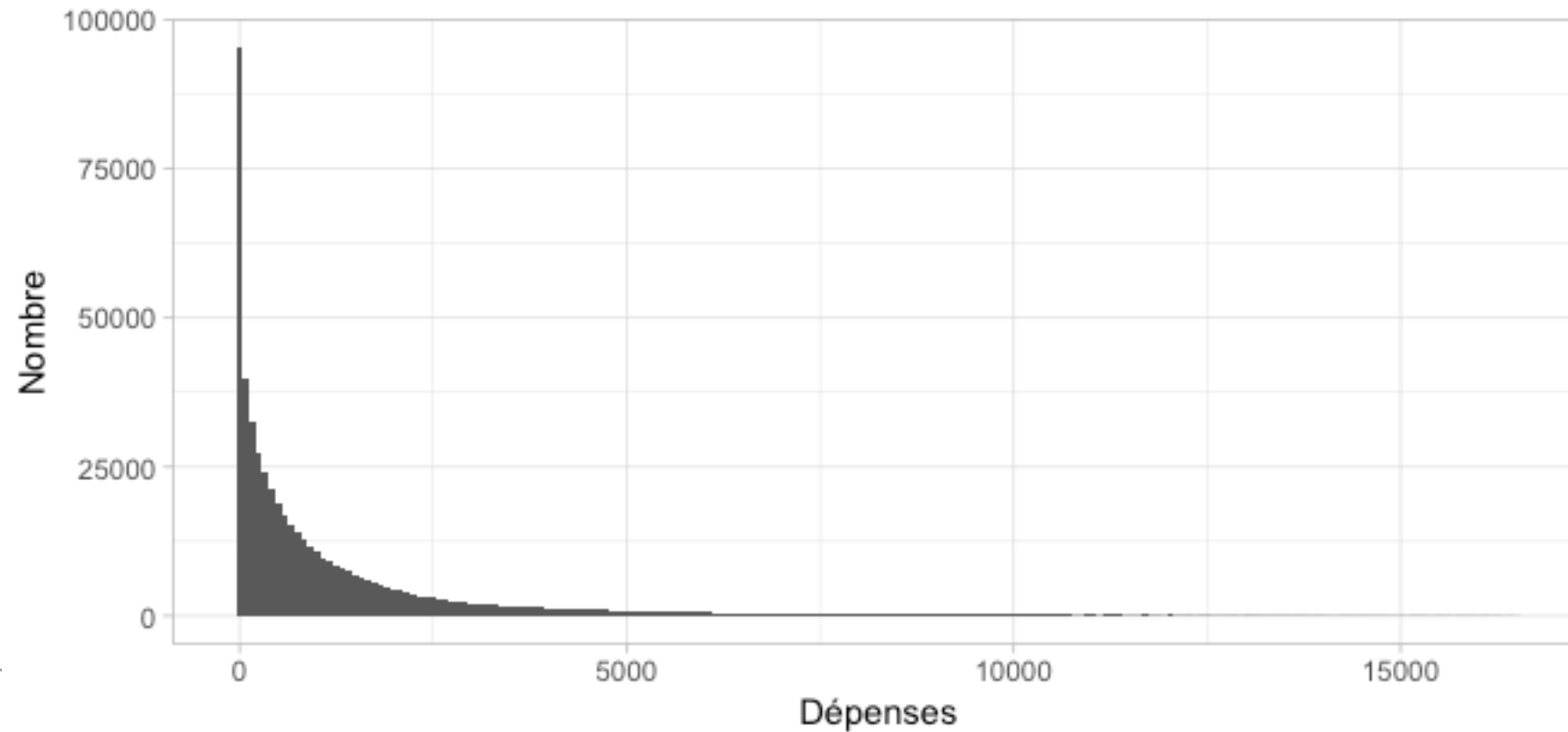
- Les dépenses d'hospitalisation "en budget global" (PSY, SSR...)
- Les MIGAC des hôpitaux
- Les dépenses directement versées aux médecins (rémunérations sur objectifs, primes d'installation...)

**Environ 75% des dépenses sont reliées à des bénéficiaires et font l'objet de l'étude**

# Quelques statistiques descriptives

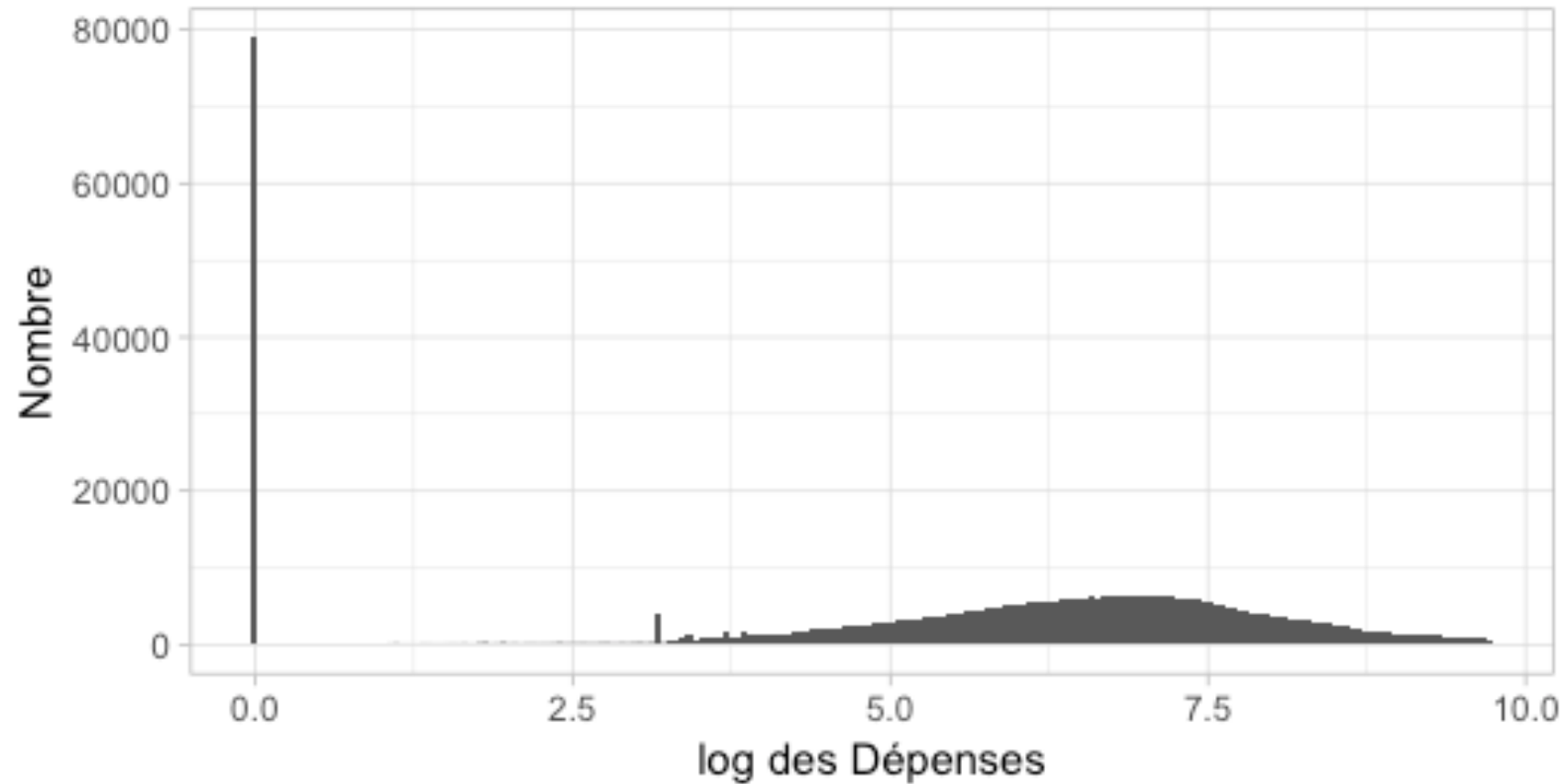


## Distribution des dépenses



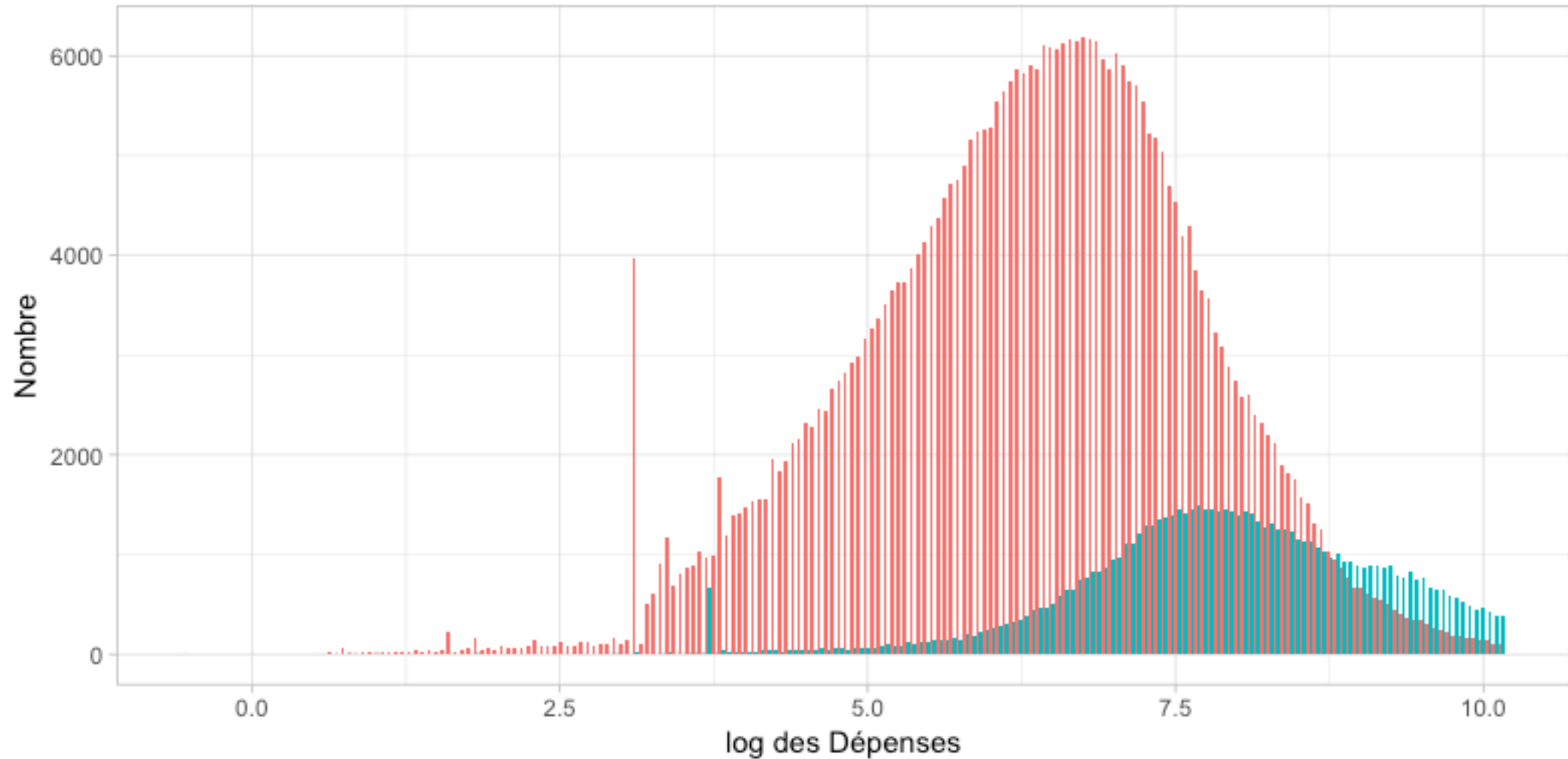
# Quelques statistiques descriptives

Distribution du logarithme des dépenses

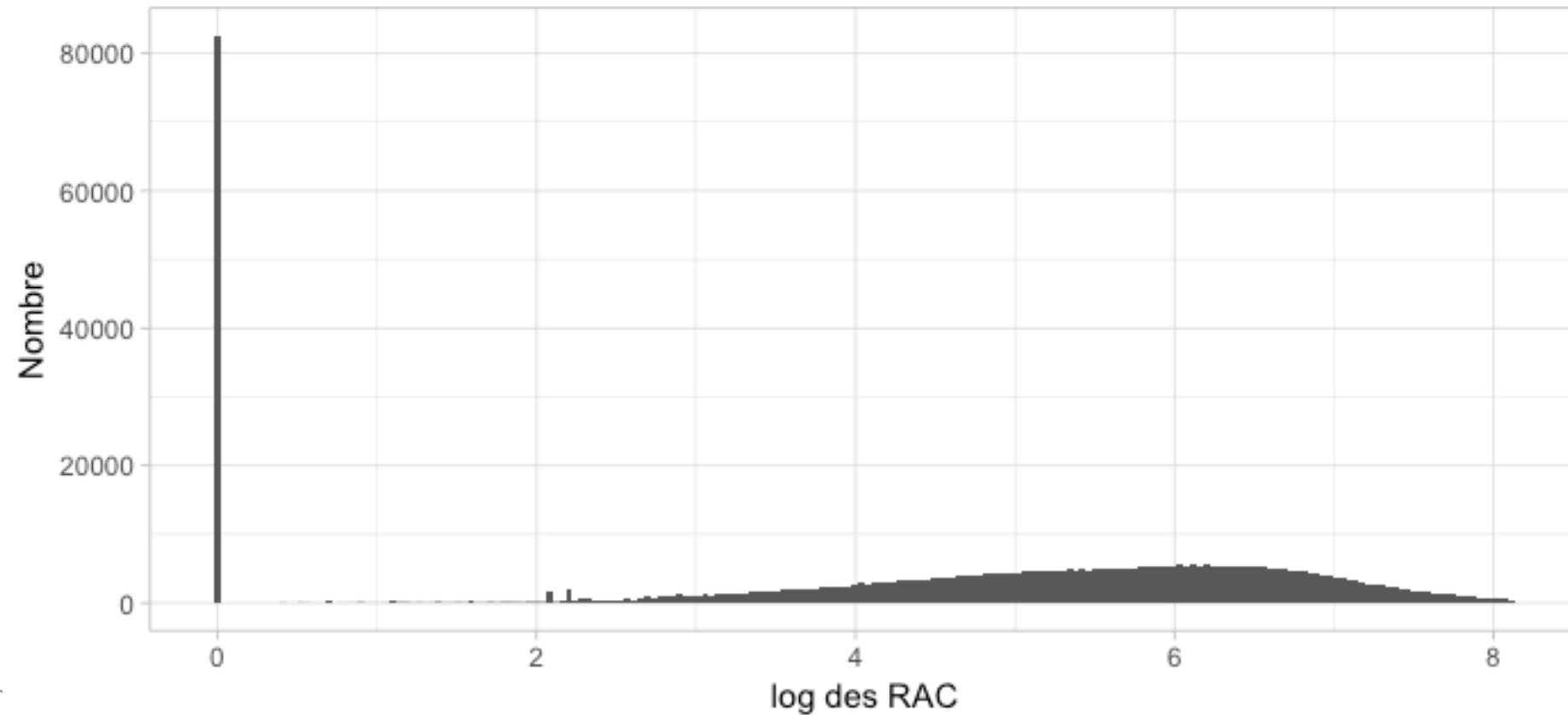


# Quelques statistiques descriptives

Distribution du logarithme des dépenses ALD / non ALD

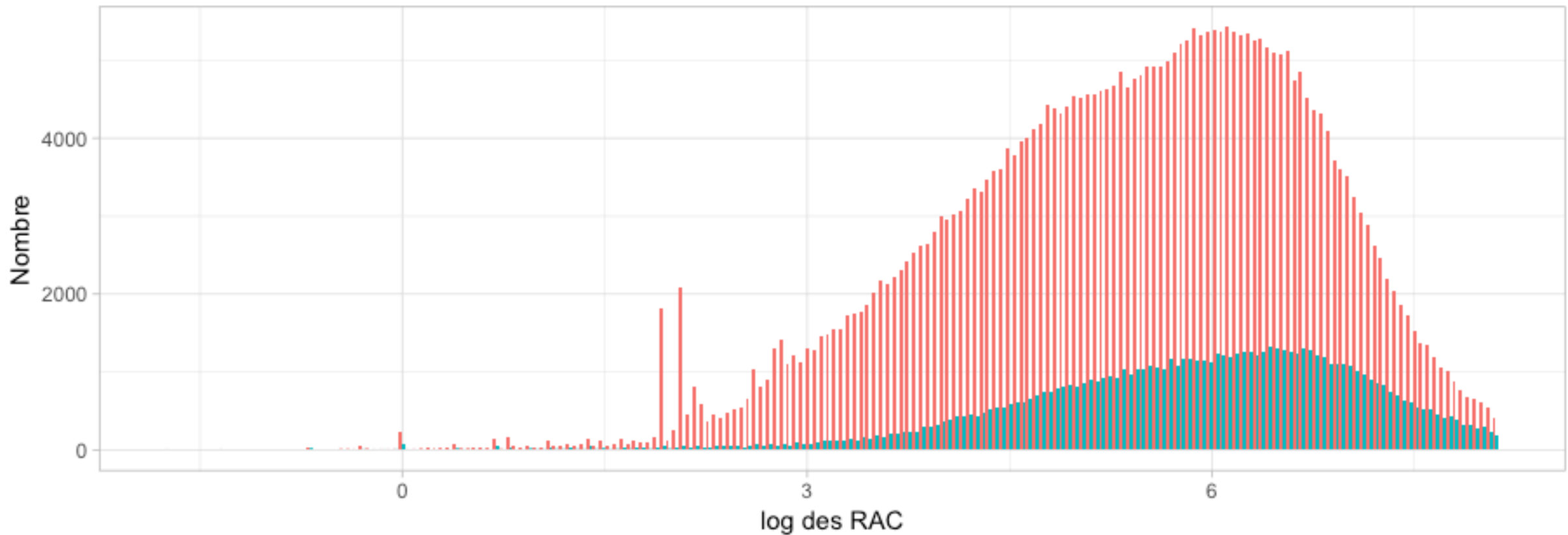


## Distribution du Reste à charge FR-RO

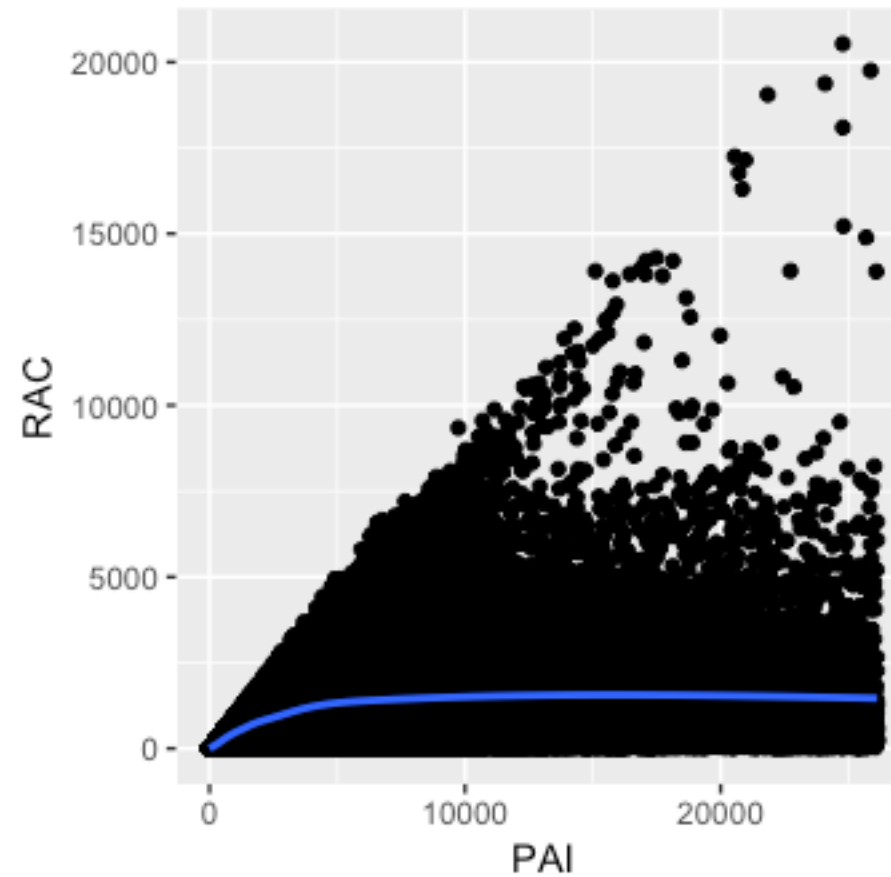


# Quelques statistiques descriptives

Distribution du Reste à charge FR-RO ALD / non ALD



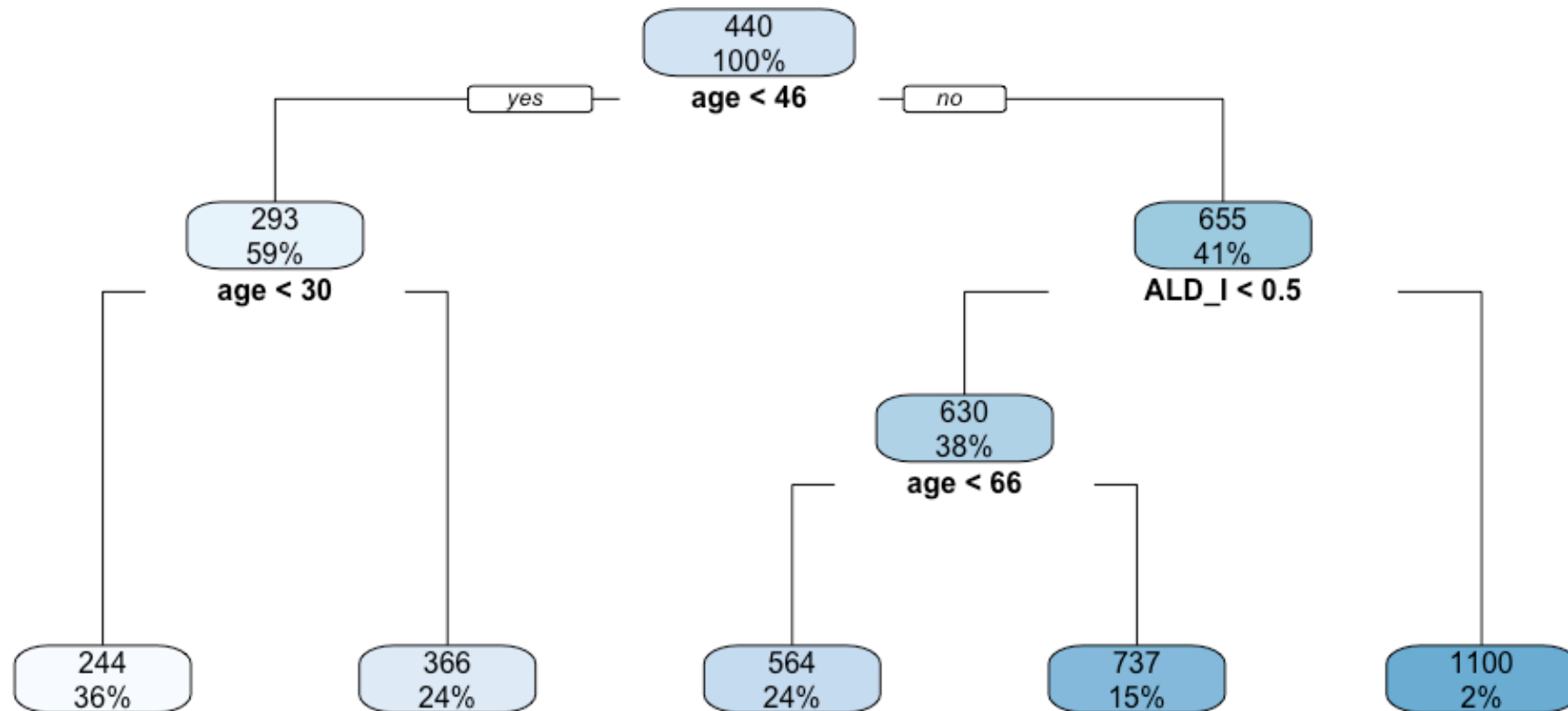
Reste à charge en fonction des dépenses



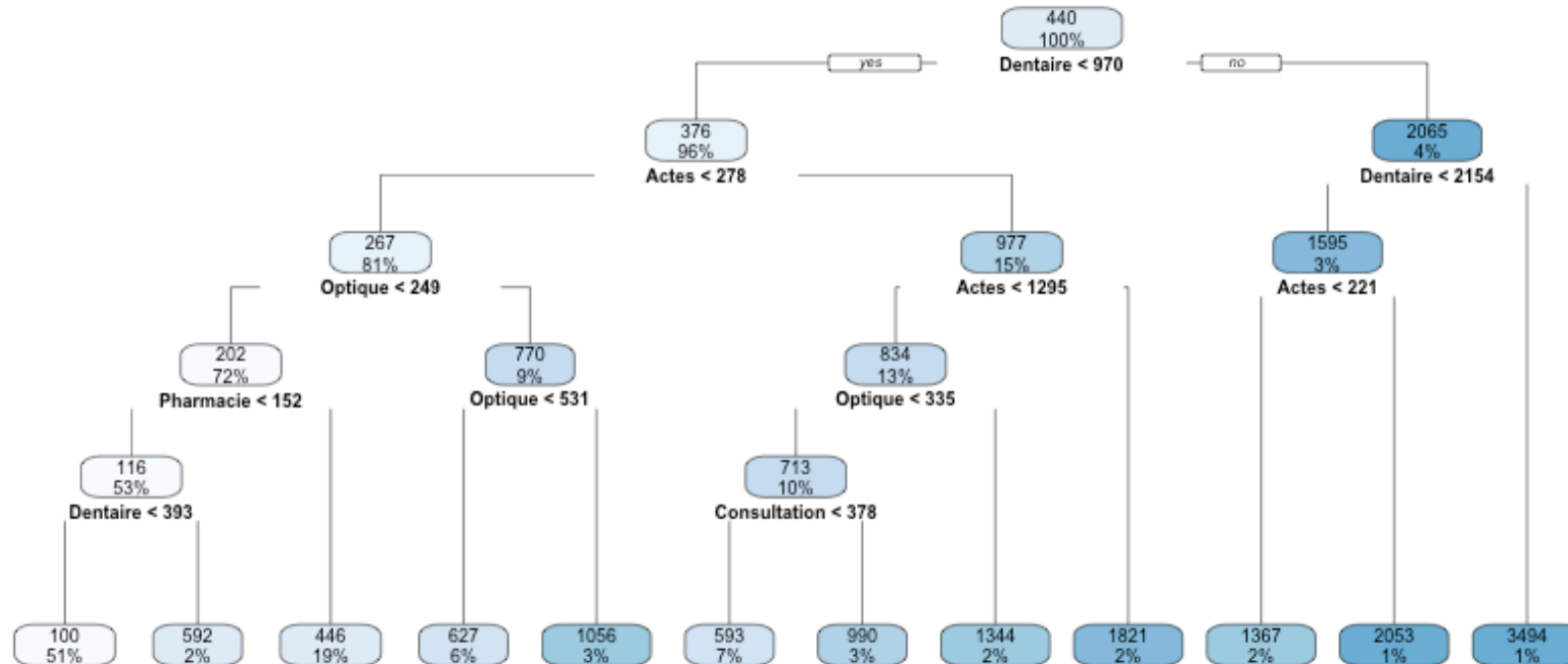


# Quelques statistiques descriptives

## Profil patients avec restes à charge



## Origine des restes à charge



## **II. Projection des populations**

# Objectifs

- Les personnes en ALD représentant un poids important dans les dépenses, estimer leur prévalence
- L'idée est d'estimer la proportion de personnes en ALD dans les prochaines années

# Méthodologie

Étape 1

- Segmentation des assurés

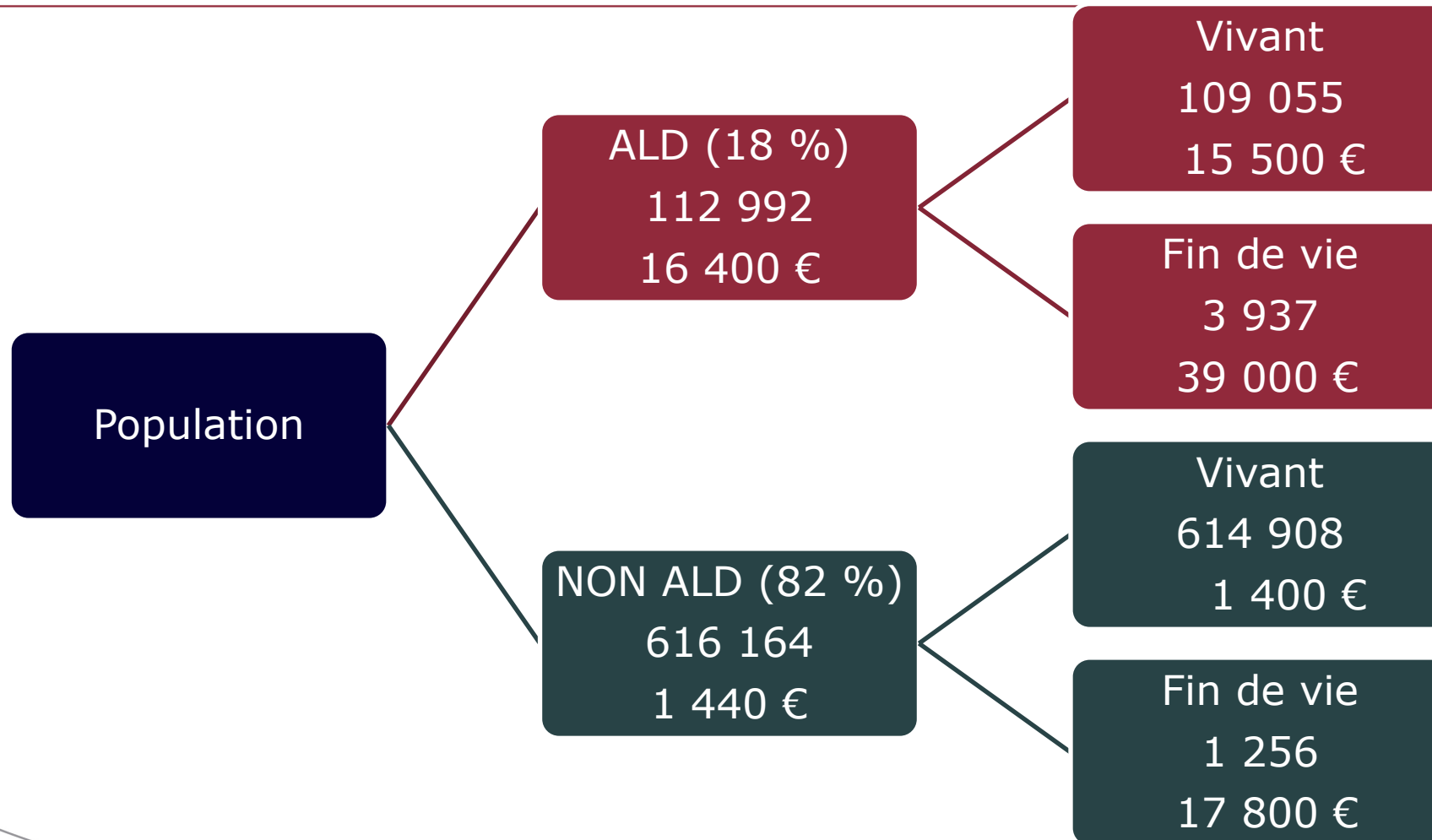
Étape 2

- Projections démographiques

Étape 3

- Projection des dépenses

## II. Segmentation des assurés



### Principe de l'algorithme CART

- CART : Classification And Regression Trees
- Méthode supervisée
- Arbre binaire : racine, nœuds intermédiaires, feuilles et branches
- Objectif : découper les observations en sous groupes homogènes en fonction de la variable à prédire (dépenses)

- Variable explicative quantitative :

$$X_j < x \quad \text{ou} \quad X_j \geq x$$

- Variable explicative qualitative :

$$X_j \in \{X_{jk} \dots X_{jK}\}$$

## a. Algorithme CART : groupes non ALD vivants

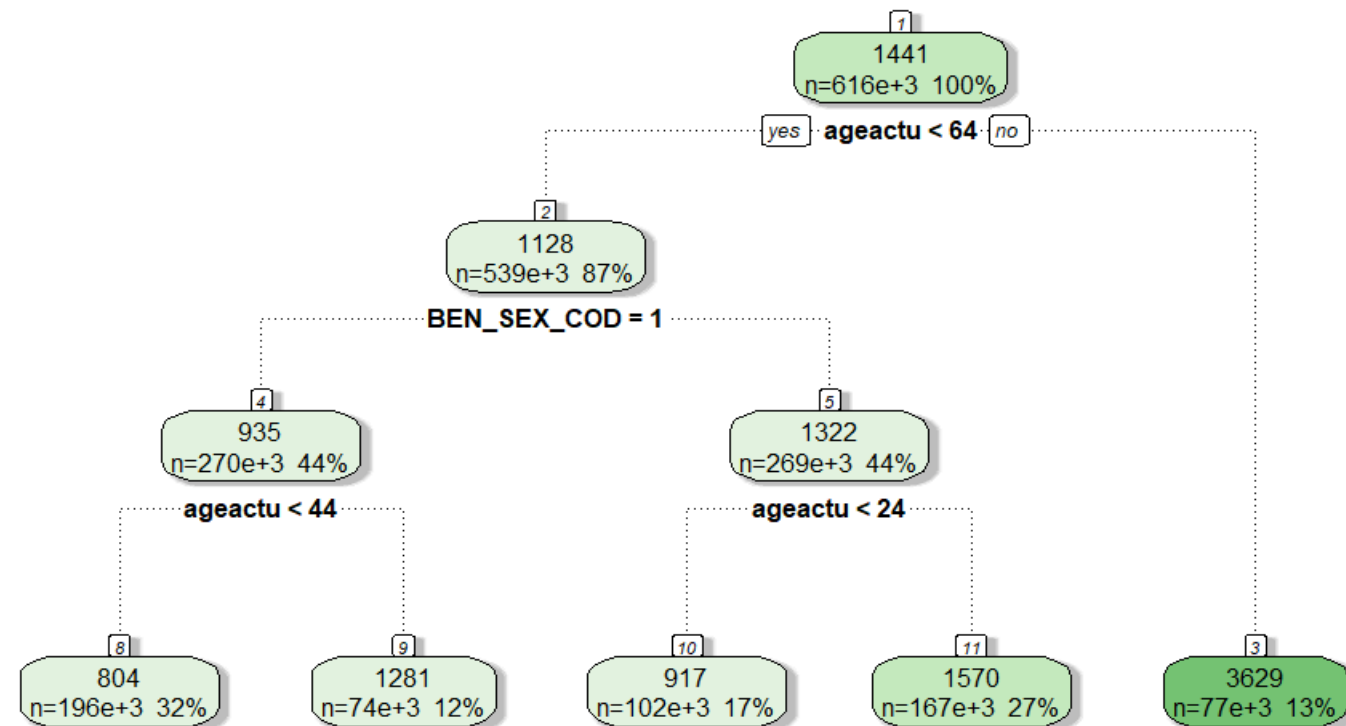
- Mise en place de l'algorithme :
  - Critère optimal
  - règle d'arrêt
  - Affectation de classes
- 3 étapes successives :
  - Construction de l'arbre maximal (ou saturé)
  - Élagage de l'arbre saturé
  - Sélection finale (arbre optimal)



## a. Algorithme CART : groupes non ALD vivants

Population non ALD « vivants » en 2015

Variables explicatives : âge et sexe / Variable cible : dépenses totales



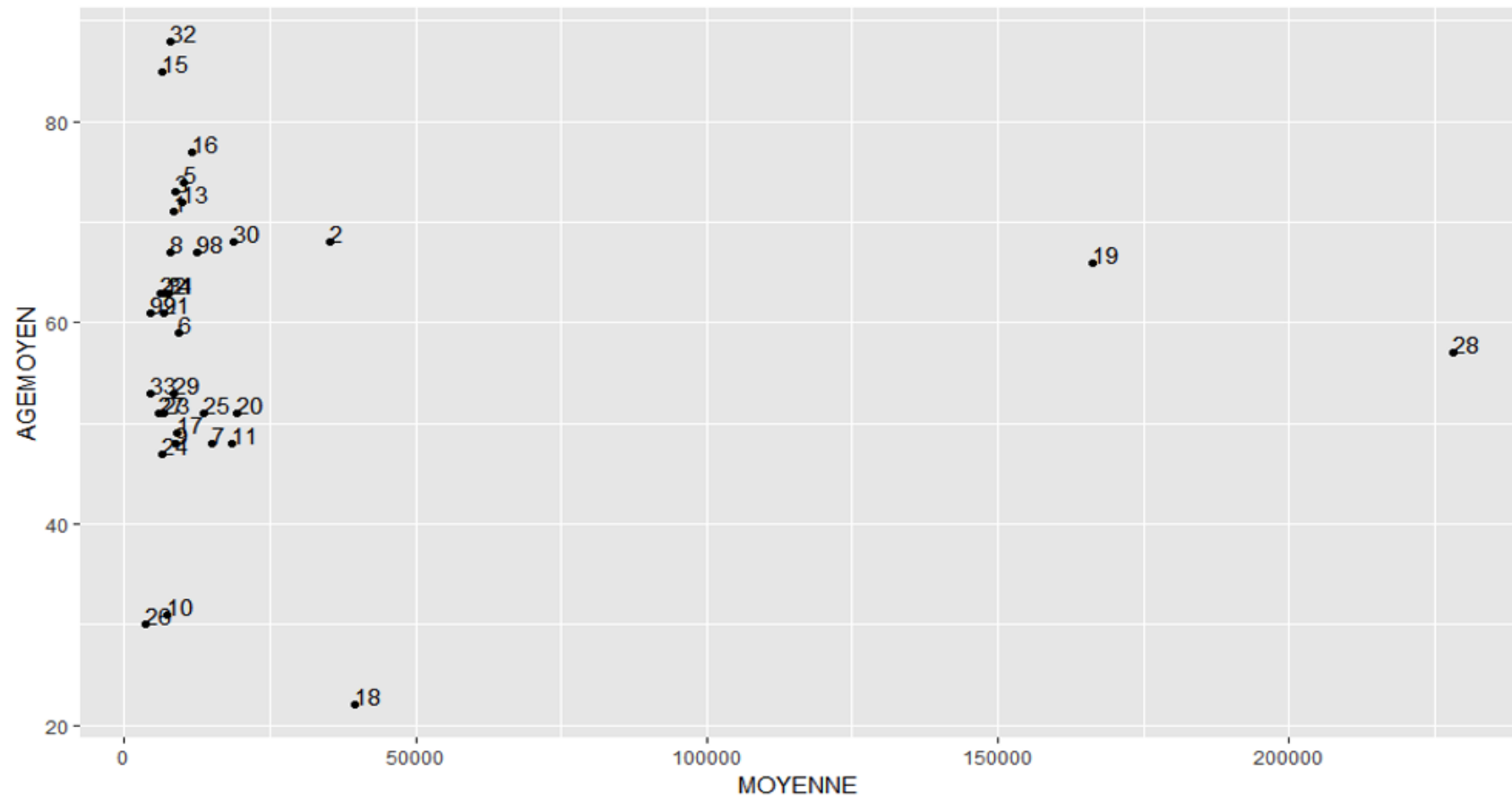
## a. Algorithme CART : groupes non ALD vivants

- 5 Groupes non ALD "vivants"
- 1 Groupe non ALD en fin de vie NONALD 6 (1 256) dépense moyenne 17 800 €

Groupe	Âge	Sexe	Effectifs	dépenses moyennes	Pourcentage Effectifs
NONALD 1	$\geq 64$	1-2	77 000	3 629 €	13 %
NONALD 2	$< 44$	1	196 000	804 €	32 %
NONALD 3	44-64	1	74 000	1 281 €	12 %
NONALD 4	$< 24$	2	102 000	917 €	17 %
NONALD 5	24-64	2	167 000	1 510 €	26 %

## b. Méthode CAH : groupes ALD vivants

Dépenses moyennes et âges moyens en fonction des ALD

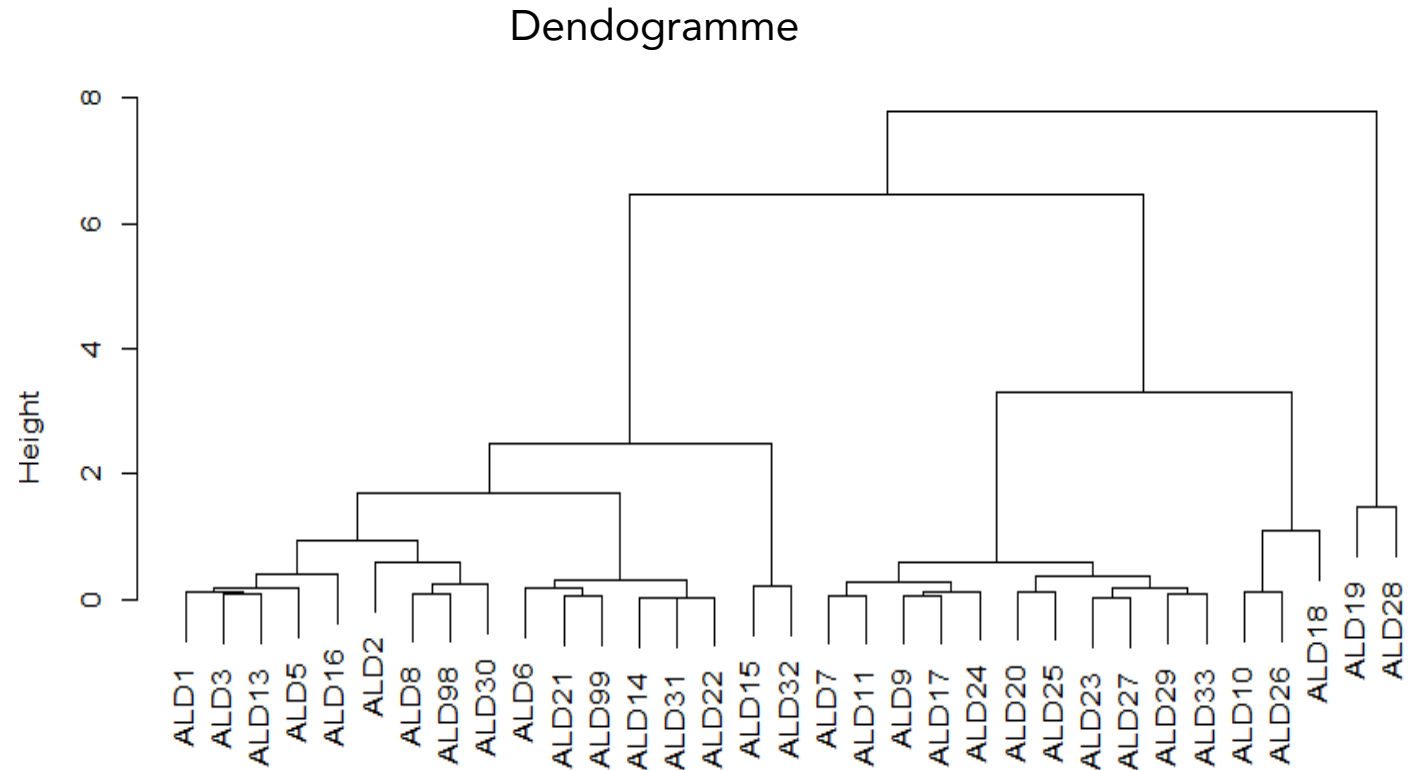


## b. Méthode CAH : groupes ALD vivants

- CAH : Classification ascendante hiérarchique
- Homogénéité intra-classe / hétérogénéité inter-classe
- Matrice de distance
- Méthode Ward :
  - Minimisation de l'inertie inter-classe
  - Maximisation de l'inertie intra-classe

## b. Méthode CAH : groupes ALD vivants

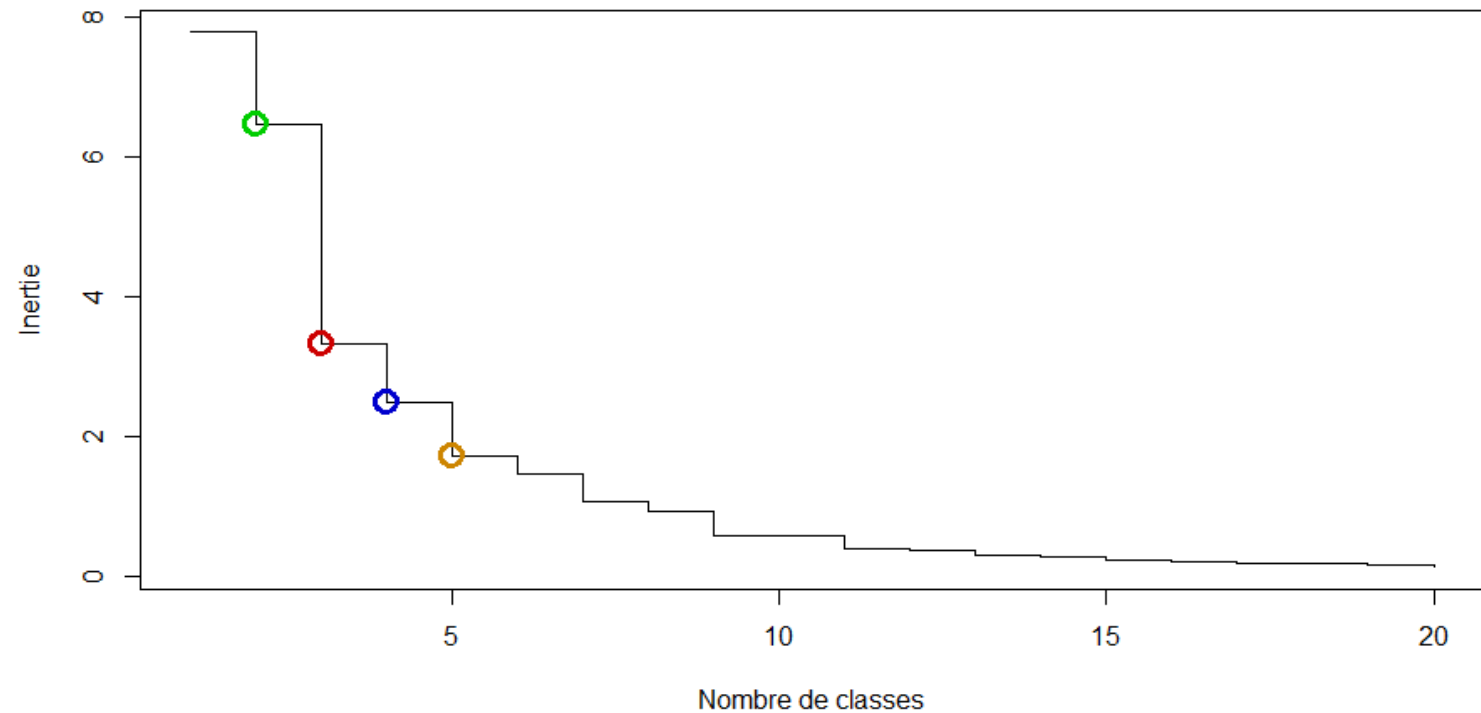
- ALD 19 et ALD 28
- ALD 10 ALD 18 et ALD 26



## b. Méthode CAH : groupes ALD vivants

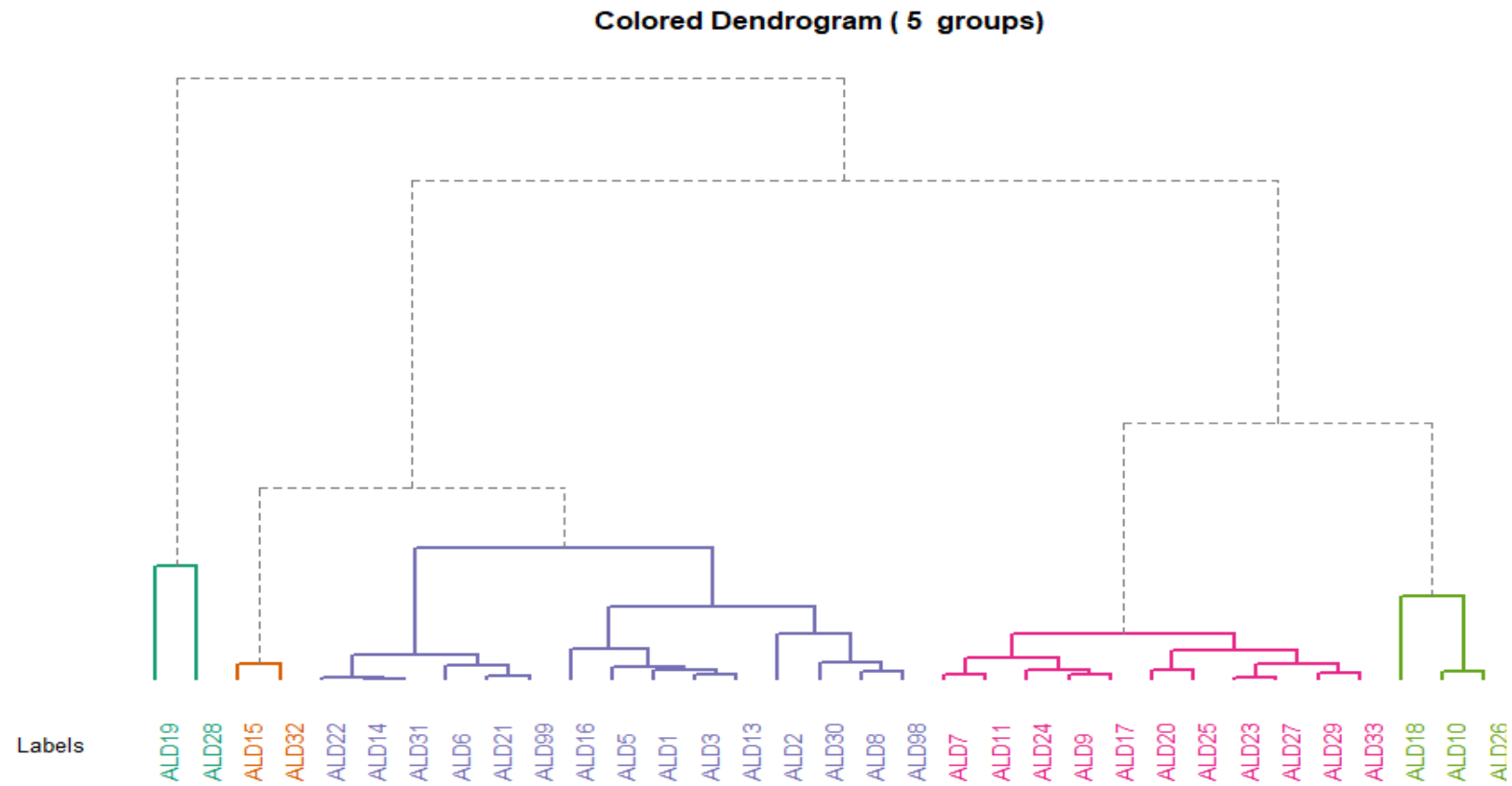
Sauts d'inertie en fonction du nombre de classes

- 4 sauts importants
- 5 classes



## b. Méthode CAH : groupes ALD vivants

Partition de l'arbre en 5 classes



## b. Méthode CAH : groupes ALD vivants

- 5 Groupes d'ALD "vivants"
- 1 Groupe d'ALD en fin de vie

Groupe ALD	ALD	Effectifs	dépenses moyennes	Pourcentage Effectifs
ALDG 1	19-28	2 020	191 000 €	1,8 %
ALDG 2	15-32	2 348	7 834 €	2 %
ALDG 3	18 - 10 - 26	720	8 110 €	0,5 %
ALDG 4	7-9-11-17-20-23-24-25-27-29-33	20 805	9 237 €	18,4 %
ALDG 5	Autres ALD dont 8 et 30	83 274	13 178 €	73,7 %
ALDG 6	ALD Fin de vie	3 937	39 987 €	3,8 %



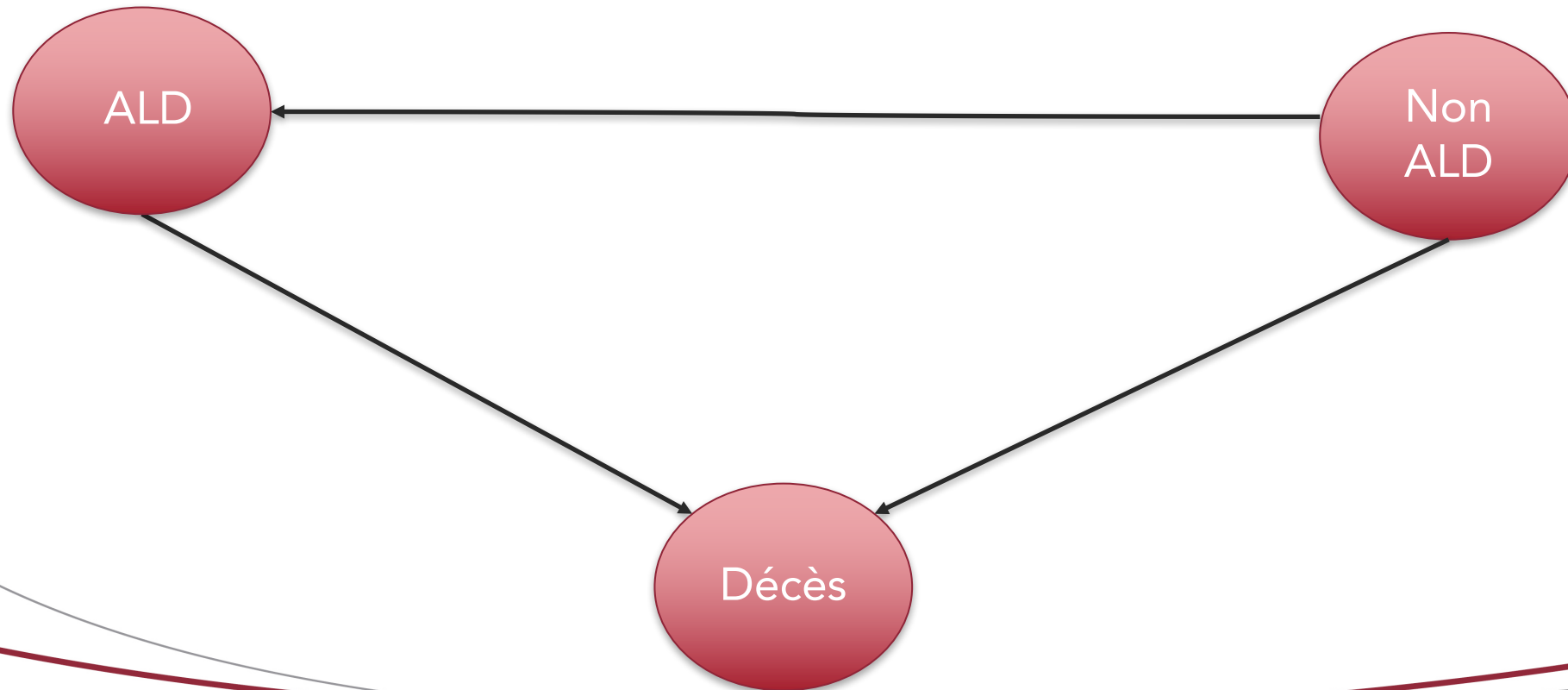
# Récapitulatif groupes

12 classes homogènes de dépenses : 6 ALD et 6 non ALD



## Transitions entre les groupes

Hypothèse : pas de transition possible entre les ALD



## Projections démographiques

1. Considérer la population totale année N
2. Appliquer les différentes probabilités de transition
3. Les survivants ont 1 an de plus
4. Nouveaux-nés année N+1

Population N+1

## Projection des dépenses

1. Considérer la population projetée en N
2. Appliquer les dépenses moyennes par individus associée à chaque groupe au nombre d'individus dans le groupe en N
3. Les individus qui décèdent en N auront un coût lié à la fin de vie (année de décès = fin de vie) en N+1
4. Ces individus considérés « morts » auront un coût = 0 à partir de N+2

## a. Estimation des paramètres de projection

### Taux de mortalité par âge et groupe

Hypothèses :

- taux de mortalité et taux de passages en ALD constants dans le temps
- Base agrégée 2013-2015 et regroupement classes ALD 1, 2 et 5 ainsi que 3 et 4

$$q_{x,ALD_i} = \frac{\text{Nombre de décès } (x, ALD_i)}{\text{Nombre de personnes } (x, ALD_i)}$$

$i = 1..5$  pour chaque groupe d'ALD

$$q_{x,HORSALD} = \frac{\text{Nombre de décès } (x, HORSALD)}{\text{Nombre de personnes } (x, HORSALD)}$$

## a. Estimation des paramètres de projection

- Taux de passage en ALD

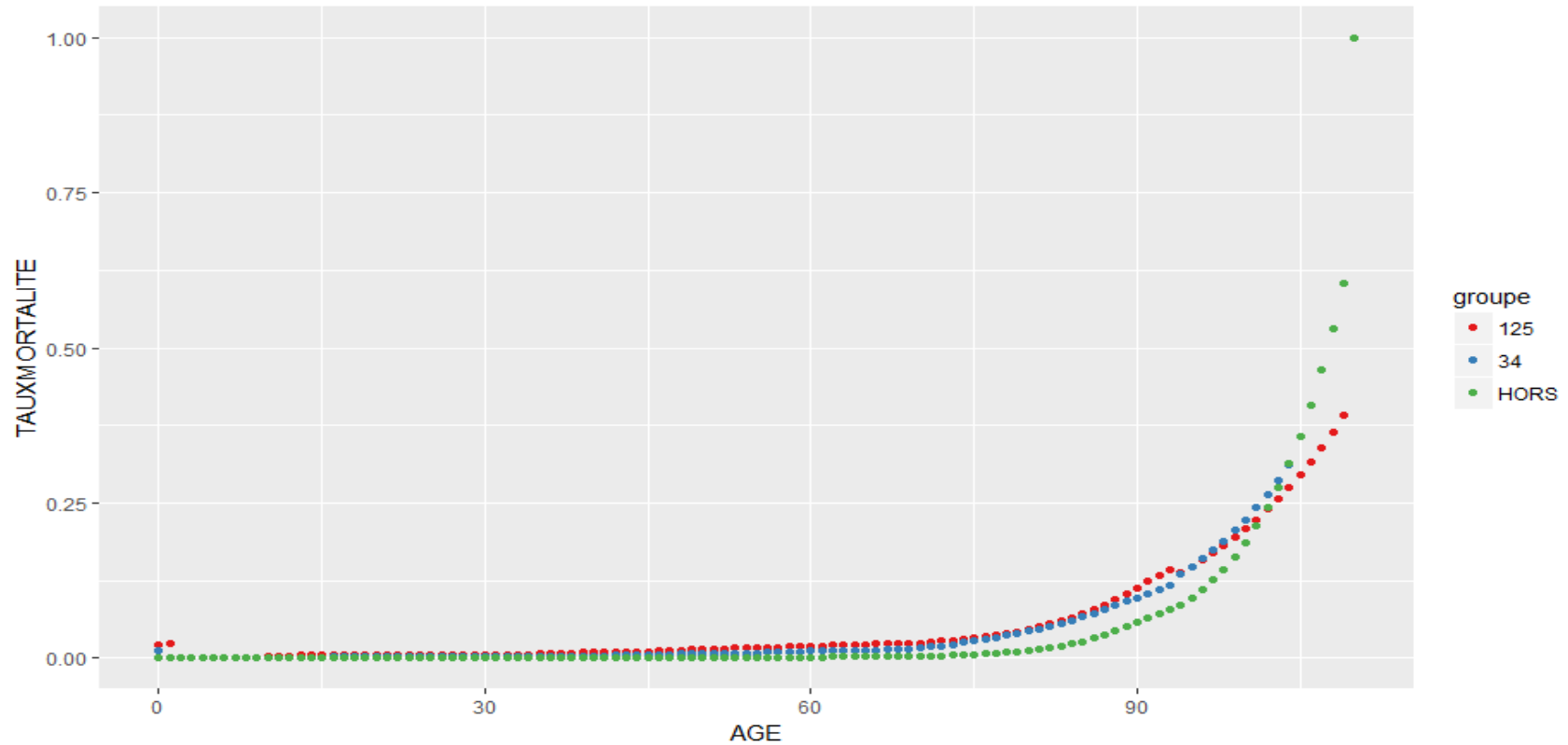
$$\mathbb{P}(HORSALD_i \rightarrow ALD_i) = \frac{N_{Passage,i,x}}{N_{HORSALD_i}}$$

- Taux de fécondité de l'INSEE (femmes 15-50 ans)
  - Proportion de femmes dans chaque groupe par rapport à la population totale
  - Prise en compte de la mortalité infantile : naissances en « fin de vie »

$$f(t) = \frac{N_{naissance,t}}{N_{f,t}}$$

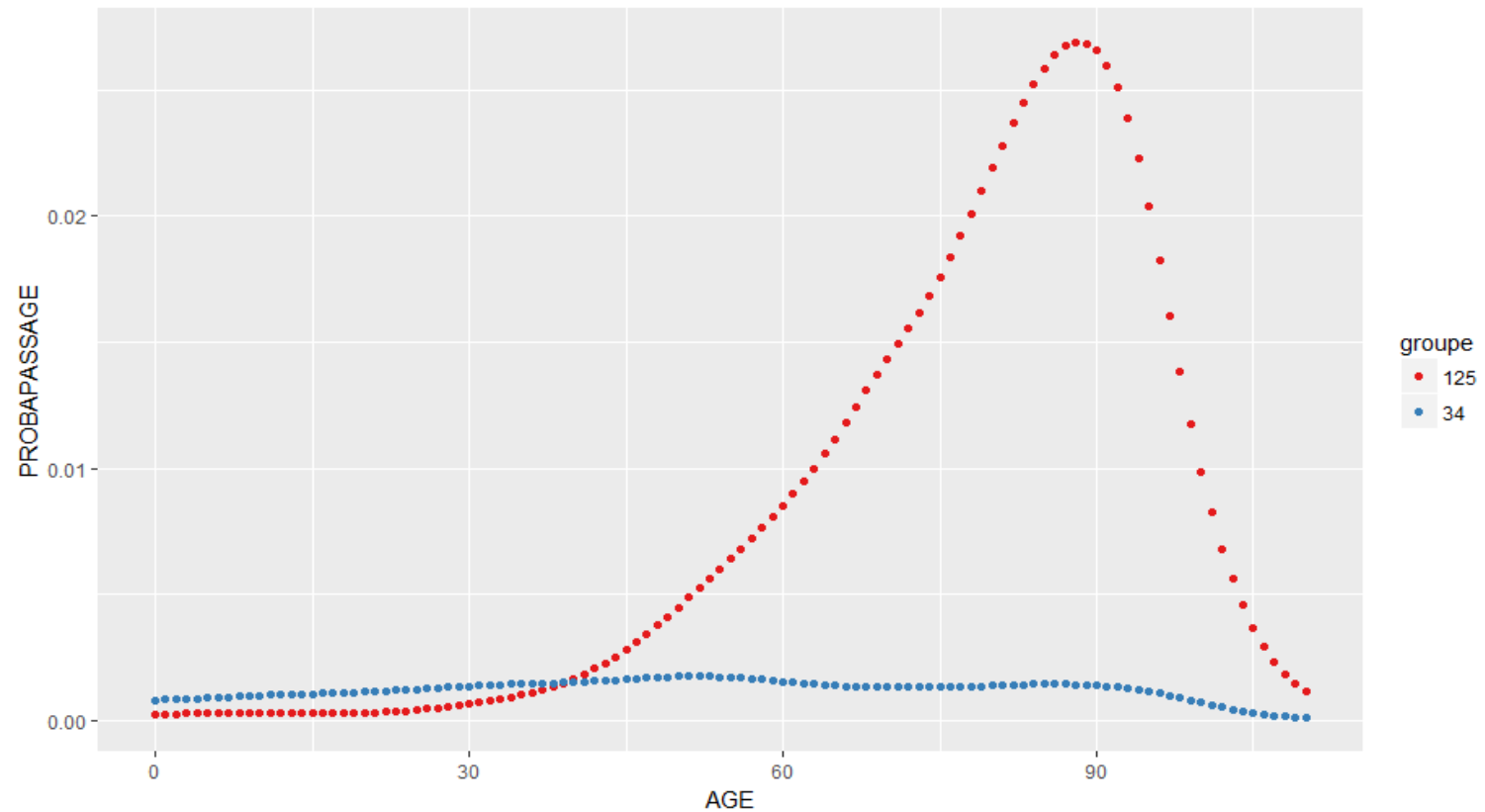
# a. Estimation des paramètres de projection

## Comparaison des taux de mortalité lissés par groupe



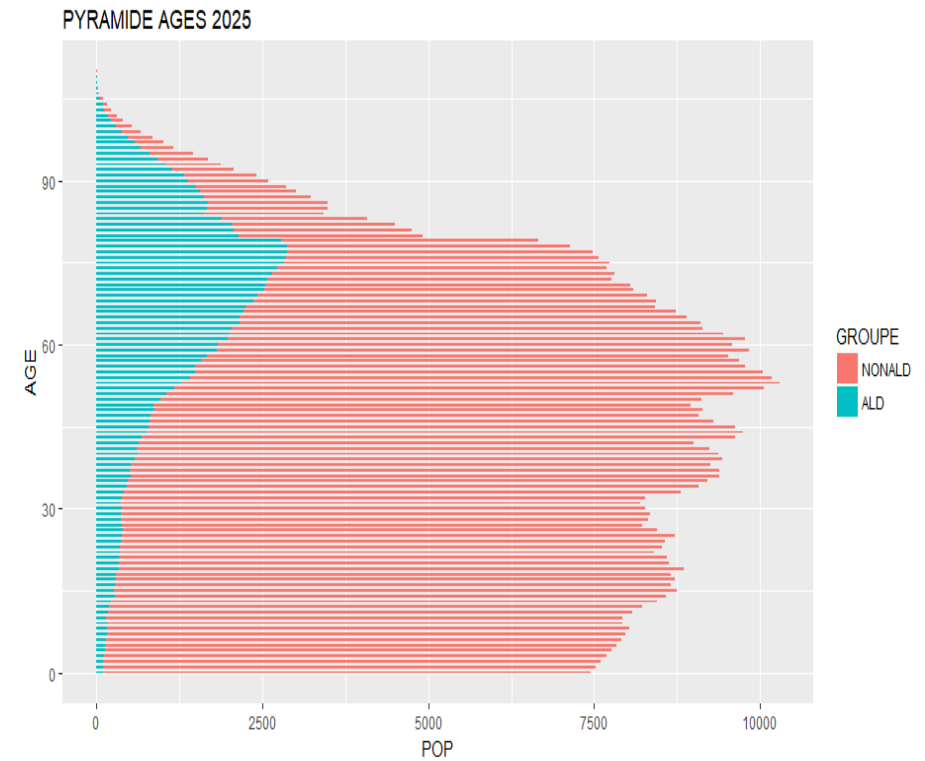
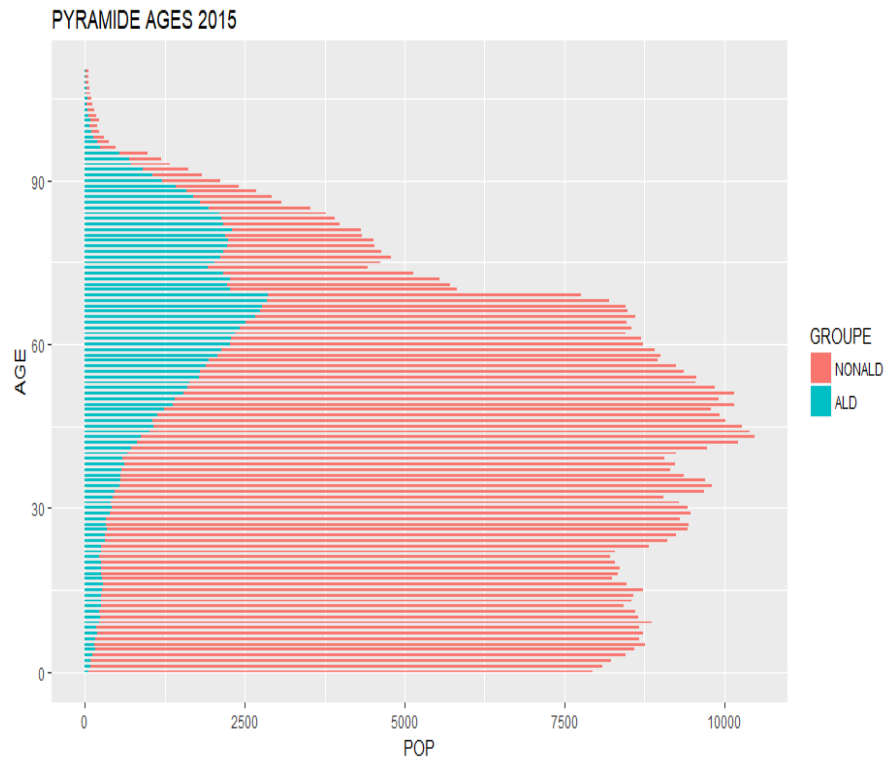
# a. Estimation des paramètres de projection

## Comparaison des lois de passage en ALD (taux lissés)



## b. Projections démographiques

### Population ALD et non ALD en 2015 et 2025





- Limites liées aux données
  - ✓ échantillon au 1/97ème
  - ✓ Regroupement des profils dans le calcul des lois de passages
- Limites liées au modèle
  - ✓ Hypothèse d'homogénéité
  - ✓ Hypothèse Markovienne
  - ✓ Surestimation des taux d'évolution des dépenses
- Perspectives
  - ✓ Analyse par postes de dépenses
  - ✓ Méthodes stochastiques pour la détermination des lois de passage

The logo consists of three white geometric shapes: a small triangle, a larger triangle, and a wide trapezoid, all pointing towards the right. The text 'INSTITUT DES ACTUAIRES' is positioned to the left of these shapes.

**INSTITUT DES  
ACTUAIRES**