

Analyse factorielle des correspondances

AUTHOR
Mariam SYLLA

Introduction:

L'analyse factorielle des correspondances (AFC) est une méthode statistique d'analyse des données qui permet d'étudier l'association entre deux variables catégorielles avec plusieurs modalités. L'objectif est de représenter un maximum de l'inertie totale sur le premier axe factoriel, un maximum de l'inertie résiduelle sur le second axe, et ainsi de suite jusqu'à la dernière dimension.

Définition du problème:

Le dataset est constitué des informations sur les causes de décès en fonctions des tranches d'âge. .

Importation des données

► Code

	a0.1	a1.4	a5.14	a15.24	a25.34
Abus_alcool	0	0	0	29	214
Accidents_transport	52	194	655	4653	2451
Arthrite_rhumatoide_osteoarthrite	0	0	3	2	6
Asthme	9	6	9	29	39
Autres_maladies_toute_cause	1561	894	1145	2906	3255
Autres_malformations_congen_anom_chromosomiques	798	83	50	44	51
	a35.44	a45.54	a55.64	a65.74	
Abus_alcool	763	1839	1543	1203	
Accidents_transport	1841	1927	1409	1479	
Arthrite_rhumatoide_osteoarthrite	6	26	59	301	
Asthme	59	183	266	535	
Autres_maladies_toute_cause	5150	12494	20624	52970	
Autres_malformations_congen_anom_chromosomiques	50	114	132	76	
	a75.84	a85.94	a95.		
Abus_alcool	636	95	5		
Accidents_transport	1211	272	12		
Arthrite_rhumatoide_osteoarthrite	526	301	57		
Asthme	846	464	67		
Autres_maladies_toute_cause	118780	98187	19663		
Autres_malformations_congen_anom_chromosomiques	79	45	1		

Nettoyage des données

Analyse exploratoire

► Code

a0.1	a1.4	a5.14	a15.24
Min. : 0.0	Min. : 0.00	Min. : 0	Min. : 0.0
1st Qu.: 0.0	1st Qu.: 1.00	1st Qu.: 2	1st Qu.: 14.0
Median : 10.0	Median : 8.00	Median : 13	Median : 29.0
Mean : 254.7	Mean : 55.02	Mean : 85	Mean : 333.1
3rd Qu.: 52.0	3rd Qu.: 41.00	3rd Qu.: 60	3rd Qu.: 127.0
Max. : 3519.0	Max. : 894.00	Max. : 1145	Max. : 4653.0

a25.34	a35.44	a45.54	a55.64
Min. : 0	Min. : 0.0	Min. : 0	Min. : 0
1st Qu.: 22	1st Qu.: 50.0	1st Qu.: 52	1st Qu.: 79
Median : 76	Median : 149.0	Median : 257	Median : 351
Mean : 381	Mean : 674.4	Mean : 1735	Mean : 2594
3rd Qu.: 214	3rd Qu.: 462.0	3rd Qu.: 977	3rd Qu.: 1408
Max. : 3255	Max. : 6457.0	Max. : 26660	Max. : 45796

a65.74	a75.84	a85.94	a95.
Min. : 0	Min. : 0	Min. : 0	Min. : 0
1st Qu.: 99	1st Qu.: 105	1st Qu.: 45	1st Qu.: 2
Median : 486	Median : 636	Median : 301	Median : 31
Mean : 4923	Mean : 8250	Mean : 5657	Mean : 1059
3rd Qu.: 1995	3rd Qu.: 4895	3rd Qu.: 4060	3rd Qu.: 574
Max. : 72566	Max. : 118780	Max. : 98187	Max. : 19663

► Code

```

rows columns discrete_columns continuous_columns all_missing_columns
1  41      12                0                12                0
total_missing_values complete_rows total_observations memory_usage
1                0                41                492            5288

```

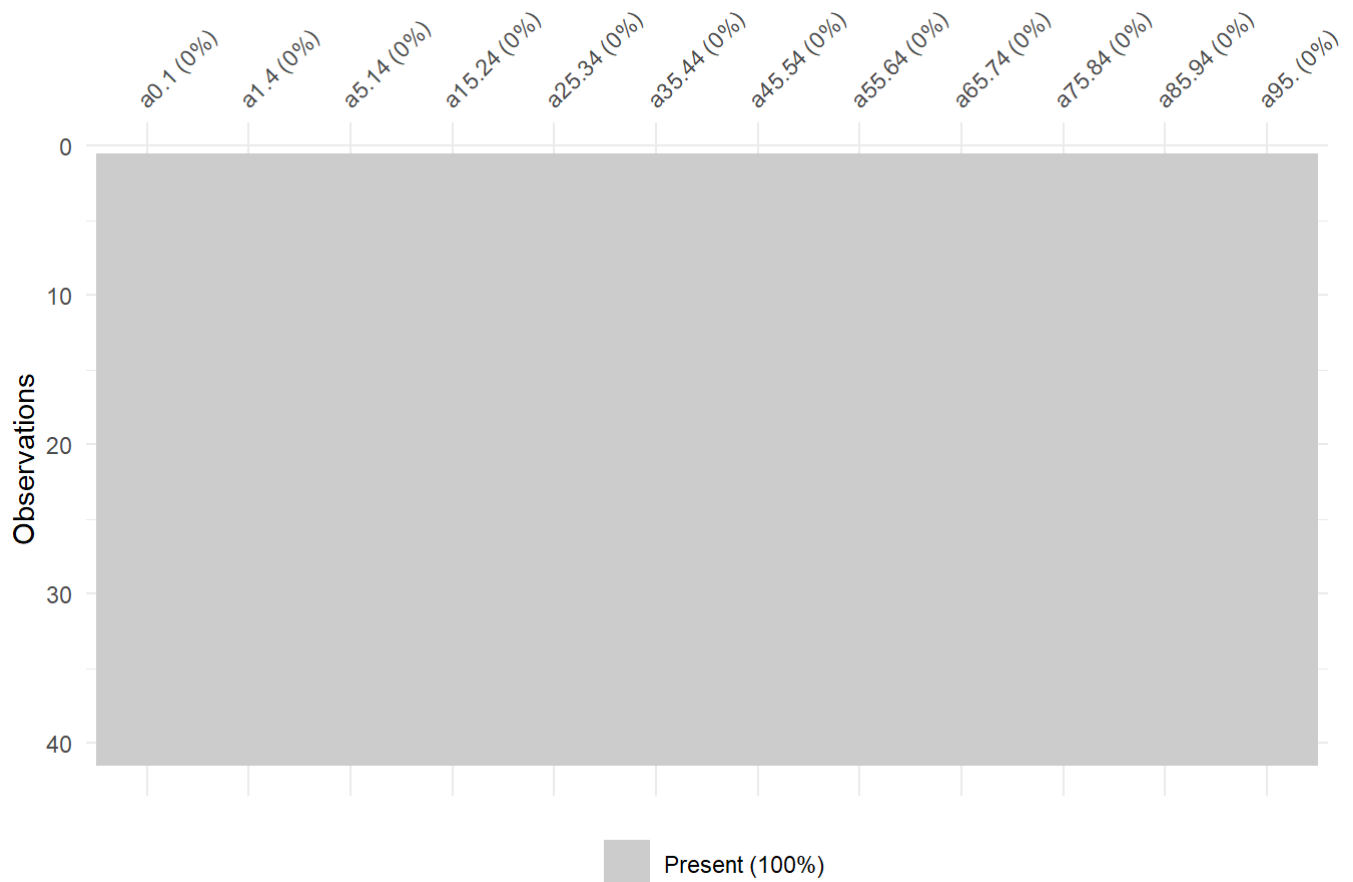
► Code

```
[1] 41 12
```

D'après ce résumé il n'y a aucune valeur manquantes et nous avons 41 observation.

Détection des valeurs manquantes

```
[1] 0
```



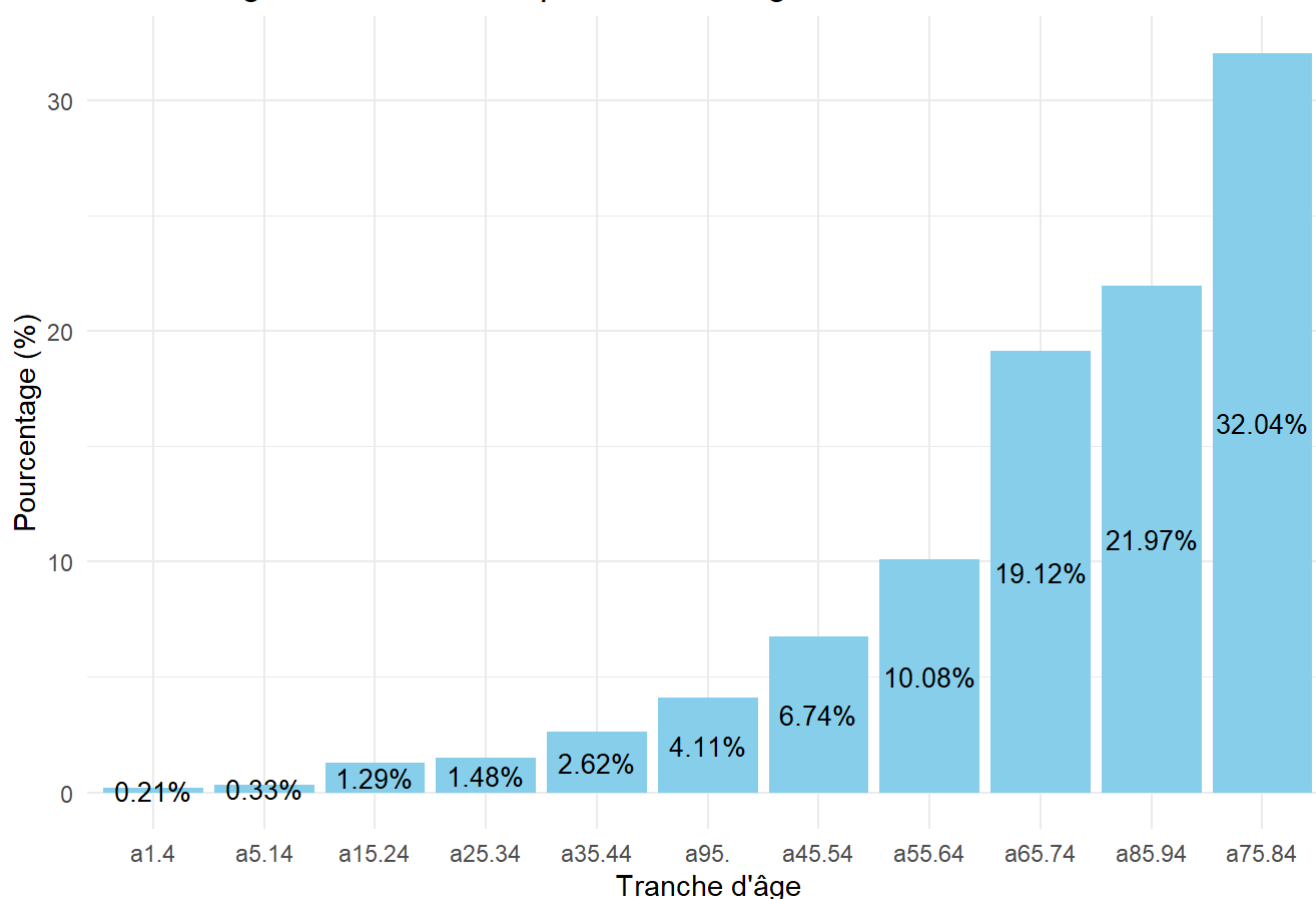
Pas de données manquantes.

Description des données

Profil colonne

	Total pourcentage	
a1.4	2256	0.2137110
a5.14	3485	0.3301343
a15.24	13659	1.2939180
a25.34	15623	1.4799679
a35.44	27649	2.6191917
a45.54	71126	6.7377711
a55.64	106369	10.0763430
a65.74	201834	19.1197492
a75.84	338243	32.0417835
a85.94	231948	21.9724506
a95.	43439	4.1149796

Pourcentage du total des cas par tranche d'âge



Profil ligne

	Total	Pourcentage
Abus_alcool	6327	0.59348656
Accidents_transport	16156	1.51546845
Arthrite_rhumatoide_osteoarthrite	1287	0.12072344
Asthme	2512	0.23563114
Autres_maladies_toute_cause	337629	31.67034528
Autres_malformations_congen_anom_chromosomiques	1523	0.14286076
Autres_symptomes_etats_morbide_mal_definis	38513	3.61260439
Autres_troubles_mentaux_comportement	17171	1.61067769
Autres_tumeurs	10596	0.99392818
Autres_tumeurs_malignes	270100	25.33597605
Cardiopathies_ischemiques	88345	8.28695596
Causes_inconnues_non_precisees	28186	2.64390900
Infections_periode_perinatale	3552	0.33318544
Chutes_accidentelles	17832	1.67268095
Complic_grossesse_accouch_puerperalite	150	0.01407033
Diabete_sucré	18090	1.69688192
evenements_dont_intention_non_determinee	2669	0.25035809
Grippe	1179	0.11059280
Hepatites_virales	1068	0.10018076
Homicides	989	0.09277038
Infection_meningocoques	147	0.01378892
Infections_peau_tissu_cellulaire_sous-cutane	3444	0.32305480
Intoxications_accidentelles	1564	0.14670665
Maladie_chronique_foie	23610	2.21467010

By Mariam Sylla

Maladies_cerebrovasculaires	98880	9.27516221
Maladies_rein_uretere	13647	1.28011872
Maladies_sang_organes_hematopoiétiques	4492	0.42135951
Malformations_congenitales_appareil_circulatoire	1893	0.17756758
Malformations_congenitales_système_nerveux	505	0.04737011
Melanome_malin_peau	2207	0.20702147
Meningites	625	0.05862638
Pharmacodependance_toxicomanie	222	0.02082409
Pneumonie	14794	1.38770985
SIDA_maladies_VIH	809	0.07588599
Suicides	20407	1.91422163
Syndrome_de_mort_subite_nourrisson	854	0.08010708
Tuberculose	2802	0.26283378
Ulcere_gastro-duodenal	2981	0.27962438
06_Abus_alcool_et_psychose_alcoolique	2956	0.27727932
06_Accidents_de_transport	4781	0.44846835
06_Arthrite_rhumatoide_osteoarthrite	579	0.05431148

Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Variance	0.546	0.246	0.125	0.030	0.016	0.011	0.003
% of var.	55.716	25.109	12.765	3.070	1.673	1.099	0.338
Cumulative % of var.	55.716	80.825	93.590	96.660	98.333	99.432	99.770
	Dim.8	Dim.9	Dim.10	Dim.11			
Variance	0.001	0.001	0.000	0.000			
% of var.	0.110	0.081	0.029	0.010			
Cumulative % of var.	99.880	99.961	99.990	100.000			

Rows (the 10 first)

	Iner*1000	Dim.1	ctr
Abus_alcool	10.247	-0.064	0.004
Accidents_transport	128.358	0.164	0.075
Arthrite_rhumatoide_osteoarthrite	0.177	-0.121	0.003
Asthme	0.046	-0.062	0.002
Autres_maladies_toute_cause	25.361	-0.060	0.211
Autres_malformations_congen_anom_chromosomiques	41.616	5.336	7.447
Autres_symptomes_etats_morbide_mal_definis	21.099	-0.087	0.050
Autres_troubles_mentaux_comportement	8.719	-0.123	0.044
Autres_tumeurs	0.342	-0.061	0.007
Autres_tumeurs_malignes	48.326	-0.103	0.488

	cos2	Dim.2	ctr
Abus_alcool	0.002	0.764	1.409
Accidents_transport	0.003	2.637	42.809
Arthrite_rhumatoide_osteoarthrite	0.100	-0.263	0.034
Asthme	0.200	0.031	0.001
Autres_maladies_toute_cause	0.045	-0.182	4.264
Autres_malformations_congen_anom_chromosomiques	0.978	0.184	0.020
Autres_symptomes_etats_morbide_mal_definis	0.013	-0.380	2.120
Autres_troubles_mentaux_comportement	0.028	-0.419	1.147
Autres_tumeurs	0.109	0.041	0.007
Autres_tumeurs_malignes	0.055	0.075	0.586

	cos2	Dim.3	ctr
Abus_alcool	0.338	-0.856	3.472
Accidents_transport	0.821	1.070	13.865
Arthrite_rhumatoide_osteoarthrite	0.475	0.090	0.008
Asthme	0.050	-0.085	0.014
Autres_maladies_toute_cause	0.414	0.206	10.770
Autres_malformations_congen_anom_chromosomiques	0.001	-0.058	0.004
Autres_symptomes_etats_morbide_mal_definis	0.247	0.493	7.010
Autres_troubles_mentaux_comportement	0.324	0.512	3.374
Autres_tumeurs	0.048	-0.047	0.018
Autres_tumeurs_malignes	0.030	-0.402	32.765

	cos2	
Abus_alcool	0.424	
Accidents_transport	0.135	
Arthrite_rhumatoide_osteoarthrite	0.055	
Asthme	0.375	
Autres_maladies_toute_cause	0.532	
Autres_malformations_congen_anom_chromosomiques	0.000	
Autres_symptomes_etats_morbide_mal_definis	0.416	
Autres_troubles_mentaux_comportement	0.484	

Autres_tumeurs	0.065
Autres_tumeurs_malignes	0.849

Columns (the 10 first)

	Iner*1000	Dim.1	ctr
a0.1	535.155	7.380	97.649
a1.4	18.626	1.262	0.617
a5.14	12.580	0.549	0.180
a15.24	126.515	0.219	0.112
a25.34	61.785	0.143	0.055
a35.44	38.808	0.032	0.005
a45.54	37.775	-0.038	0.018
a55.64	32.966	-0.070	0.089
a65.74	19.506	-0.092	0.295
a75.84	19.819	-0.098	0.555

	cos2	Dim.2	ctr
a0.1	0.997	-0.387	0.594
a1.4	0.181	0.690	0.410
a5.14	0.078	1.298	2.238
a15.24	0.005	2.726	38.682
a25.34	0.005	1.869	20.788
a35.44	0.001	1.064	11.936
a45.54	0.003	0.498	6.731
a55.64	0.015	0.228	2.101
a65.74	0.083	-0.004	0.001
a75.84	0.153	-0.192	4.732

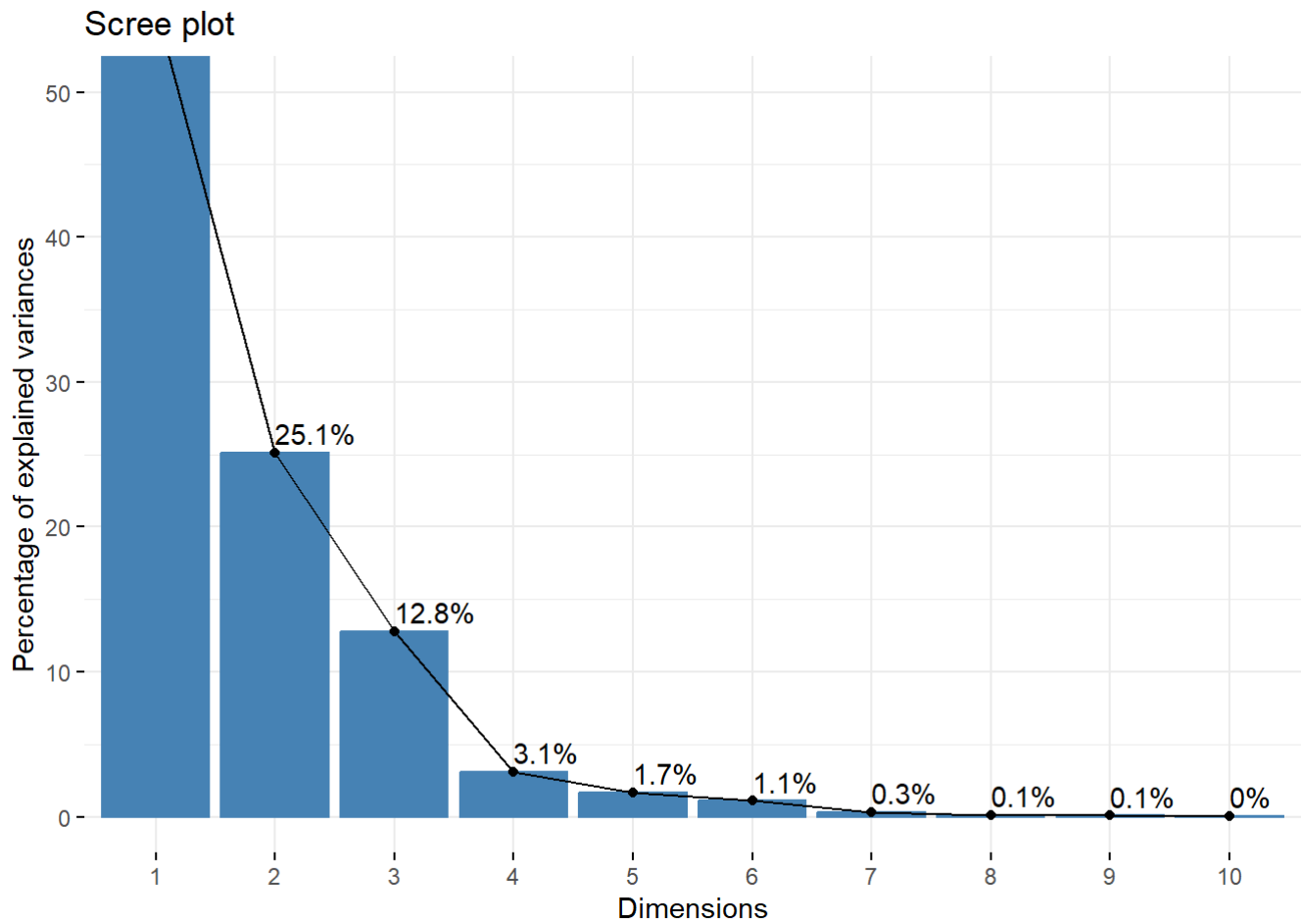
	cos2	Dim.3	ctr
a0.1	0.003	-0.096	0.072
a1.4	0.054	0.625	0.659
a5.14	0.438	0.749	1.467
a15.24	0.753	1.360	18.946
a25.34	0.828	0.559	3.658
a35.44	0.757	-0.146	0.441
a45.54	0.439	-0.499	13.291
a55.64	0.157	-0.508	20.565
a65.74	0.000	-0.273	11.286
a75.84	0.588	0.052	0.687

	cos2
a0.1	0.000
a1.4	0.044
a5.14	0.146
a15.24	0.187
a25.34	0.074
a35.44	0.014
a45.54	0.440
a55.64	0.781
a65.74	0.724
a75.84	0.043

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	5.463308e-01	55.716025129	55.71603
Dim.2	2.462091e-01	25.108943873	80.82497
Dim.3	1.251685e-01	12.764963867	93.58993
Dim.4	3.010493e-02	3.070167134	96.66010
Dim.5	1.640435e-02	1.672952041	98.33305

By Mariam Sylla

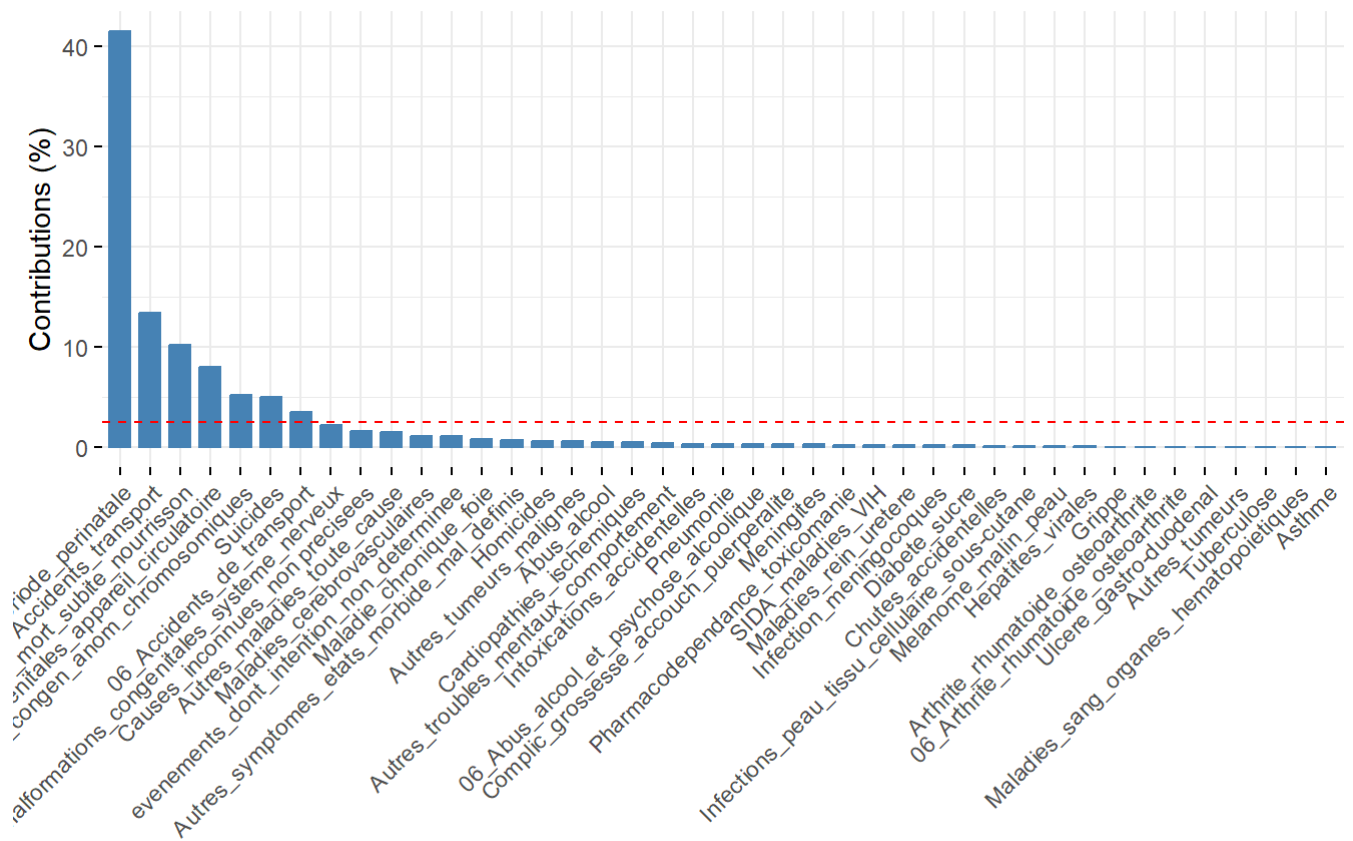
Dim.6	1.077731e-02	1.099094033	99.43215
Dim.7	3.316987e-03	0.338273612	99.77042
Dim.8	1.078932e-03	0.110031867	99.88045
Dim.9	7.929661e-04	0.080868434	99.96132
Dim.10	2.842606e-04	0.028989520	99.99031
Dim.11	9.502139e-05	0.009690491	100.00000



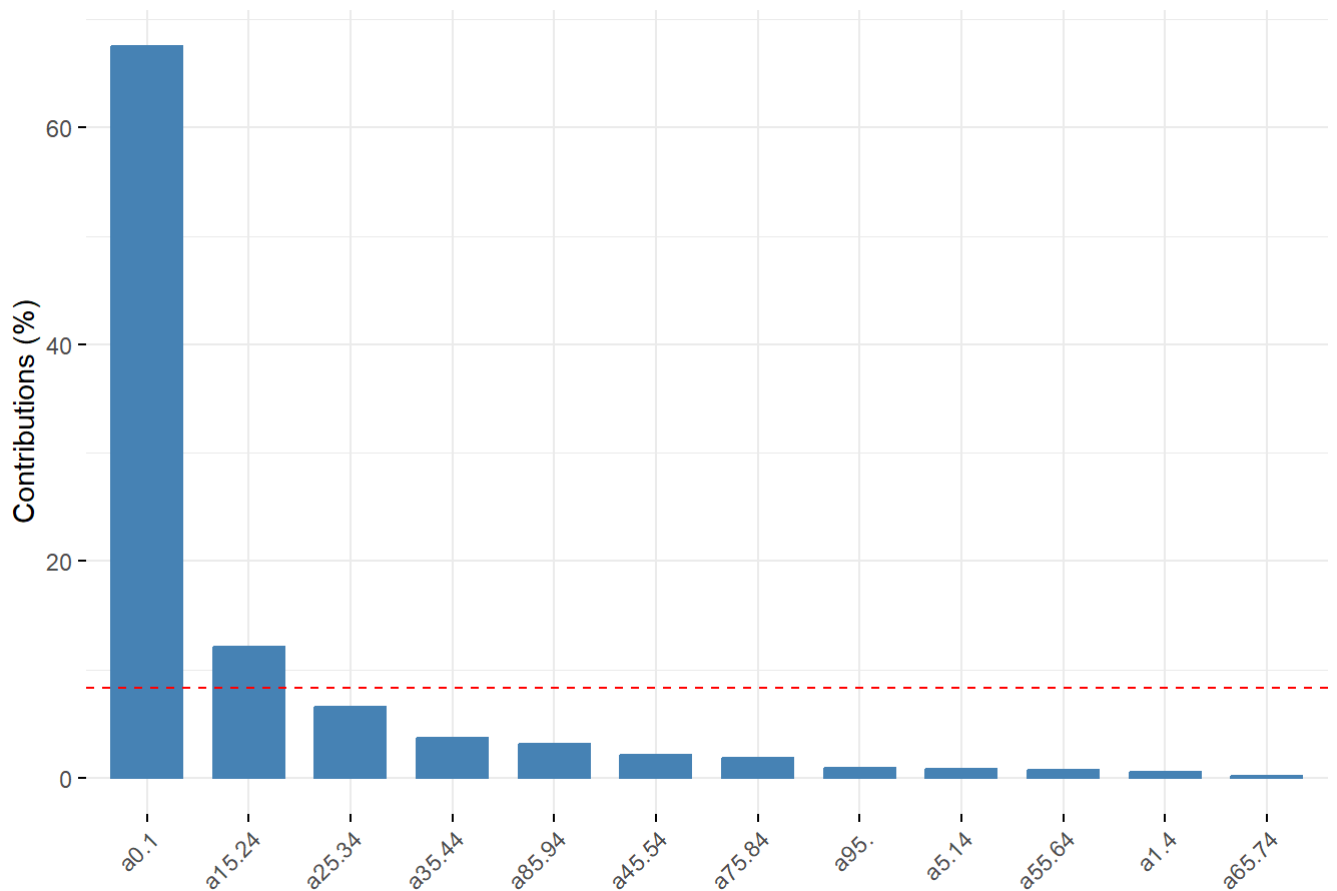
Les deux premiers axes expliquent à peu près 81% des informations ce qui représente un bon taux.
Donc on peut choisir les deux axes.

Contribution des variables sur les deux premiers axes

Contribution of rows to Dim-1-2



Contribution of columns to Dim-1-2



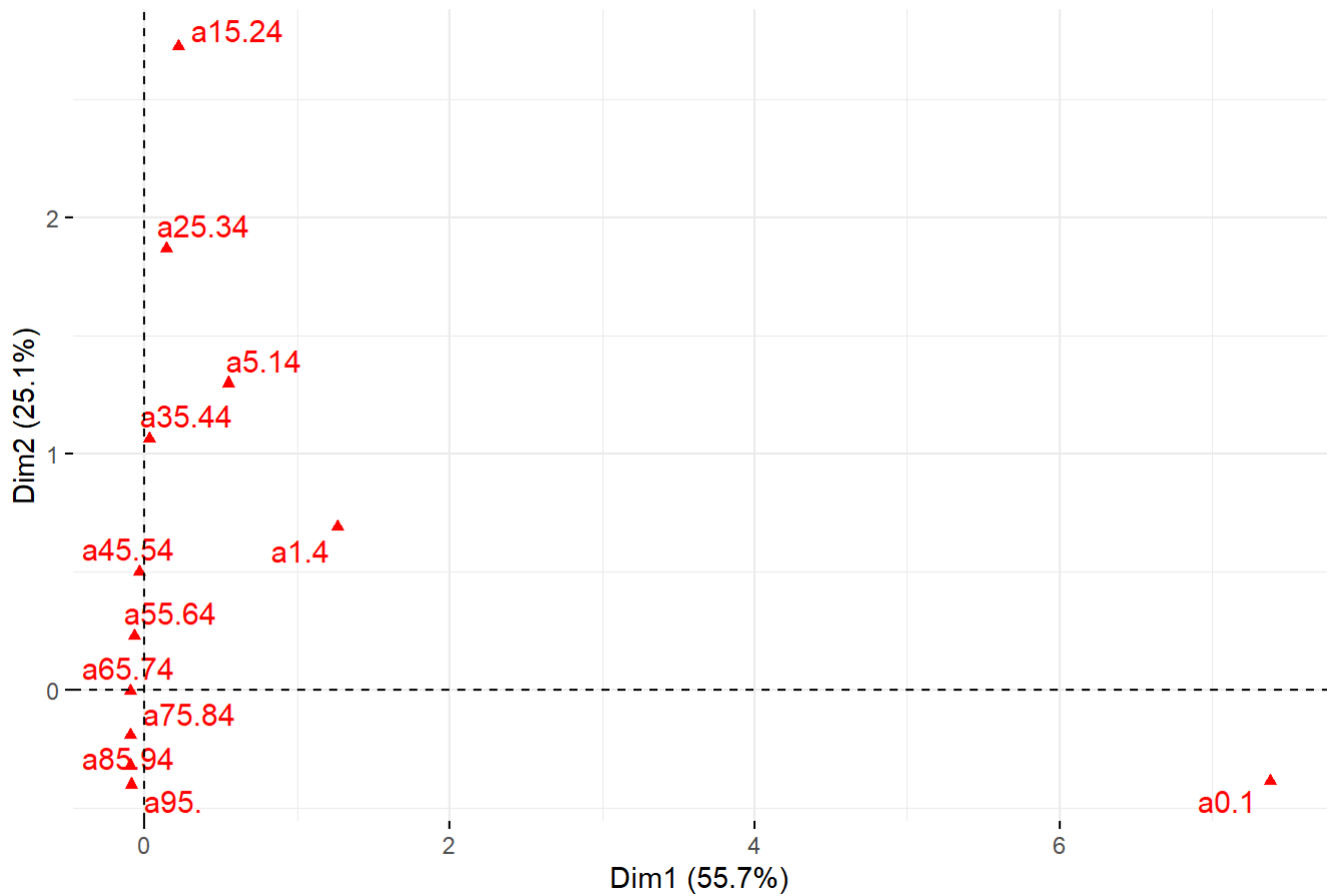
Les graphiques

Graphique des maladies (profil ligne)



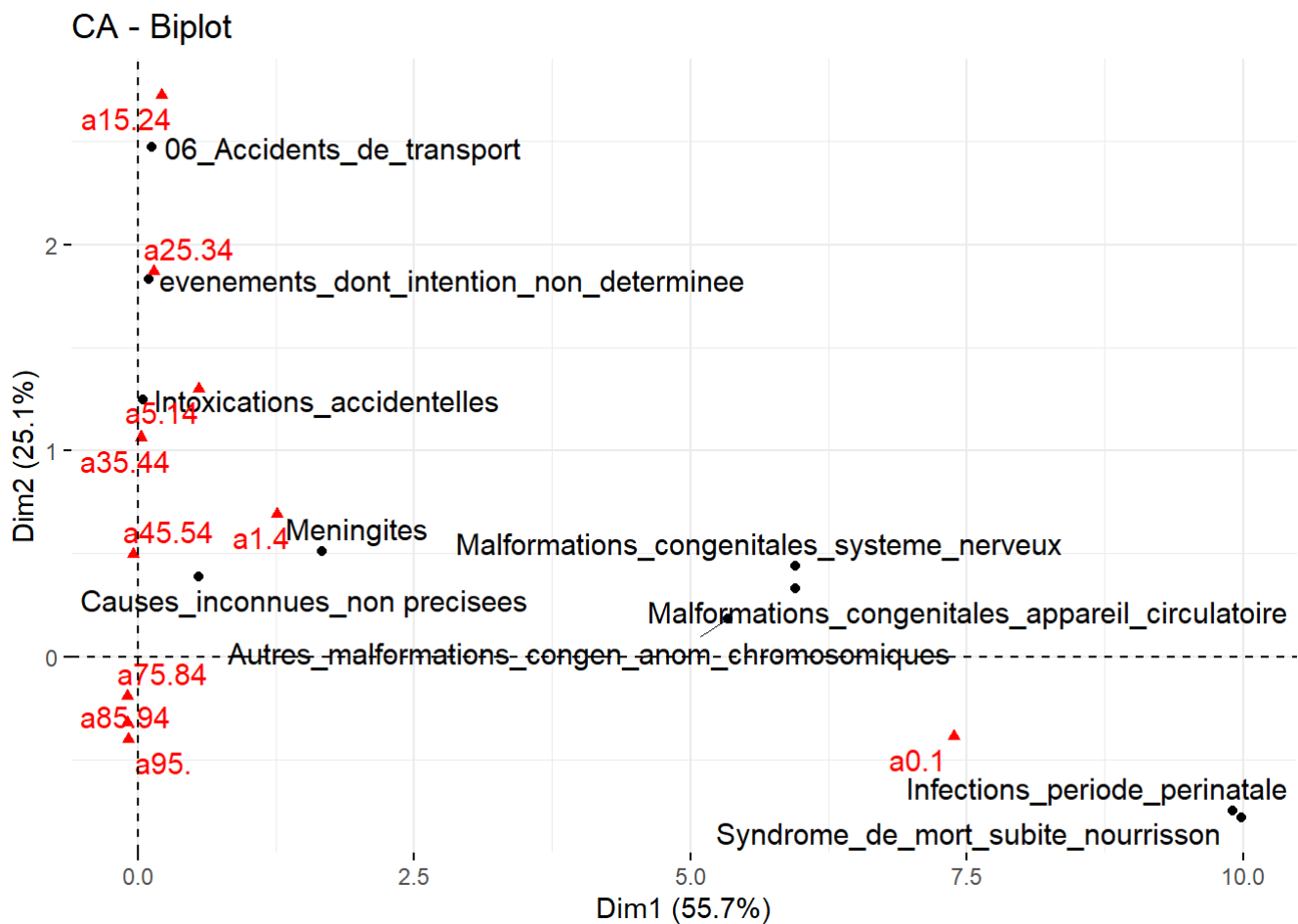
Graphique des tranches d'age (profil colonne)

graphique des tranches d'age



Les individus de la tranche d'age "a15-24" sont opposés à ceux de "a0-1"

Graphique final



Pour ce graphique, étant donné le nombre élevé de variables, le top 10 des variables ont été sélectionnées selon le cos2 et la contribution

Conclusion [↗](#)

Les individus de la tranche d'âge "a15-24" opposés à ceux de "a0-1" ont des causes de décès différentes. La première tranche meurent souvent d'accident de transport et la deuxième d'infection période périnatale et de syndrome de mort subite de nourrisson. La tranche d'âge "a25-34" a pour cause de décès événement dont l'intention non déterminée.