

Présentation Datamining et datavisualisation

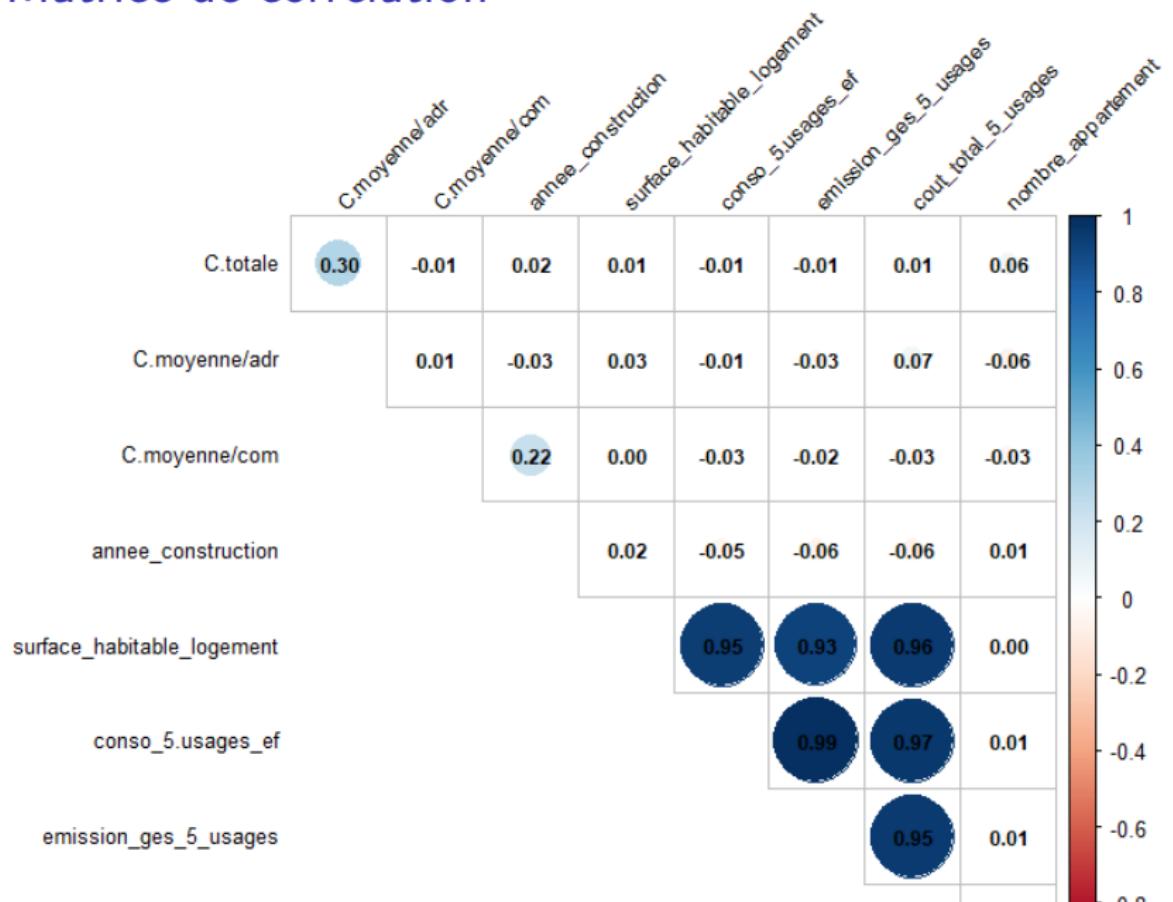
Kenny Jean-elie Ibrahima Caba Bah Fatou Diop Ndeye

2025-12-12

Section 1

Analyse en Composantes Principales

Matrice de corrélation

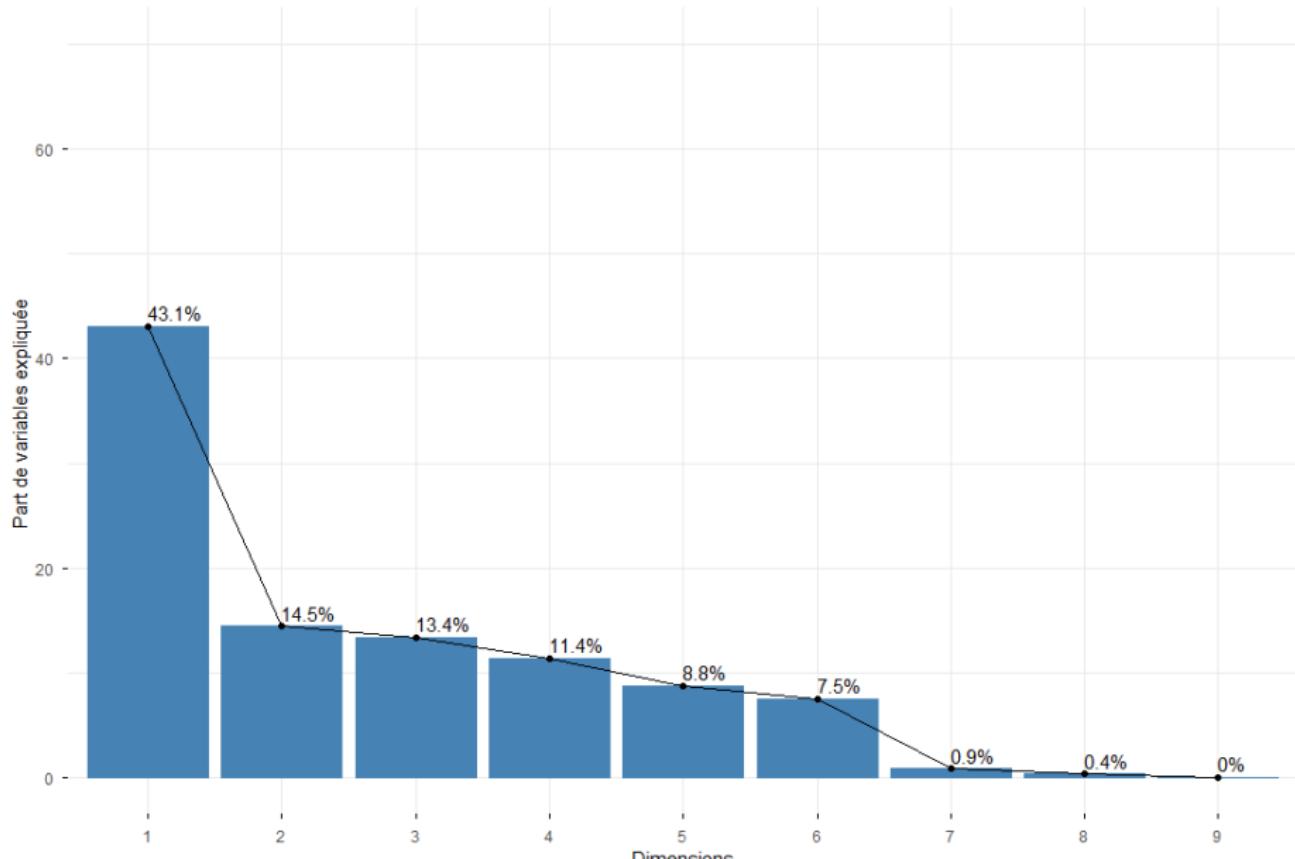


Etude des inerties

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3.88	43.1	43.1
comp 2	1.30	14.5	57.5
comp 3	1.21	13.4	70.9
comp 4	1.03	11.4	82.4
comp 5	0.79	8.8	91.2
comp 6	0.68	7.5	98.7
comp 7	0.08	0.9	99.6
comp	0.03	0.4	100.0

Représentation graphique

Représentation des inerties



Observation des variables

	Dim.1	Dim.2	Dim.3
C.totale	0.001	0.801	0.036
C.moyenne/adr	0.021	0.809	0.005
C.moyenne/com	-0.031	-0.019	0.774
annee_construction	-0.056	-0.024	0.773
surface_habitable_logement	0.972	0.013	0.071
conso_5.usages_ef	0.993	-0.022	0.002
emission_ges_5_usages	0.984	-0.046	0.000
cout_total_5_usages	0.986	0.036	-0.004
nombre_appartement	0.005	-0.003	-0.074

Qualité de représentation des variables sur l'axe F1F2

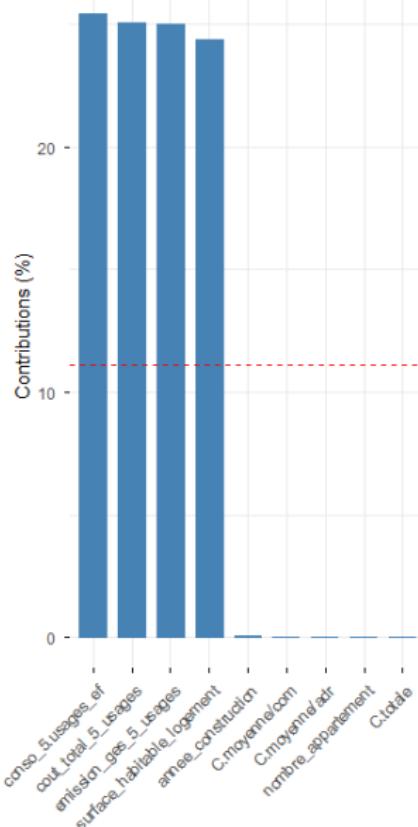
	x
C.totale	0.642
C.moyenne/adr	0.655
C.moyenne/com	0.001
annee_construction	0.004
surface_habitable_logement	0.944
conso_5.usages_ef	0.987
emission_ges_5_usages	0.971
cout_total_5_usages	0.974
nombre_appartement	0.000

Tableau des coordonnées des variables

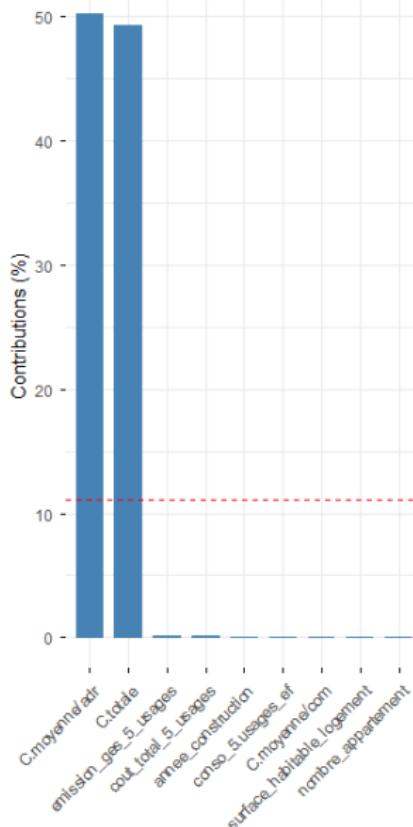
	F1	F2	F3
C.totale	0.00	0.8	0.0
C.moyenne/adr	0.02	0.8	0.0
C.moyenne/com	-0.03	0.0	0.8
annee_construction	-0.06	0.0	0.8
surface_habitable_logement	0.97	0.0	0.1
conso_5.usages_ef	0.99	0.0	0.0
emission_ges_5_usages	0.98	0.0	0.0
cout_total_5_usages	0.99	0.0	0.0
nombre_appartement	0.00	0.0	-0.1

Etude des contributions des variables

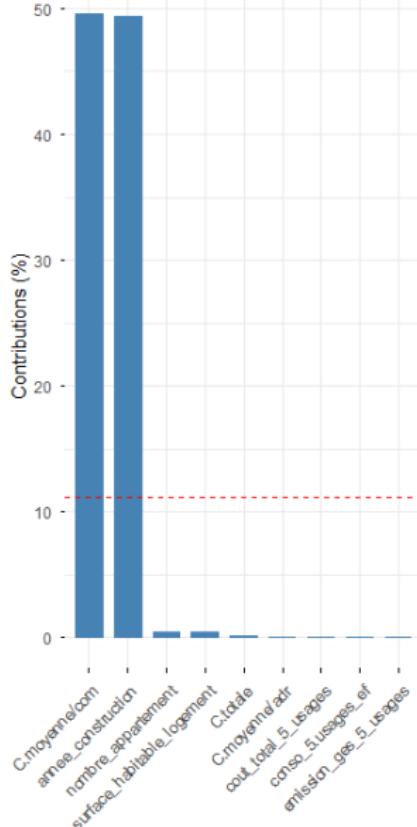
Contributions des variables Axes 1



Contributions des variables Axes 2

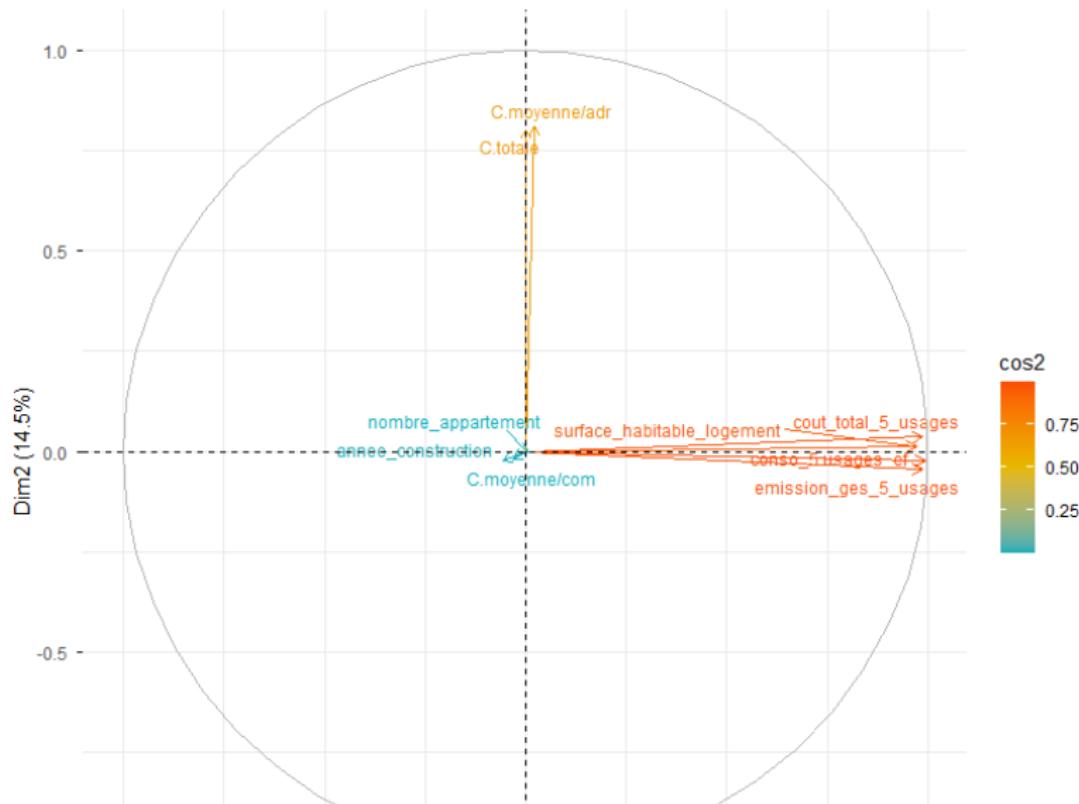


Contributions des variables Axes 3



Représentation des variables avec la qualité de représentation

Variables - PCA



Section 2

Analyse des Correspondances Factorielles

Présentation générale de l'ACF

Variables actives:

Étiquette_dpe, Étiquette_ges
pour quoi?

Test de Chi2

Resultat test

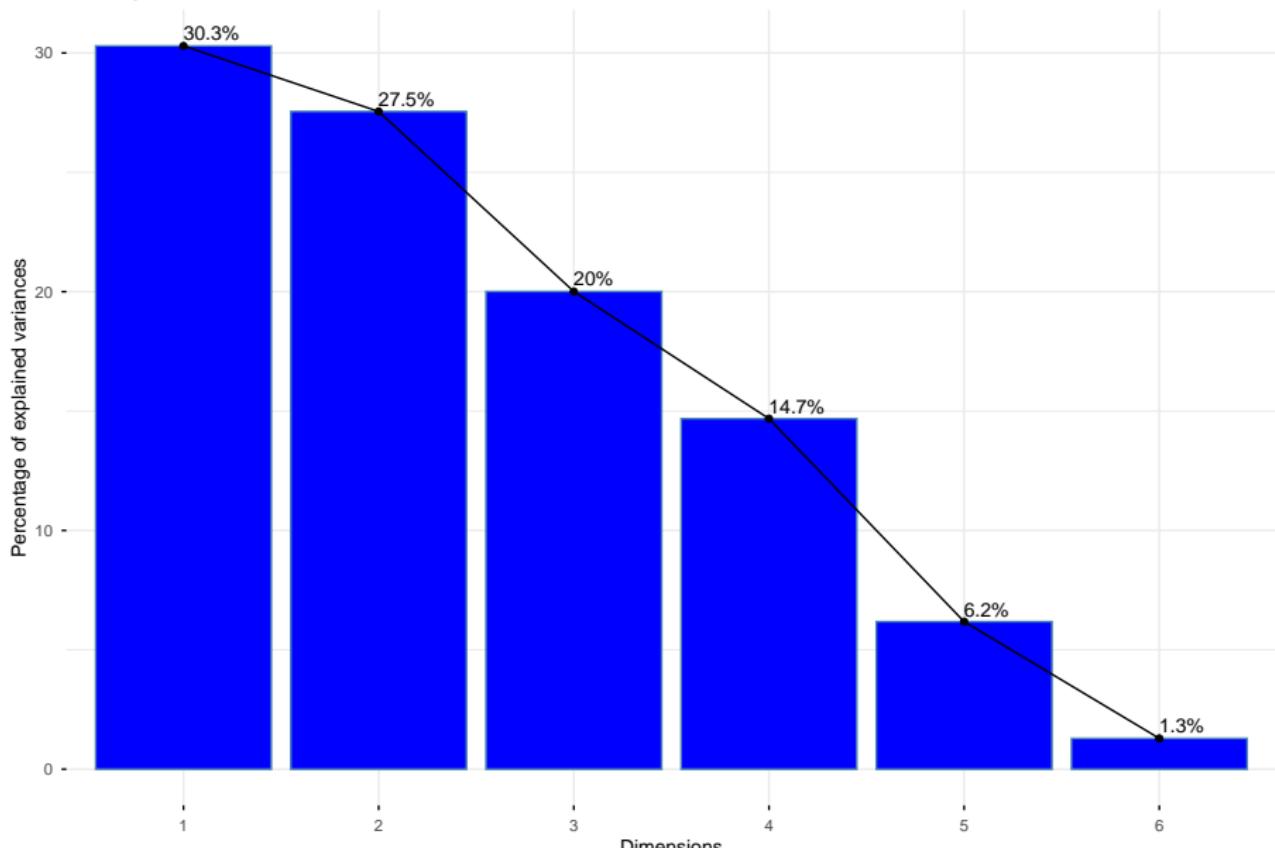
$\chi^2_{\text{obs}} = 14907.6$

df = 36

$p_{\text{value}} = 0$

Qualité globale de l'ACF

Scree plot

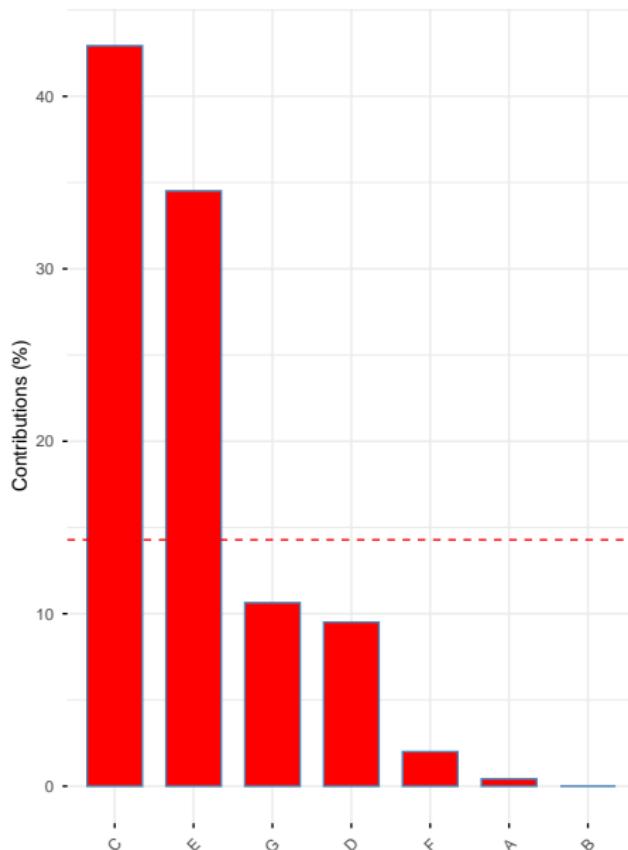


Qualité de IACF

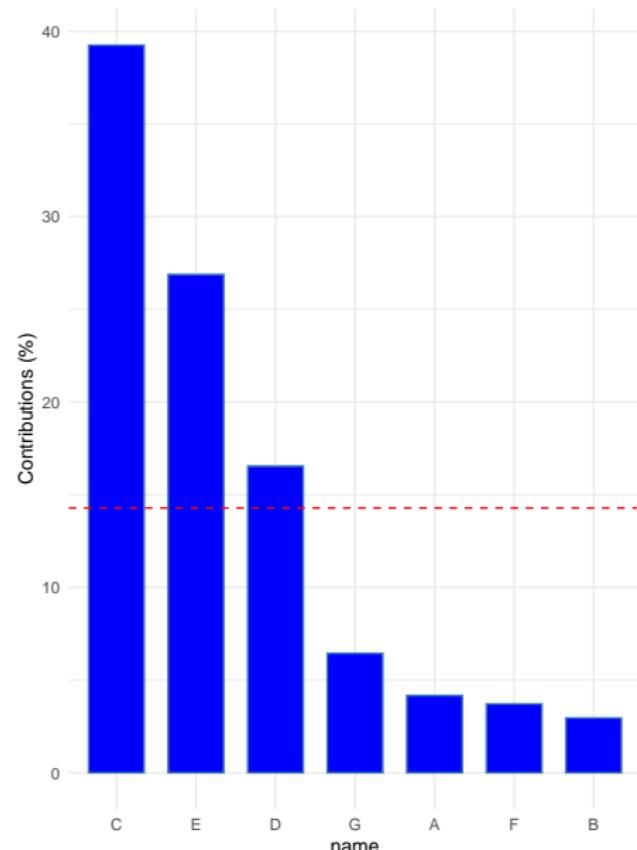
	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.63	30.3	30.3
dim 2	0.57	27.5	57.8
dim 3	0.42	20.0	77.9
dim 4	0.31	14.7	92.5
dim 5	0.13	6.2	98.7
dim 6	0.03	1.3	100.0

Interprétation de l'axe 1

Contribution des DPE à la Dime 1

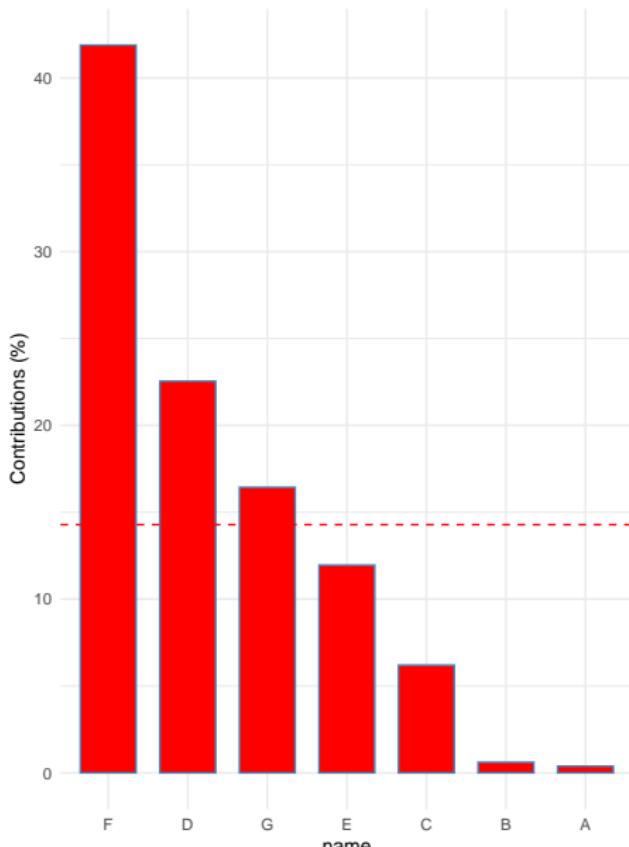


Contribution des GES à la Dim1

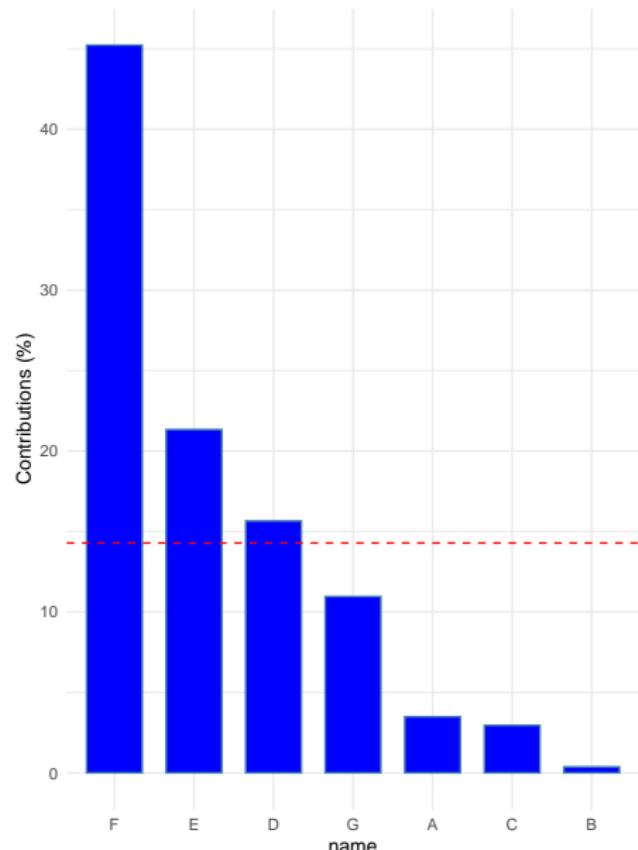


Interprétation de l'axe 2

Contribution des DPE à la Dime 2

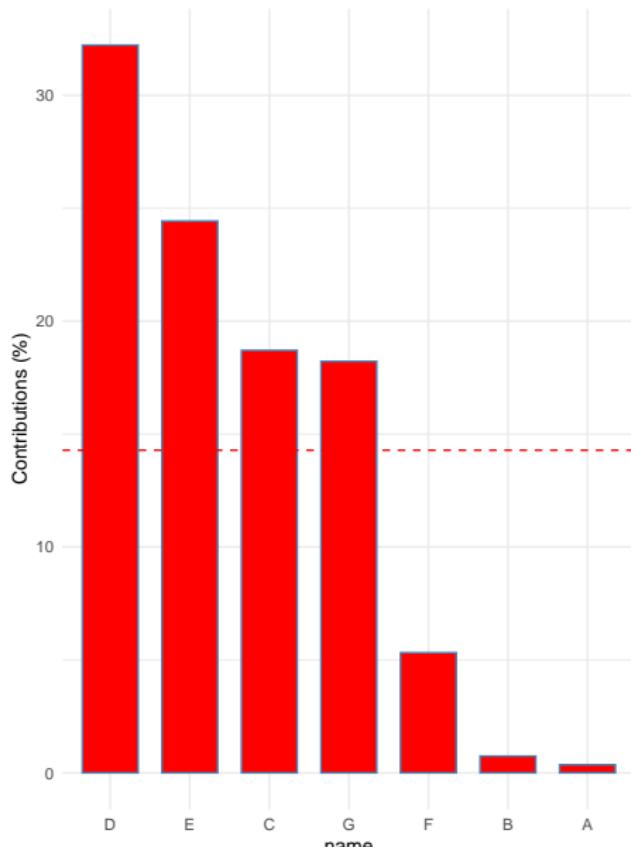


Contribution des GES à la Dim2

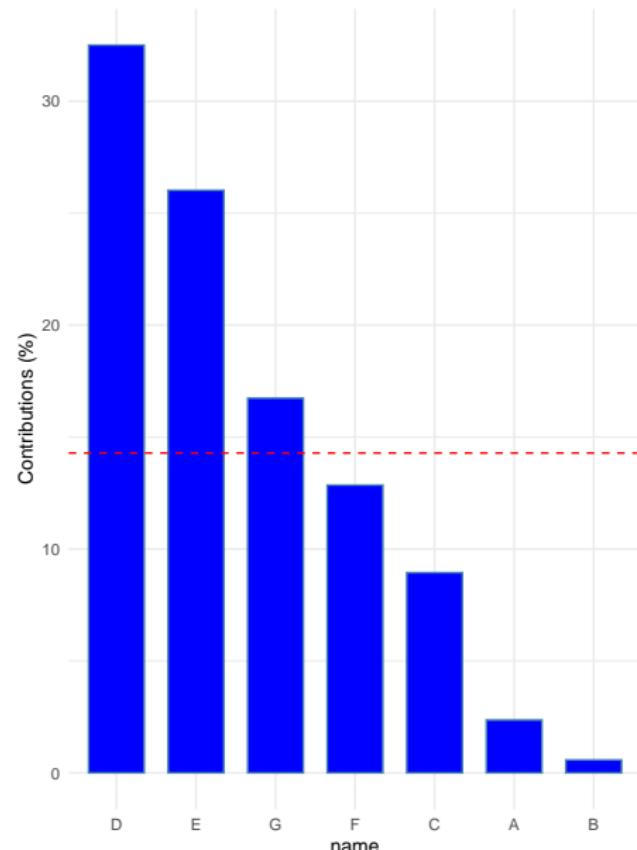


Interprétation de l'axe 3

Contribution des DPE à la Dime 3

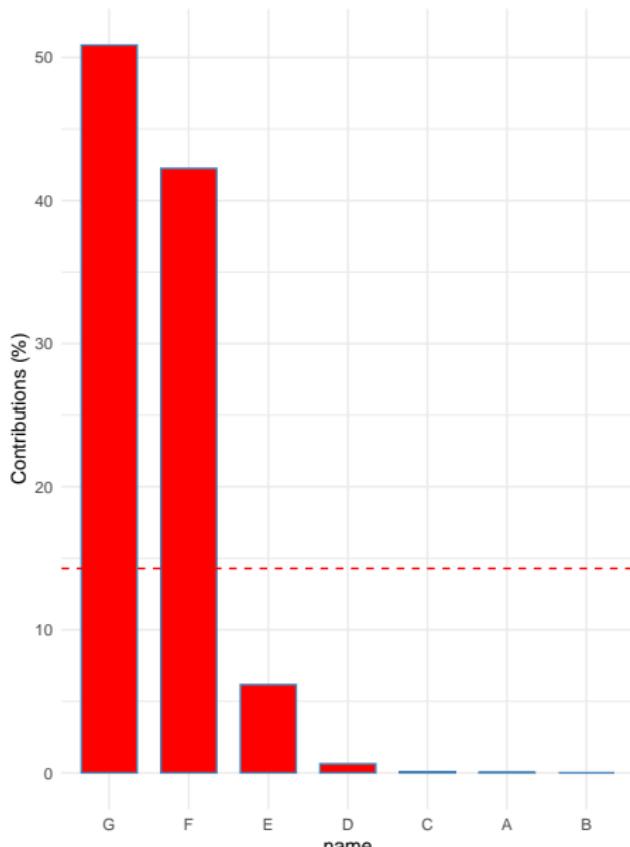


Contribution des GES à la Dim3

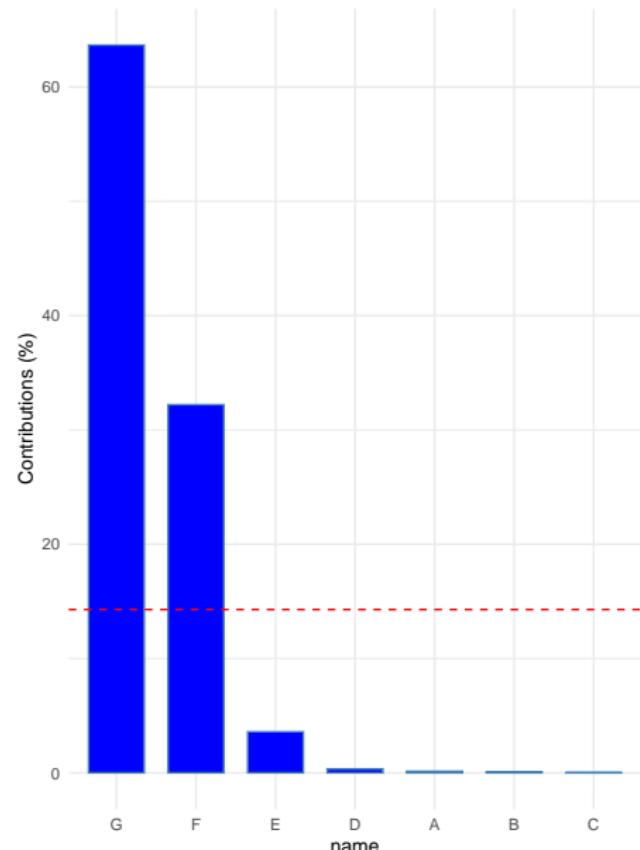


Interprétation de l'axe 4

Contribution des DPE à la Dime 3

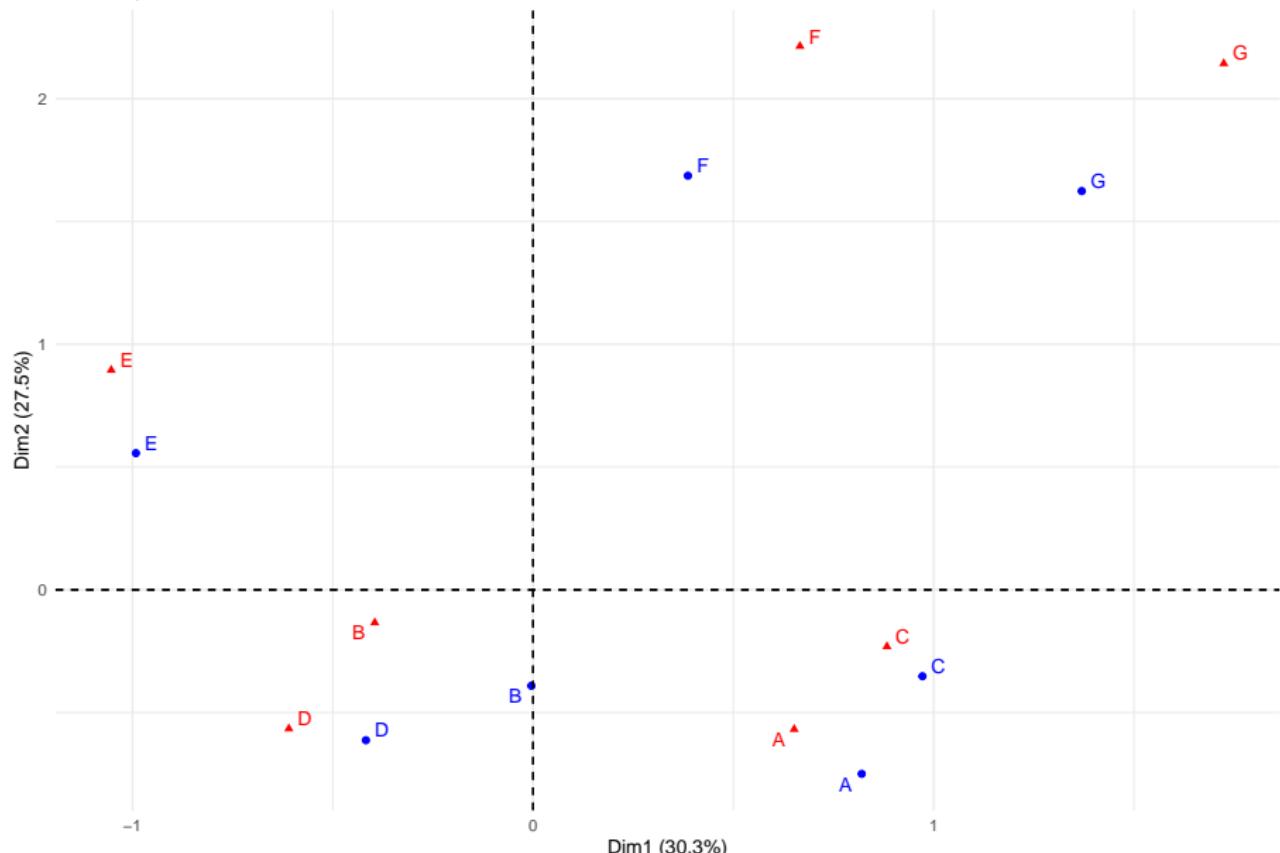


Contribution des GES à la Dim3



Synthèse finale

CA – Biplot



Section 3

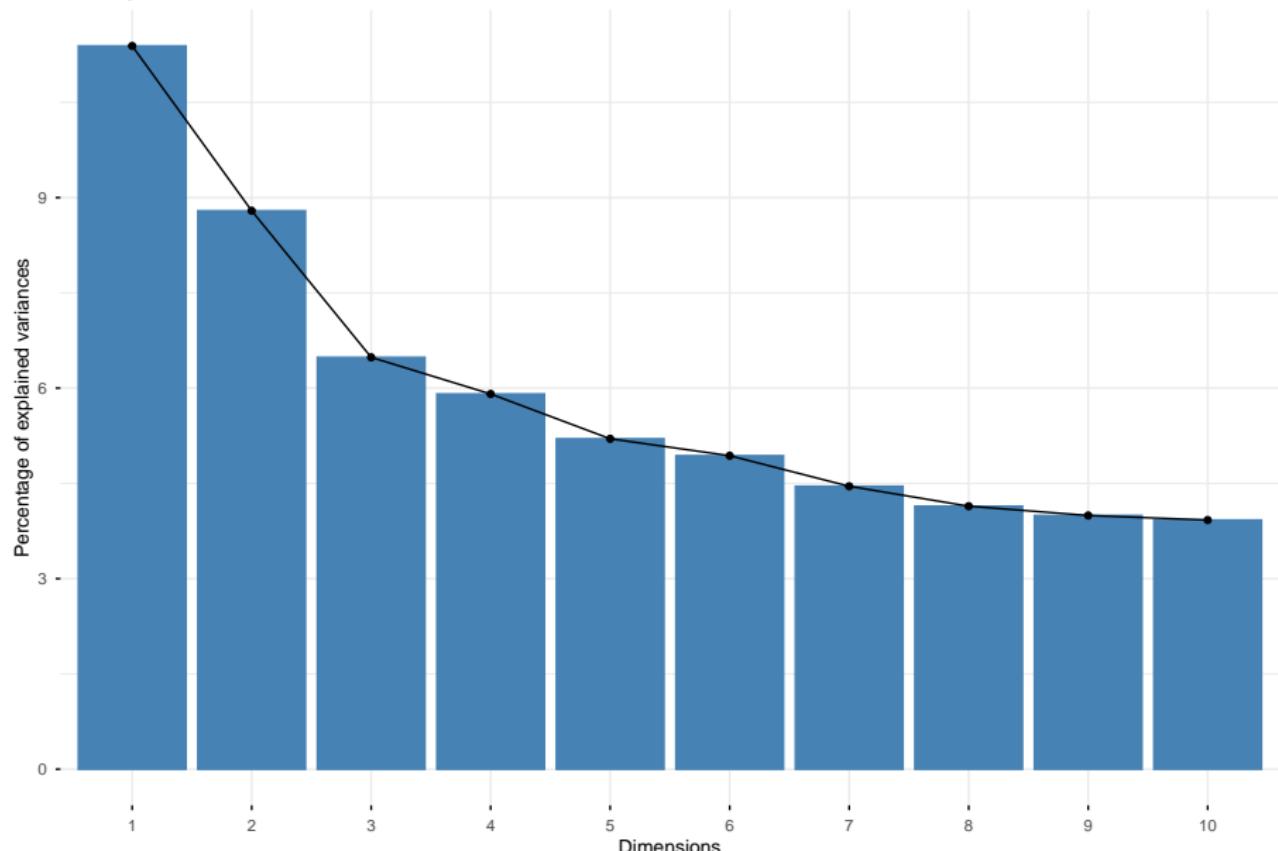
Analyse des correspondances multiples

Inertie

eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.3	11.4
dim 2	0.3	8.8
dim 3	0.2	6.5
dim 4	0.2	5.9
dim 5	0.2	5.2
dim 6	0.1	4.9
dim 7	0.1	4.5
dim 8	0.1	4.1
dim 9	0.1	4.0
dim	0.1	3.9
10		
dim	0.1	3.9
11		
dim	0.1	3.7
12		

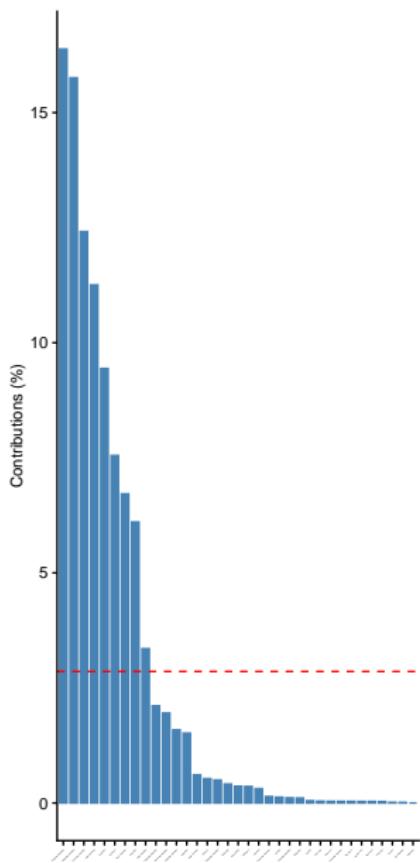
Inertie

Scree plot

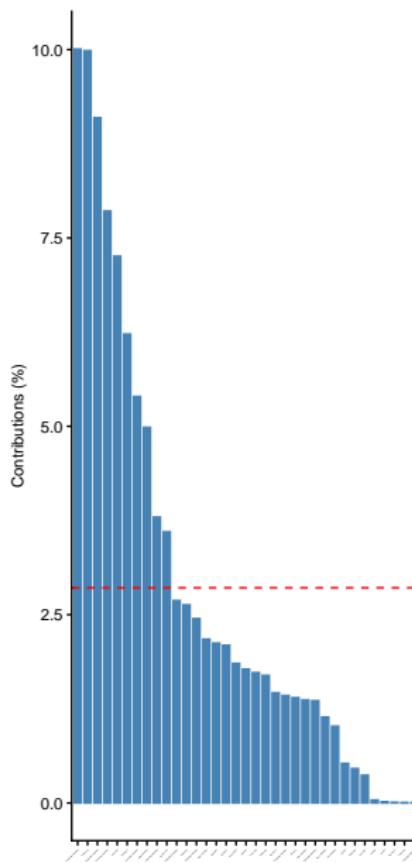


Contribution des variables

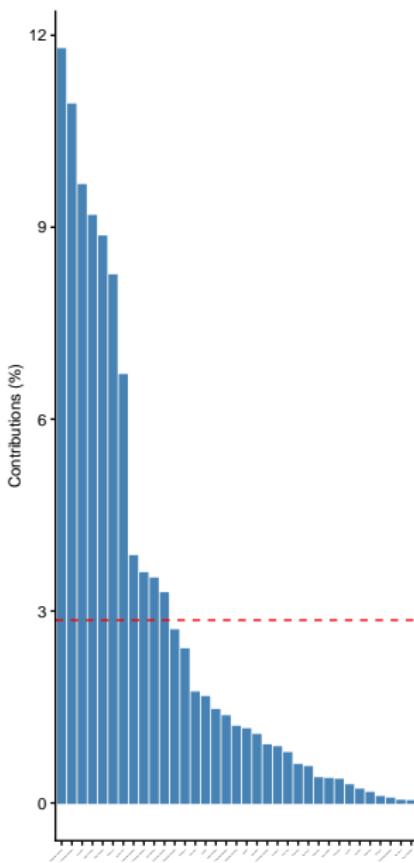
Contribution of variables to Dim-1



Contribution of variables to Dim-2



Contribution of variables to Dim-3

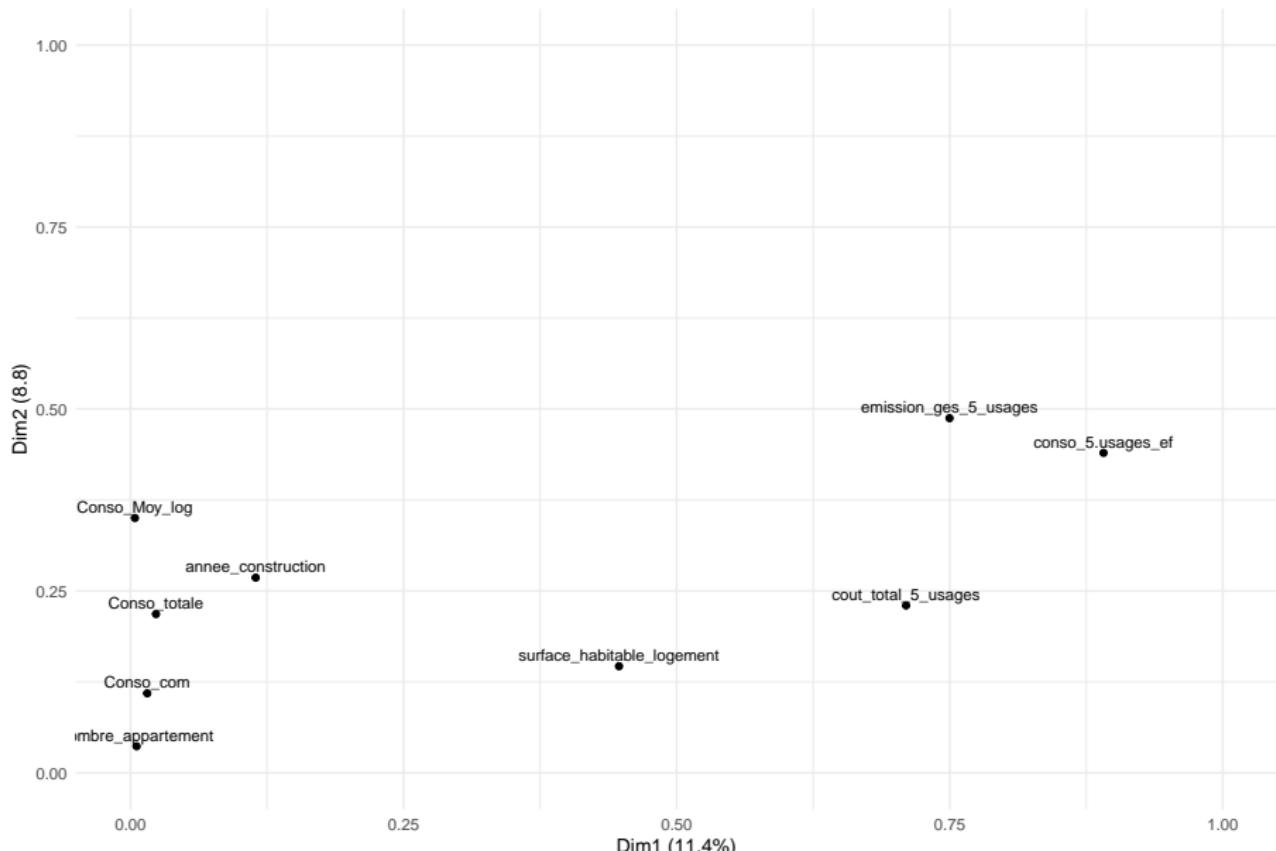


corrélation des var

	dim1	dim2	dim3	dim4	dim5
conso_5.usages_ef	0.89	0.44	0.29	0.10	0.02
emission_ges_5_usages	0.75	0.49	0.41	0.09	0.02
cout_total_5_usages	0.71	0.23	0.35	0.01	0.04
surface_habitable_logement	0.45	0.15	0.02	0.21	0.07
annee_construction	0.11	0.27	0.16	0.47	0.14
Conso_totale	0.02	0.22	0.06	0.04	0.58
Conso_com	0.02	0.11	0.02	0.25	0.01
nombre_appartement	0.01	0.04	0.11	0.21	0.37
Conso_Moy_log	0.00	0.35	0.27	0.17	0.11

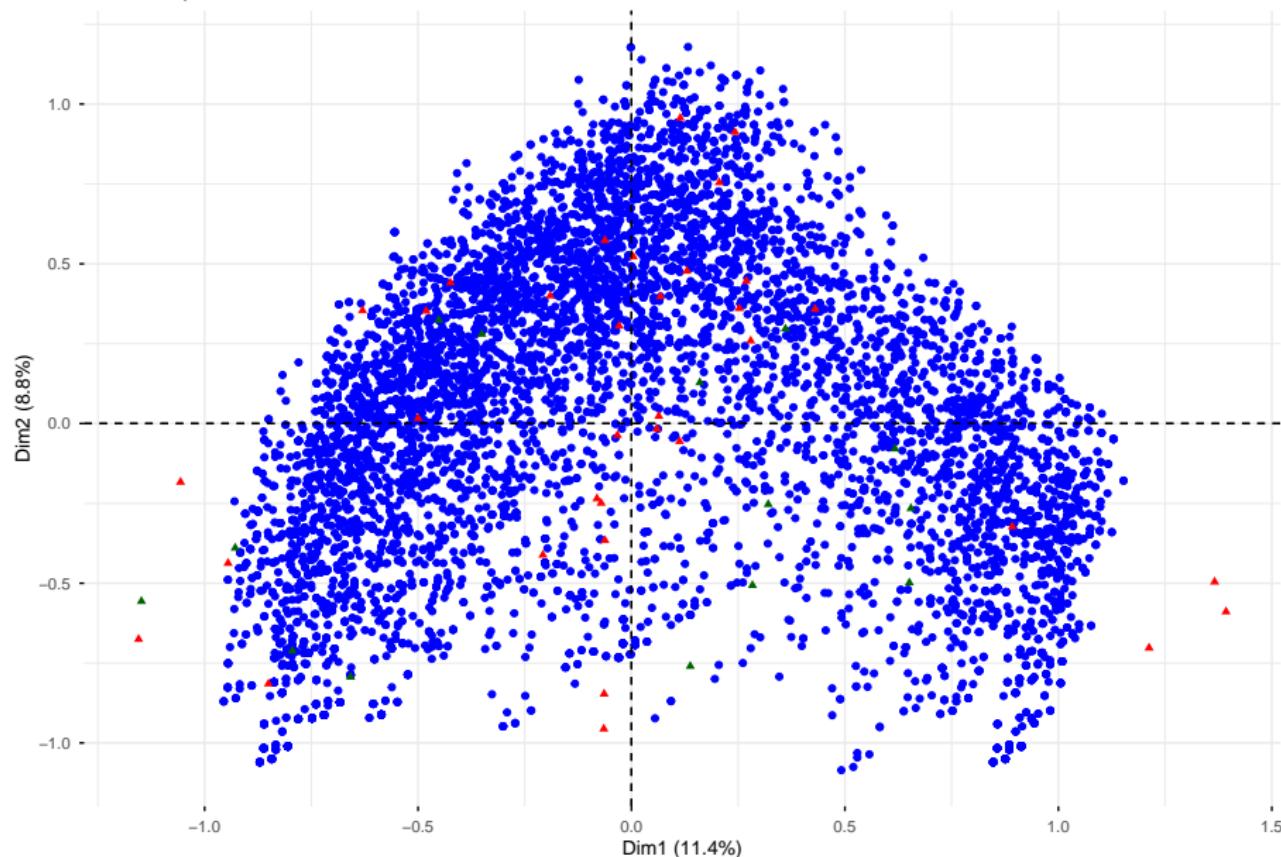
visualisation des corrélation

Variables – MCA



Nuages des points

MCA – Biplot



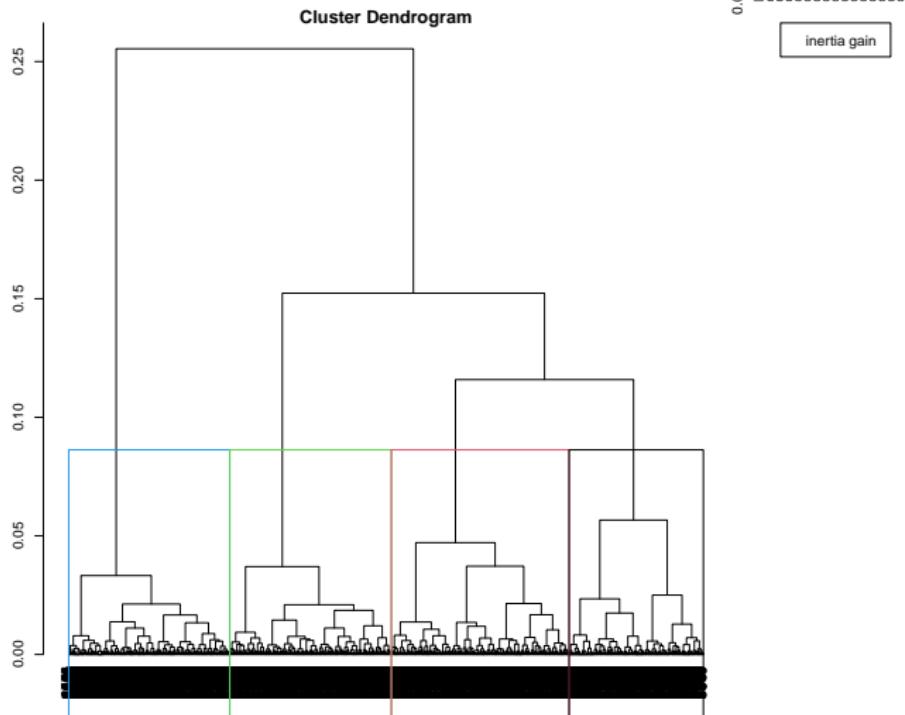
Section 4

Mise en Oeuvre de la classification non supervisée à
partir de l'ACM

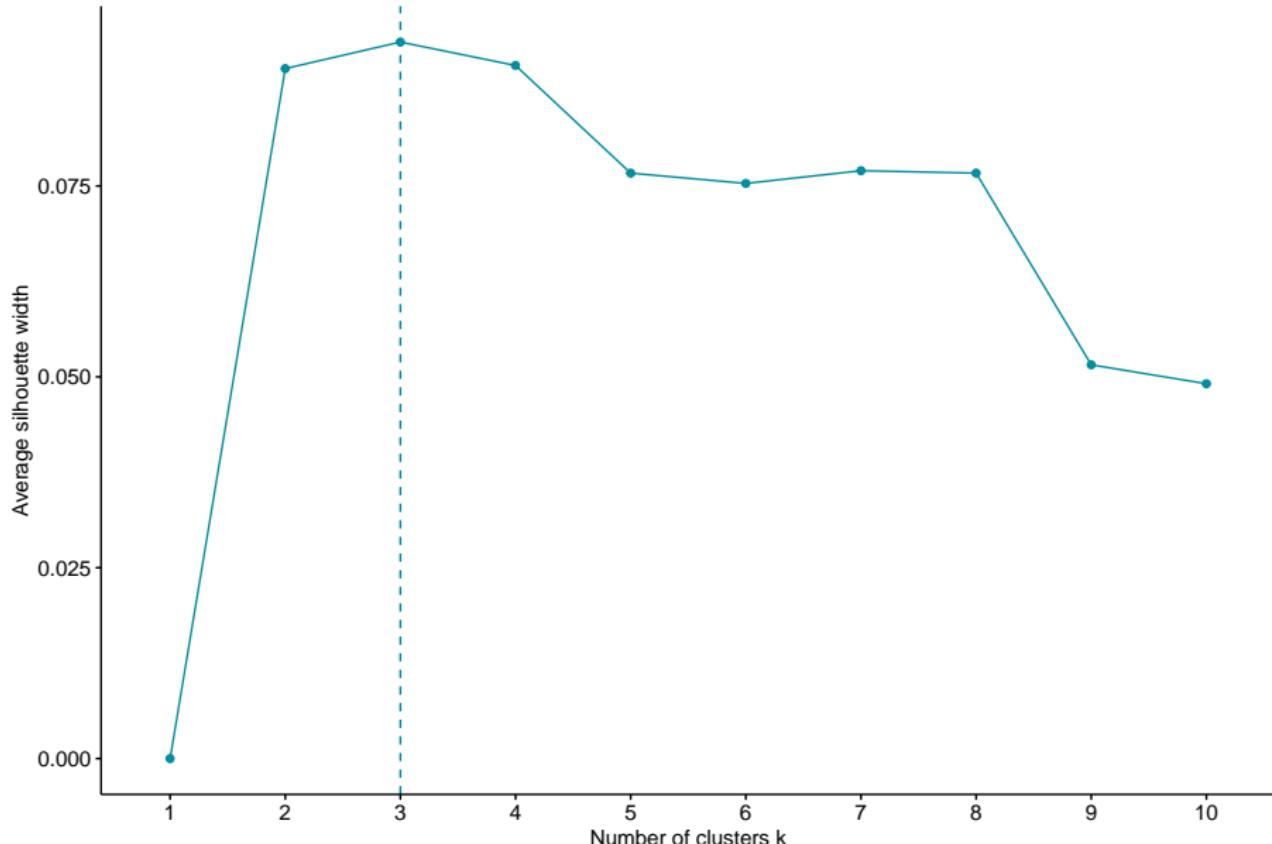
Mise en Oeuvre de la classification non supervisée à partir de l'ACM

On commence par faire une classification hiérarchique ascendante sur la sortie de l'ACM avec consolidation puis faire la méthode **k-means**

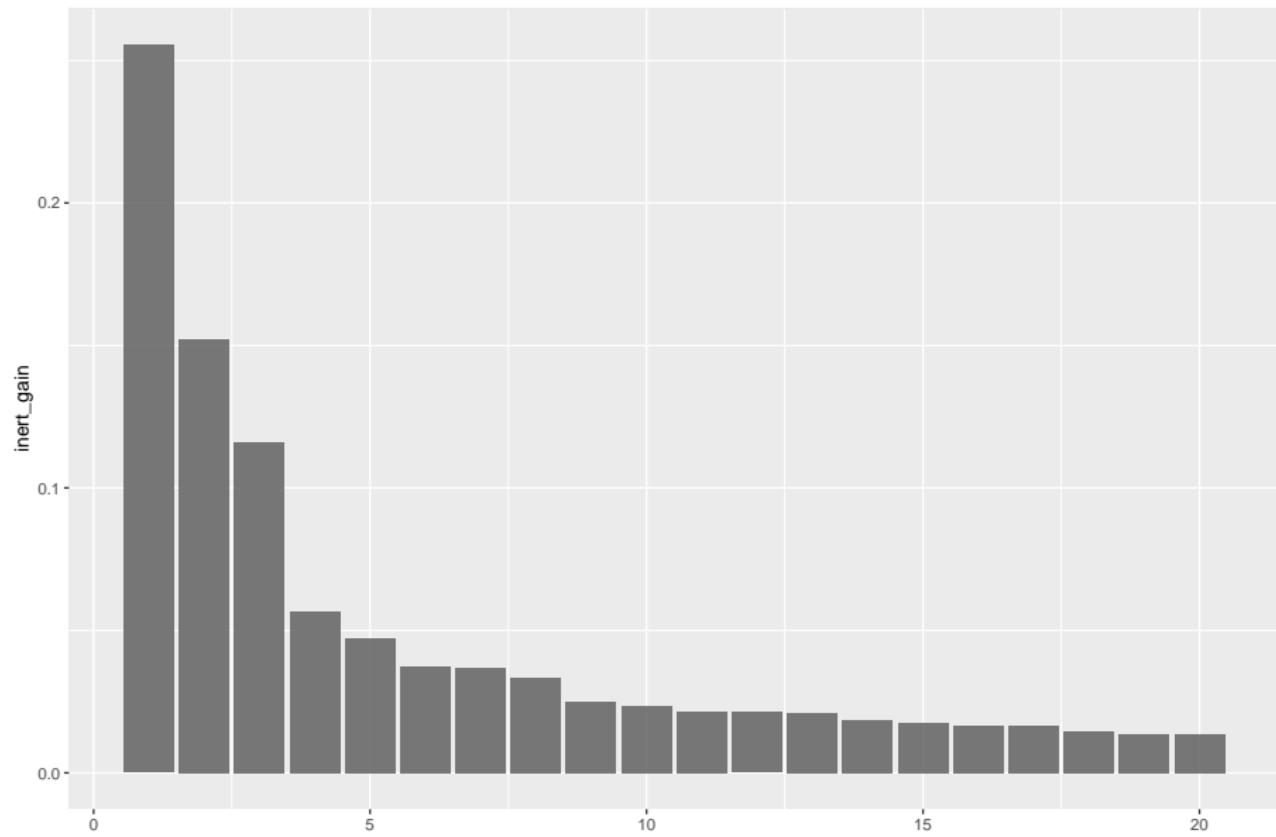
Arbre hiérarcique



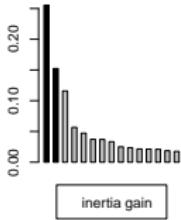
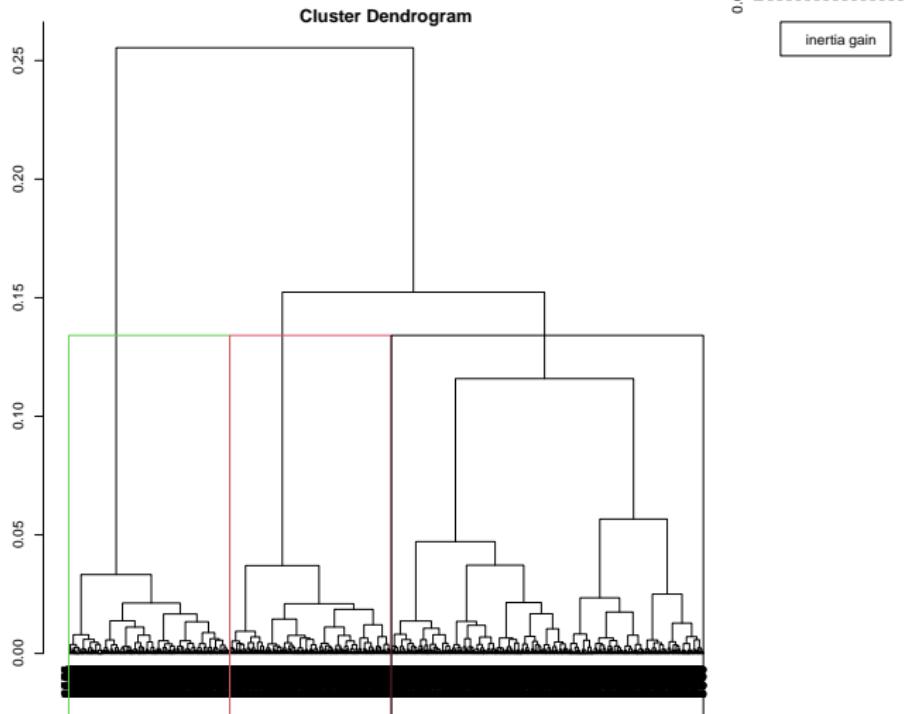
Optimal number of clusters



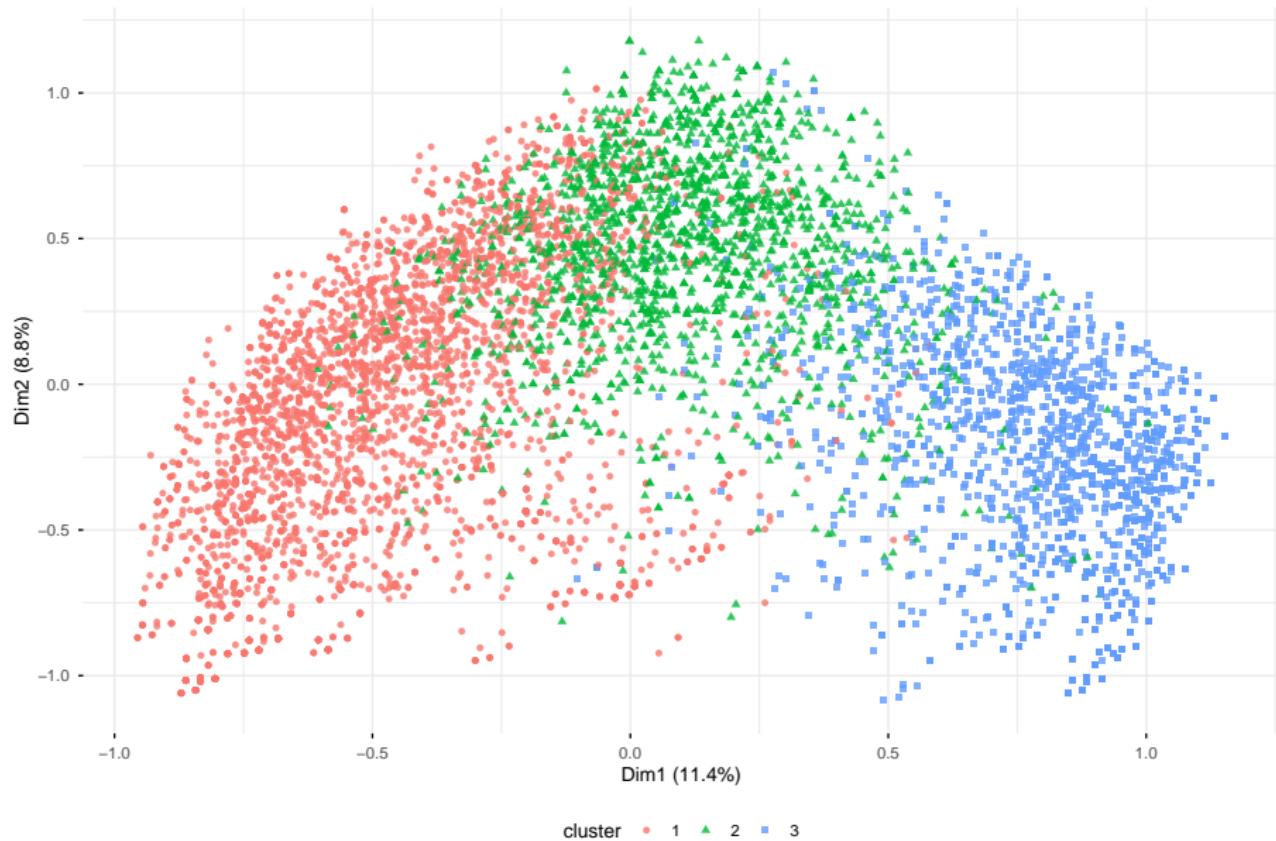
Between inertia gain



Arbre hiérarchique 2

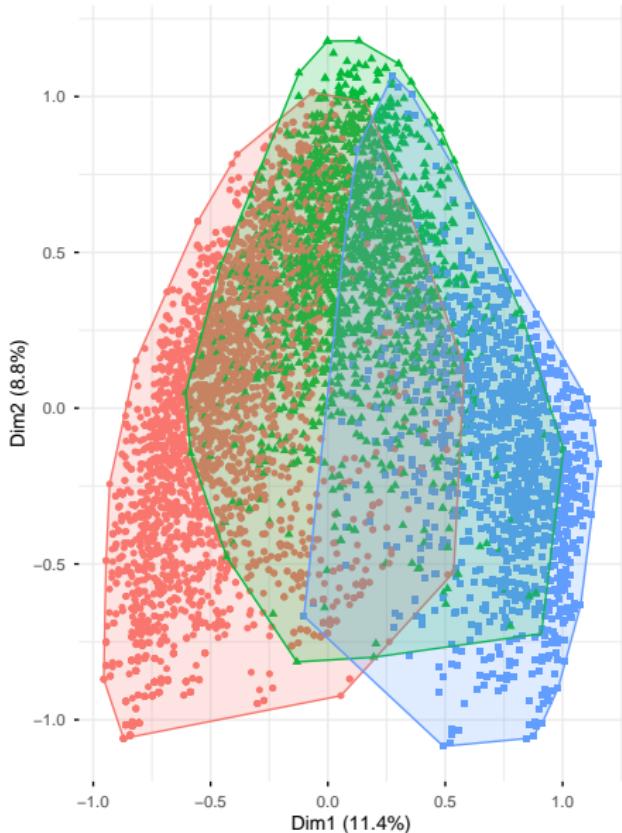


cluster visualization

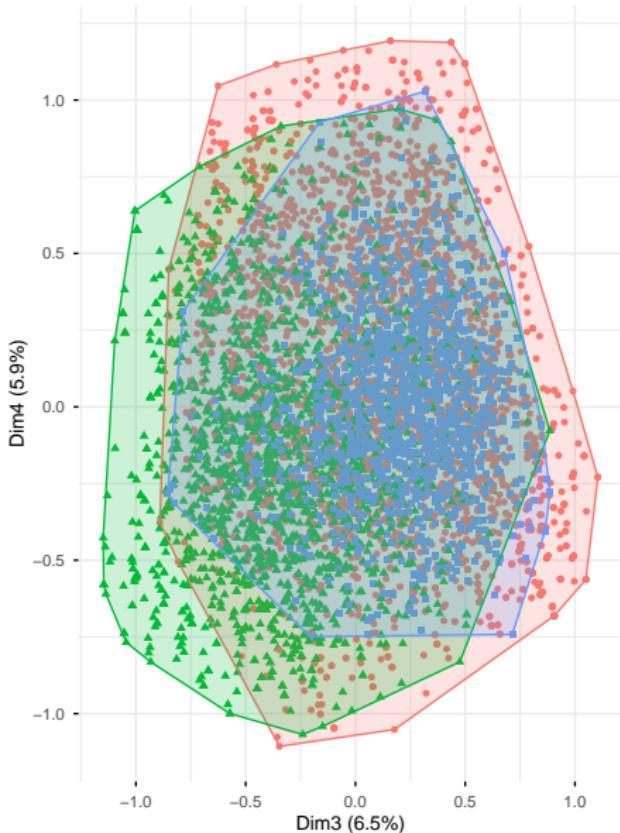


cluster ● 1 ▲ 2 ■ 3

Cluster visualization



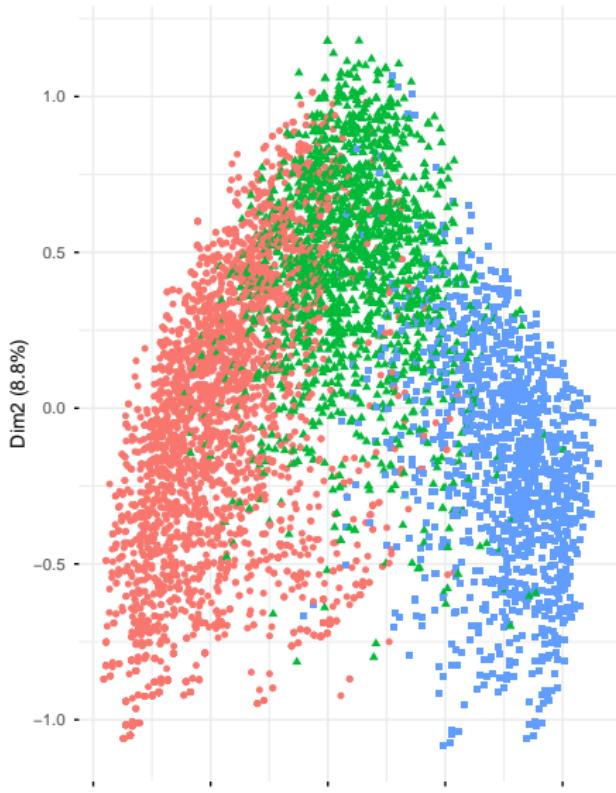
Cluster visualization



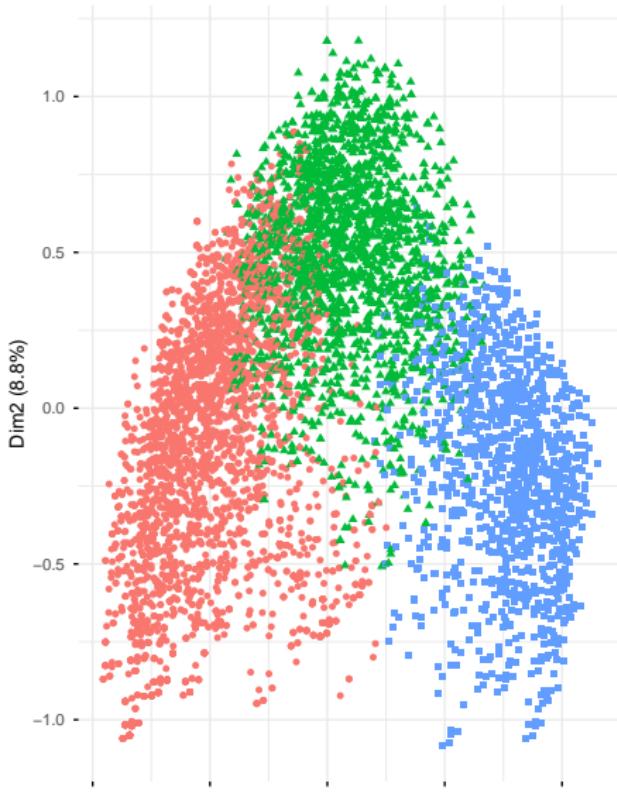
Avec consolidation

Dimension 1 et 2

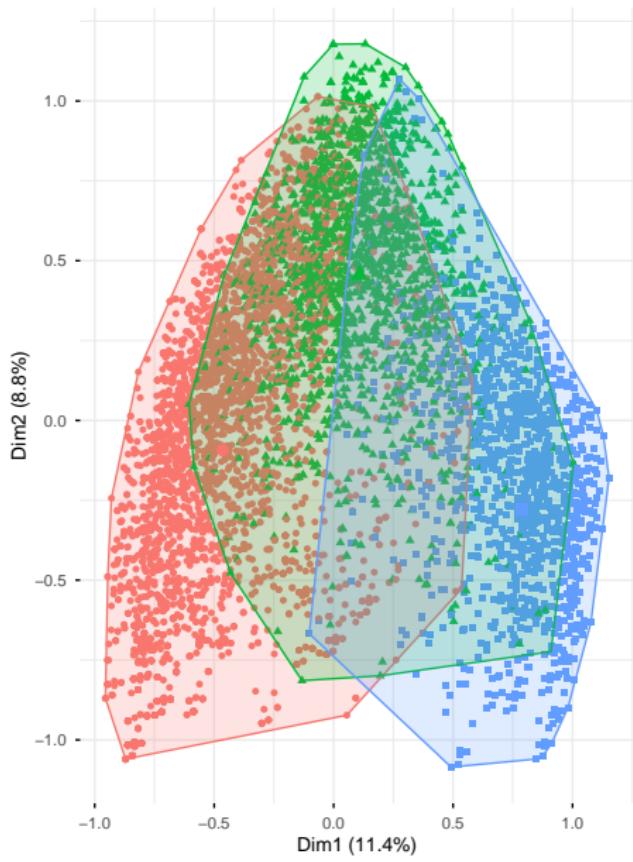
HC without consolidation



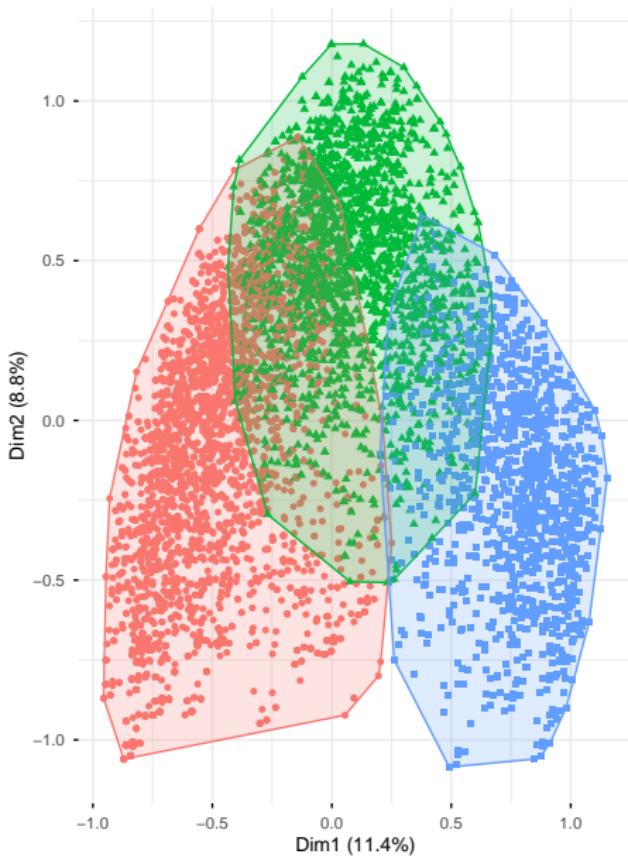
HC with consolidation



Visualisation des clusters sans consolidation

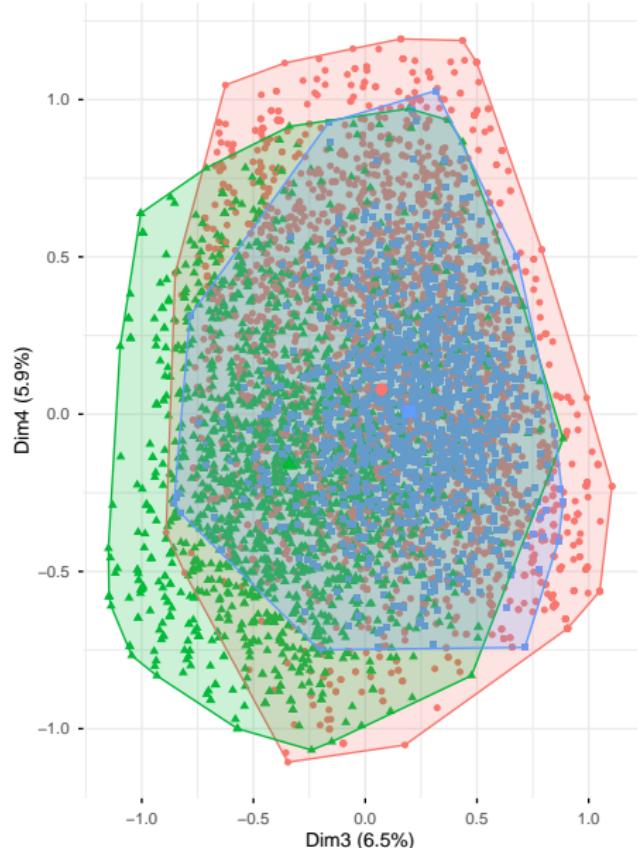


Visualisation des clusters avec consolidation

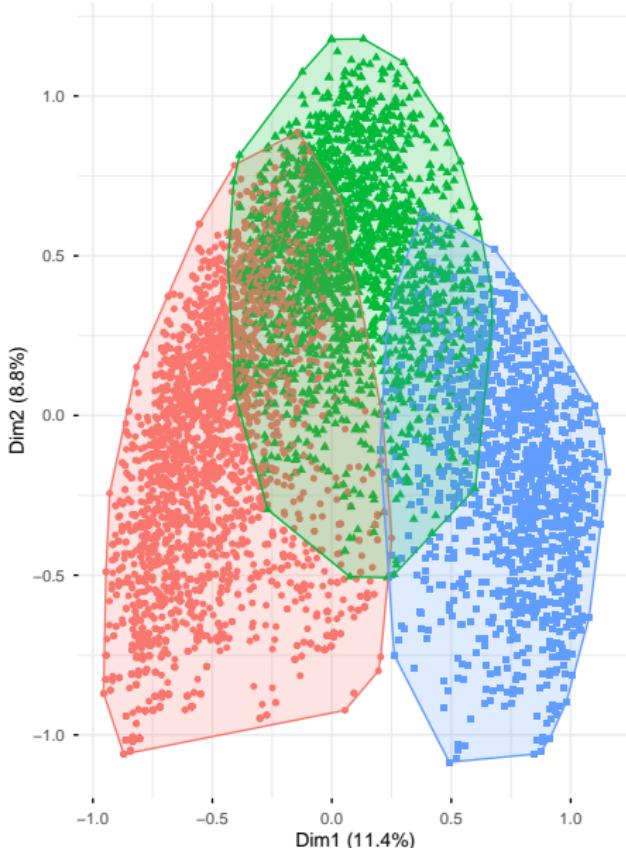


Dimension 3 et 4

Visualisation des clusters sans consolidation



Visualisation des clusters avec consolidation



	C.totale	C.moyenne/adr	C.moyenne/com	etiquette_dpe	etiquette_ges	annee_construction	surface_habitable_logement	conso_5_usages_ef	emission_ges_5_usages	cout_total_5_usages	nombre_appartement
Cluster 1											
2456	152.517	6.355	2.631	C	A	2012	32.5	2307.9	130.1	474.0	15
5505	48.794	4.436	2.631	D	B	1996	28.4	2899.4	207.0	562.0	11
7440	63.487	6.349	2.631	D	B	1989	29.0	3237.8	231.8	613.0	11
7841	60.750	4.050	2.631	C	A	1998	29.0	1982.7	143.1	533.8	15
1359	200.587	3.647	2.631	C	A	1989	27.5	2196.2	150.1	574.3	57
Cluster 2											
4522	25.529	1.276	2.619	D	D	1972	52.3	11105.0	1664.0	958.0	20
4661	5.536	0.503	2.555	E	E	1970	43.0	9870.1	2201.2	990.0	20
4690	14.933	1.358	2.403	E	E	1970	38.9	9639.7	2145.5	967.0	12
6230	14.020	1.078	2.555	D	D	1970	50.5	11134.4	2480.0	1116.0	19
1405	19.492	1.392	2.631	E	E	1975	50.8	11617.4	2587.4	1166.0	20
Cluster 3											
4441	22.837	2.076	2.239	D	D	1966	94.1	16187.1	3275.6	1931.0	11
3507	42.540	1.934	2.631	D	D	1949	248.7	53296.9	11943.1	5234.0	10
4282	44.955	2.498	2.631	C	C	1955	124.5	16938.0	2847.4	1633.0	19
7450	41.459	2.303	2.631	D	D	1957	96.8	16585.1	3070.5	1389.0	18
3246	21.280	1.637	2.631	E	E	1948	151.9	34957.6	8357.5	3368.0	18

	C.totale	C.moyenne/adr	C.moyenne/com	etiquette_dpe	etiquette_ges	annee_construction	surface_habitable_logement	conso_5_usages_ef	emission_ges_5_usages	cout_total_5_usages	nombre_appartement
Cluster 1											
3016	75.961	2.532	2.941	G	E	1920	7.3	3716.2	624.7	486.0	40
2113	228.565	2.540	2.941	C	A	2013	30.5	2180.5	154.3	478.0	96
4053	238.400	4.865	2.941	B	A	2013	86.0	2820.4	189.1	550.0	49
828	51.166	2.047	2.682	D	A	2021	37.8	3199.8	234.3	607.6	25
6454	106.146	4.083	2.941	C	A	1903	55.1	2826.8	192.4	551.0	26
Cluster 2											
2964	27.184	1.431	2.941	D	D	1967	60.1	10568.3	2342.7	1078.0	22
3894	23.902	1.138	2.941	D	D	1971	57.8	10138.8	2150.2	1182.0	49
3126	24.838	1.552	2.707	D	D	1959	52.0	10042.8	2239.6	1197.7	32
5047	60.288	1.827	2.941	D	D	1959	64.6	11325.3	2251.0	1020.0	33
5073	56.347	2.087	2.941	D	D	1956	73.5	12568.9	2487.2	1106.0	24
Cluster 3											
2980	89.811	2.994	2.941	D	D	1950	86.1	15824.2	3544.4	1556.0	28
5045	106.874	4.453	2.941	D	C	1967	202.9	22520.4	3811.9	3080.0	30
5051	35.408	3.541	2.941	D	D	1900	116.6	17272.5	3648.6	1439.0	31
7122	42.255	2.641	2.941	F	F	1930	89.2	29901.1	8313.9	2429.0	33
7146	39.035	2.602	2.941	D	D	1939	91.8	13643.5	3025.8	1465.0	26

	p.value	df
etiquette_ges	0.000000e+00	12
surface_habitable_logement	0.000000e+00	6
conso_5.usages_ef	0.000000e+00	6
emission_ges_5_usages	0.000000e+00	6
cout_total_5_usages	0.000000e+00	6
etiquette_dpe	0.000000e+00	12
annee_construction	0.000000e+00	6
Conso_totale	0.000000e+00	6
Conso_Moy_log	0.000000e+00	6
Conso_com	1.046000e-17	4
nombre_appartement	7.736142e-04	6

	Cla/Mod	Mod/Cla	Global	p.value
cout_total_5_usages=[112,679]	91.4	50.8	25.0	0.000000e+00
emission_ges_5_usages=(764,1.54e+03]	83.5	46.4	25.0	0.000000e+00
emission_ges_5_usages=[18.7,764]	91.8	50.9	25.0	0.000000e+00
conso_5.usages_ef=[405,5.57e+03]	99.9	55.4	25.0	0.000000e+00
surface_habitable_logement=[4.8,38.8]	79.0	43.9	25.0	0.000000e+00
conso_5.usages_ef=(5.57e+03,8.49e+03]	77.2	42.8	25.0	0.000000e+00
etiquette_ges=etiquette_ges_B	87.4	23.2	12.0	4.216257e-16
annee_construction=(1.98e+03,2.02e+03]	69.3	38.4	25.0	4.914476e-12
etiquette_dpe=etiquette_dpe_C	66.2	42.1	28.7	1.309104e-11
etiquette_ges=etiquette_ges_A	89.9	12.4	6.2	1.057490e-9
etiquette_ges=etiquette_ges_C	60.8	42.8	31.7	1.565536e-7
etiquette_dpe=etiquette_dpe_B	99.4	5.0	2.3	7.211058e-5
Conso_Moy_log=(2.52,11.9]	58.6	32.5	25.0	4.889148e-4
Conso_totale=(81.8,2e+03]	57.3	31.8	25.0	3.221480e-3
cout_total_5_usages=(679,946]	54.7	30.3	25.0	3.219775e-2
etiquette_dpe=etiquette_dpe_A	100.0	0.9	0.4	1.872444e-1
Conso_com=[2.09,2.63]	47.9	66.4	62.4	2.688319e-1
Conso_totale=(46.7,81.8]	49.0	27.2	25.0	9.494287e-0
nombre_appartement=(20,33]	48.5	23.4	21.7	2.221801e-0
annee_construction=[1.61e+03,1.94e+03]	47.9	26.7	25.1	5.236471e-0

	Cla/Mod	Mod/Cla	Global	p.value
emission_ges_5_usages=(1.54e+03,2.62e+03]	89.2	74.9	25.0	0.000000e+00
conso_5.usages_ef=(8.49e+03,1.28e+04]	94.1	79.0	25.0	0.000000e+00
cout_total_5_usages=(946,1.31e+03]	57.2	48.1	25.0	4.469718e-17
etiquette_ges=etiquette_ges_D	43.8	41.3	28.1	1.229669e-5
cout_total_5_usages=(679,946]	43.1	36.1	25.0	4.445709e-4
Conso_totale=[4.43,28.1]	40.3	33.8	25.0	2.490892e-2
etiquette_dpe=etiquette_dpe_D	37.9	43.8	34.5	5.537070e-2
surface_habitable_logement=(38.8,57]	38.9	33.2	25.4	2.509650e-2
annee_construction=(1.97e+03,1.98e+03]	38.3	31.5	24.5	1.158208e-1
Conso_Moy_log=[0.225,1.37]	36.4	30.6	25.1	3.479068e-1
annee_construction=(1.94e+03,1.97e+03]	35.7	30.4	25.4	3.382145e-1
Conso_Moy_log=(1.37,1.82]	35.0	29.4	25.0	3.143342e-0
surface_habitable_logement=(57,73.5]	34.7	28.6	24.6	2.441977e-0
Conso_com=(2.94,6.54]	34.9	25.8	22.0	6.369440e-0
Conso_com=(2.63,2.94]	33.8	17.6	15.6	1.767468e-0
nombre_appartement=(61,7.63e+03]	32.5	27.2	24.9	3.990285e-0
Conso_totale=(28.1,46.7]	31.9	26.8	25.0	2.293618e-0
Conso_Moy_log=(1.82,2.52]	31.8	26.6	24.9	3.335029e-0
nombre_appartement=(33,61]	27.4	22.2	24.2	1.077838e-0
surface_habitable_logement=(73.5,1.29e+04]	27.0	22.7	25.0	2.830218e-0

	Cla/Mod	Mod/Cla	Global	p.value	v.test
cout_total_5_usages=(1.31e+03,1.53e+05]	78.0	77.4	25.0	0.000000e+00	Inf
emission_ges_5_usages=(2.62e+03,4.78e+05]	87.6	87.0	25.0	0.000000e+00	Inf
conso_5.usages_ef=(1.28e+04,2.22e+06]	97.8	97.1	25.0	0.000000e+00	Inf
surface_habitable_logement=(73.5,1.29e+04]	53.3	52.9	25.0	0.000000e+00	30.27
etiquette_ges=etiquette_ges_E	54.5	33.1	15.3	2.266248e-115	22.83
etiquette_dpe=etiquette_dpe_E	43.0	37.8	22.2	4.357470e-71	17.83
etiquette_ges=etiquette_ges_F	58.3	12.3	5.3	4.403605e-45	14.09
annee_construction=(1.94e+03,1.97e+03]	35.5	35.8	25.4	2.241155e-30	11.45
etiquette_dpe=etiquette_dpe_F	45.0	15.1	8.5	1.757121e-28	11.07
etiquette_ges=etiquette_ges_D	34.0	37.9	28.1	8.763334e-26	10.50
etiquette_ges=etiquette_ges_G	62.2	3.4	1.4	7.838967e-15	7.77
annee_construction=[1.61e+03,1.94e+03]	30.5	30.4	25.1	3.572462e-09	5.90
surface_habitable_logement=(57,73.5]	29.8	29.1	24.6	3.442093e-07	5.10
Conso_com=(2.63,2.94]	30.8	19.0	15.6	4.017375e-06	4.61
etiquette_dpe=etiquette_dpe_G	37.9	5.4	3.6	4.904519e-06	4.57
etiquette_dpe=etiquette_dpe_D	27.6	37.8	34.5	6.180471e-04	3.42
annee_construction=(1.97e+03,1.98e+03]	27.5	26.8	24.5	1.024834e-02	2.57
Conso_totale=[4.43,28.1]	27.2	27.0	25.0	2.629231e-02	2.22
nombre_appartement=(20,33]	23.1	20.0	21.7	3.475434e-02	-2.11
etiquette_dpe=etiquette_dpe_A	0.0	0.0	0.4	2.913940e-04	-3.62
Conso_com=(2.94,6.54]	21.4	18.8	22.0	9.117785e-05	-3.91
Conso_totale=(81.8,2e+03]	21.1	20.9	25.0	2.875510e-06	-4.68
cout_total_5_usages=(946,1.31e+03]	20.6	20.4	25.0	1.421675e-07	-5.26
etiquette_dpe=etiquette_dpe_B	0.0	0.0	2.3	1.544377e-21	-9.53