Projet de Statistiques - Nutrition et consommation d'alcool

PUISSANT Claire, DERACHE Gabin, LACOSTE Louis

28/11/2021

Contents

Résumé des données

Nous avons importé et préparé les données afin de pouvoir appliquer des analyses statistiques dessus (conversion en variable catégorielle de données malinterprétées) et également pour faciliter la lecture et la compréhension (changement de noms des colonnes pour des noms plus explicites).

- Données de INCA2

```
    nomen         sexe_ps        v2_age            bmi
110006 :   1    homme:1087    Min.   :18.00    Min.    :15.24
110007 :   1    femme:1537    1st Qu.:34.00    1st Qu.:21.67
110020 :   1                  Median :45.00    Median :24.09
110021 :   1                  Mean   :45.77    Mean    :24.81
110025 :   1                  3rd Qu.:57.00    3rd Qu.:27.17
110034 :   1                  Max.   :79.00    Max.    :56.77
(Other):2618                                   NA's    :29
          reg          q_tot            q_vin            q_biere
Ouest         :398  Min.   :    0.0  Min.   :    0.0  Min.   :     0.0
Nord Ouest    :395  1st Qu.:    0.0  1st Qu.:    0.0  1st Qu.:     0.0
Ile de France:353   Median :  300.0  Median :  120.0  Median :     0.0
Sud Est       :328  Mean   :  839.6  Mean   :  530.9  Mean   :   173.8
Centre Est    :310  3rd Qu.: 1140.0  3rd Qu.:  700.0  3rd Qu.:     0.0
Est           :293  Max.   :10660.0  Max.   : 8160.0  Max.   :10500.0
(Other)       :547
    q_cidre          q_spirit          q_cocktail        q_lundi
Min.   :   0.00  Min.   :    0.00  Min.   :    0.00  Min.   :    0.00
1st Qu.:   0.00  1st Qu.:    0.00  1st Qu.:    0.00  1st Qu.:    0.00
Median :   0.00  Median :    0.00  Median :    0.00  Median :    0.00
Mean   :  24.62  Mean   :   37.84  Mean   :   72.48  Mean   :   84.84
3rd Qu.:   0.00  3rd Qu.:   40.00  3rd Qu.:    0.00  3rd Qu.:  100.00
Max.   :3240.00  Max.   : 1510.00  Max.   : 4200.00  Max.   : 2210.00

    q_mardi          q_mercredi         q_jeudi          q_vendredi
```

```
   Min.   :   0.00    Min.   :   0.00    Min.   :   0.0    Min.   :   0.0
   1st Qu.:   0.00    1st Qu.:   0.00    1st Qu.:   0.0    1st Qu.:   0.0
   Median :   0.00    Median :   0.00    Median :   0.0    Median :   0.0
   Mean   :  90.04    Mean   :  91.84    Mean   :  97.4    Mean   : 123.3
   3rd Qu.: 100.00    3rd Qu.: 100.00    3rd Qu.: 100.0    3rd Qu.: 150.0
   Max.   :3650.00    Max.   :2500.00    Max.   :1900.0    Max.   :2600.0

     q_samedi          q_dimanche       energie_totale      proteines
   Min.   :   0.0    Min.   :   0.0    Min.   : 255.9    Min.   : 12.97
   1st Qu.:   0.0    1st Qu.:   0.0    1st Qu.:1509.1    1st Qu.: 61.83
   Median :   0.0    Median :   0.0    Median :1885.4    Median : 76.32
   Mean   : 189.9    Mean   : 162.3    Mean   :1945.0    Mean   : 79.12
   3rd Qu.: 250.0    3rd Qu.: 240.0    3rd Qu.:2316.1    3rd Qu.: 93.53
   Max.   :2860.0    Max.   :3950.0    Max.   :5209.7    Max.   :229.51

   glucides_dispos     lipides            alcool          ag_monoinsatures
   Min.   : 15.7    Min.   :  8.938    Min.   :  0.00000    Min.   : 2.975
   1st Qu.:152.4    1st Qu.: 62.178    1st Qu.:  0.00855    1st Qu.:21.492
   Median :198.9    Median : 78.500    Median :  3.91976    Median :27.912
   Mean   :206.0    Mean   : 80.846    Mean   : 10.26428    Mean   :28.903
   3rd Qu.:248.4    3rd Qu.: 96.692    3rd Qu.: 14.04332    3rd Qu.:35.189
   Max.   :752.1    Max.   :243.495    Max.   :129.47290    Max.   :83.593

   ag_polyinsatures    ag_satures          amidon          glucides_simples
   Min.   : 0.7379    Min.   :  2.648    Min.   :  9.344    Min.   :  2.074
   1st Qu.: 8.0069    1st Qu.: 24.181    1st Qu.: 80.494    1st Qu.: 59.316
   Median :10.9785    Median : 31.526    Median :108.121    Median : 80.838
   Mean   :11.9263    Mean   : 33.150    Mean   :114.876    Mean   : 85.774
   3rd Qu.:14.5498    3rd Qu.: 40.353    3rd Qu.:139.859    3rd Qu.:107.026
   Max.   :50.1500    Max.   :110.004    Max.   :685.843    Max.   :358.841

      fibres           cholesterol          eau              calcium
   Min.   : 1.681    Min.   : 26.11    Min.   :  355.3    Min.   : 103.3
   1st Qu.:11.559    1st Qu.:223.57    1st Qu.: 1633.0    1st Qu.: 620.6
   Median :15.069    Median :298.86    Median : 2088.3    Median : 812.8
   Mean   :15.891    Mean   :313.00    Mean   : 2152.2    Mean   : 846.2
   3rd Qu.:19.355    3rd Qu.:383.09    3rd Qu.: 2574.5    3rd Qu.:1033.1
   Max.   :60.906    Max.   :944.48    Max.   :12601.9    Max.   :2785.5

       fer              sodium           magnesium          manganese
   Min.   : 1.862    Min.   :  508.3    Min.   :  51.64    Min.   : 0.2157
   1st Qu.: 8.712    1st Qu.: 1987.8    1st Qu.: 207.12    1st Qu.: 1.8298
   Median :11.234    Median : 2561.7    Median : 254.37    Median : 2.4261
   Mean   :11.992    Mean   : 2696.2    Mean   : 267.34    Mean   : 2.6595
   3rd Qu.:14.298    3rd Qu.: 3250.2    3rd Qu.: 314.74    3rd Qu.: 3.2325
   Max.   :61.158    Max.   :12378.0    Max.   :1393.87    Max.   :13.7595

     phosphore         potassium           cuivre             zinc
   Min.   : 204.1    Min.   : 493.6    Min.   :0.2176    Min.   : 1.512
   1st Qu.: 916.3    1st Qu.:2127.1    1st Qu.:0.9089    1st Qu.: 7.444
   Median :1123.3    Median :2620.6    Median :1.1714    Median : 9.391
   Mean   :1158.3    Mean   :2704.9    Mean   :1.3329    Mean   : 9.810
   3rd Qu.:1359.1    3rd Qu.:3203.0    3rd Qu.:1.5188    3rd Qu.:11.606
   Max.   :3174.1    Max.   :6965.4    Max.   :8.9918    Max.   :29.361
```

```
    selenium              iode            retinol          beta-carotene
 Min.   :  8.099   Min.   : 22.69   Min.   :   5.676   Min.   :   99.16
 1st Qu.: 36.809   1st Qu.: 84.58   1st Qu.: 252.156   1st Qu.: 1536.03
 Median : 46.541   Median :109.04   Median : 376.715   Median : 2654.54
 Mean   : 48.419   Mean   :116.08   Mean   : 634.582   Mean   : 3047.82
 3rd Qu.: 57.606   3rd Qu.:137.66   3rd Qu.: 627.585   3rd Qu.: 4037.05
 Max.   :135.609   Max.   :558.59   Max.   :8006.581   Max.   :25886.94

    vitamineC          vitamineD          vitamineE         thiamine-B1
 Min.   :  4.162   Min.   : 0.09571   Min.   : 1.072   Min.   : 0.1755
 1st Qu.: 48.631   1st Qu.: 1.31032   1st Qu.: 6.611   1st Qu.: 0.8425
 Median : 75.874   Median : 1.94360   Median : 9.316   Median : 1.0572
 Mean   : 85.506   Mean   : 2.37479   Mean   :10.196   Mean   : 1.1280
 3rd Qu.:110.599   3rd Qu.: 2.98656   3rd Qu.:12.486   3rd Qu.: 1.3283
 Max.   :381.950   Max.   :44.83060   Max.   :45.383   Max.   :10.6242

 riboflavine-B2     niacine-B3      acide_panthothenique-B5 pyridoxine-B6
 Min.   :0.2414   Min.   : 3.039   Min.   : 0.9443          Min.   :0.2738
 1st Qu.:1.3060   1st Qu.:12.708   1st Qu.: 3.9473          1st Qu.:1.1800
 Median :1.6637   Median :16.027   Median : 4.9695          Median :1.5011
 Mean   :1.7294   Mean   :17.087   Mean   : 5.1711          Mean   :1.5849
 3rd Qu.:2.0595   3rd Qu.:20.526   3rd Qu.: 6.1376          3rd Qu.:1.8815
 Max.   :5.7477   Max.   :80.164   Max.   :20.1084          Max.   :6.3944

   folates-B9      colobamine-B12
 Min.   :  34.34   Min.   : 0.1012
 1st Qu.: 192.04   1st Qu.: 3.0478
 Median : 251.39   Median : 4.2319
 Mean   : 261.85   Mean   : 5.3231
 3rd Qu.: 313.78   3rd Qu.: 5.8607
 Max.   :1090.67   Max.   :37.4638
```

- Données de consommation associées

```
    nomen            nomjour            nomgr           qte_nette
 113036 :  21   dimanche:1657   biere  :1073   Min.   :   1.0
 170467 :  20   jeudi   : 986   cidre  : 218   1st Qu.: 100.0
 115284 :  18   lundi   : 891   cocktail: 906  Median : 200.0
 171011 :  18   mardi   : 928   spirit :1219   Mean   : 264.3
 112304 :  17   mercredi: 935   vin    :4919   3rd Qu.: 330.0
 110889 :  16   samedi  :1723                  Max.   :3750.0
 (Other):8225   vendredi:1215
```
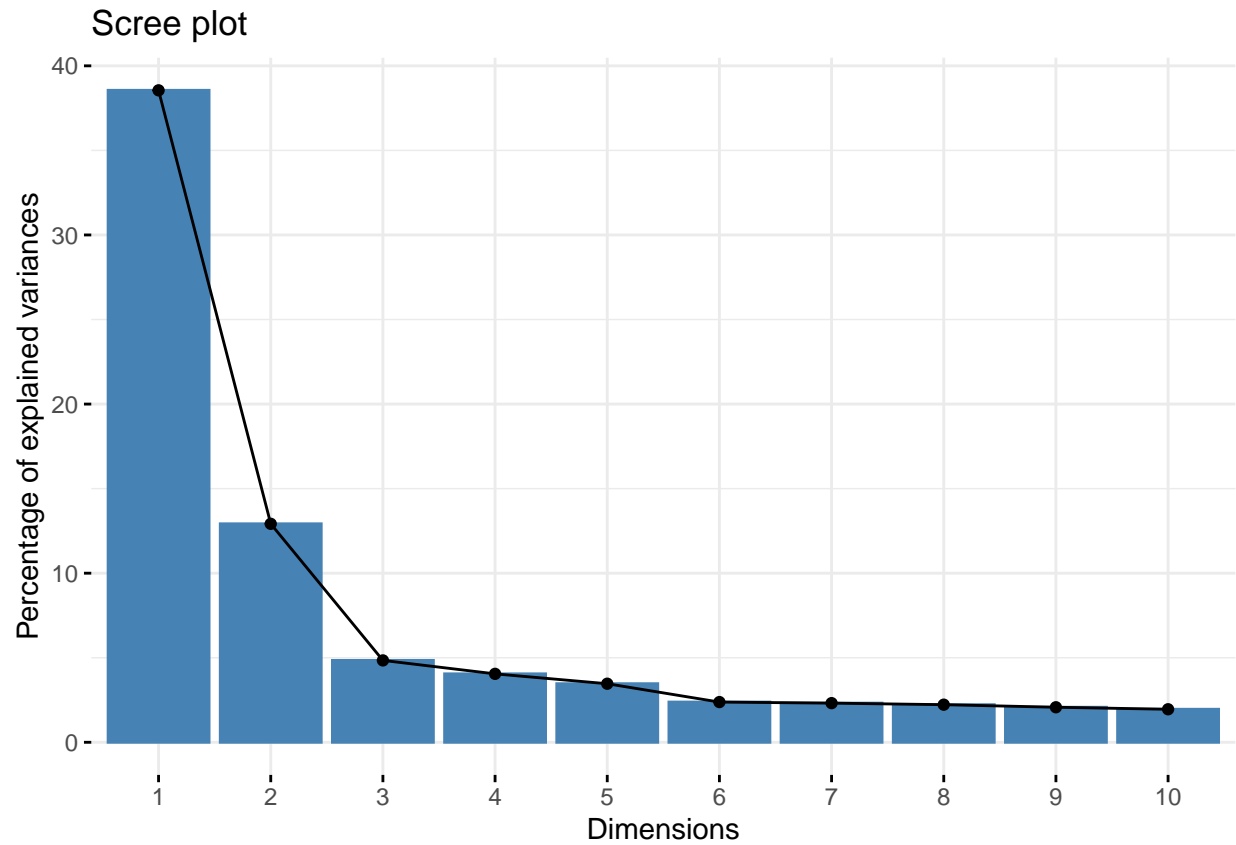
# ACP sur les données de INCA2

ACP sur toutes les données
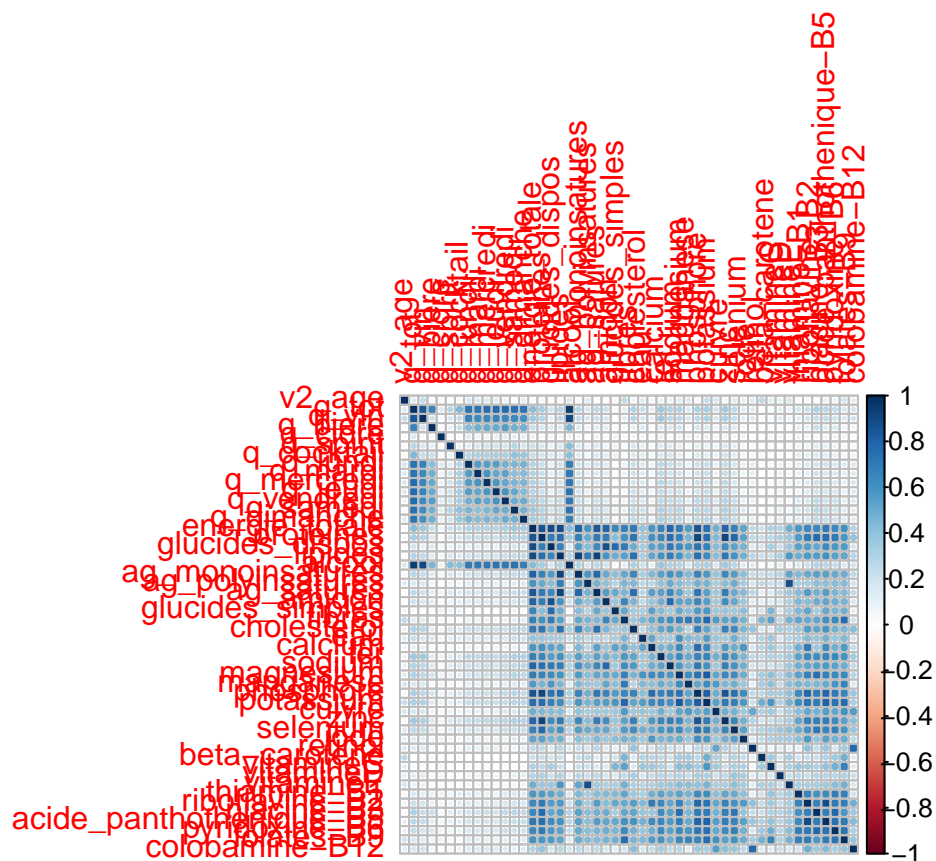
	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	19.275727	38.551454	38.55145
Dim.2	6.459741	12.919482	51.47094
Dim.3	2.421394	4.842788	56.31372
Dim.4	2.023478	4.046956	60.36068
Dim.5	1.731966	3.463932	63.82461
Dim.6	1.189660	2.379320	66.20393

4

Scree plot

ACP sur les sous catégories

On repère sur le graphique des corrélations plusieurs groupes qui semblent être faiblement corrélés en dehors des groupes, nous allons donc les étudier en les séparant.

Sous bloc 1 : Par semaine
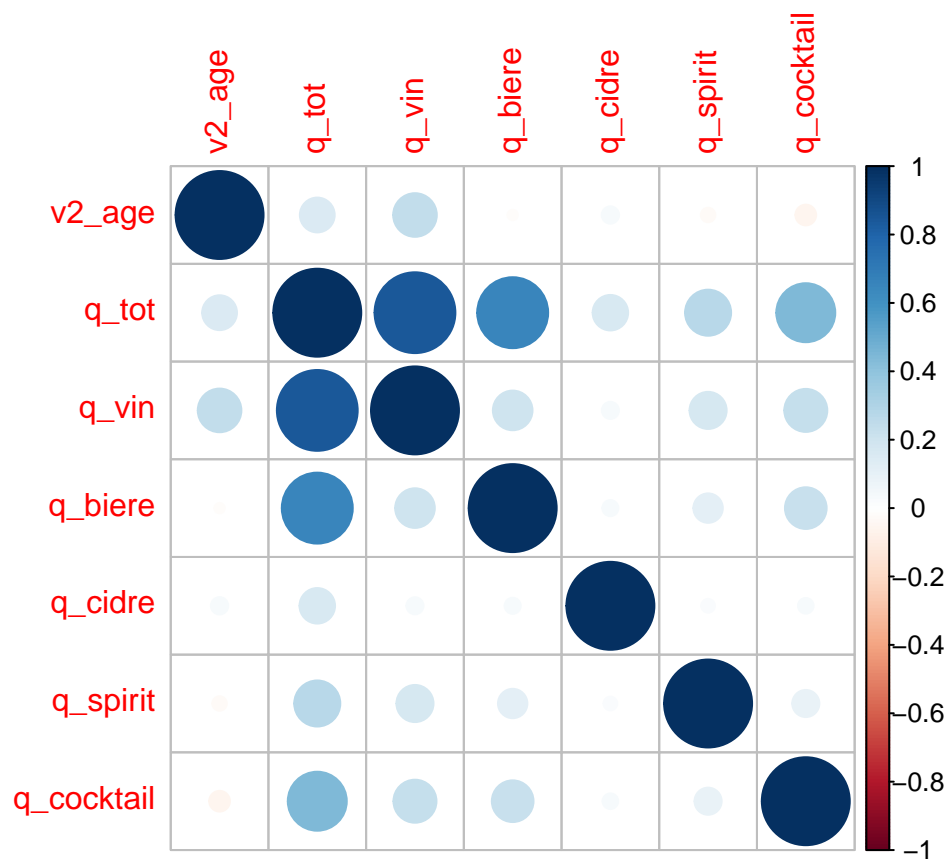
	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.5093727	35.848181	35.84818
Dim.2	1.1341915	16.202736	52.05092
Dim.3	0.9886685	14.123836	66.17475
Dim.4	0.9346412	13.352018	79.52677
Dim.5	0.7909750	11.299643	90.82641
Dim.6	0.6421510	9.173586	100.00000
Dim.7	0.0000000	0.000000	100.00000

Scree plot

Variables – PCA

Variables – PCA

Variables – PCA
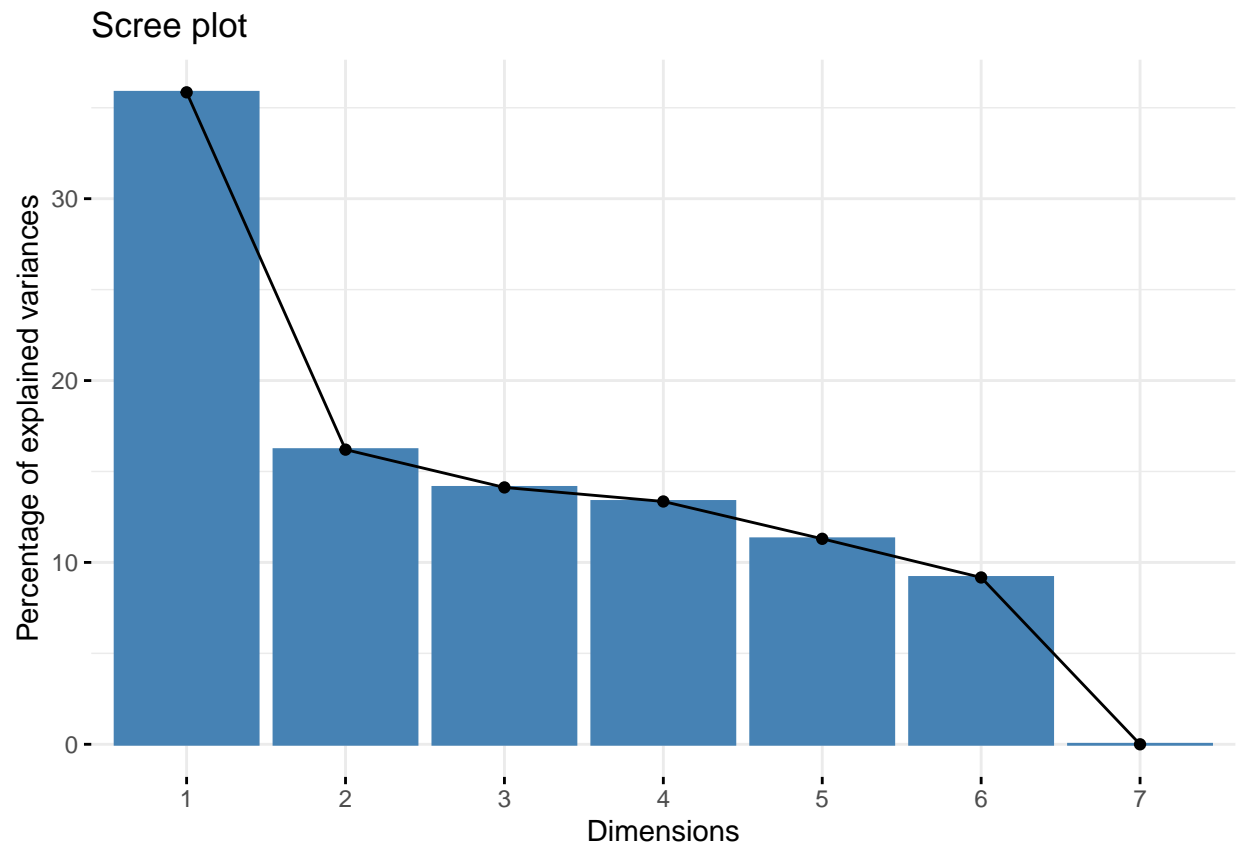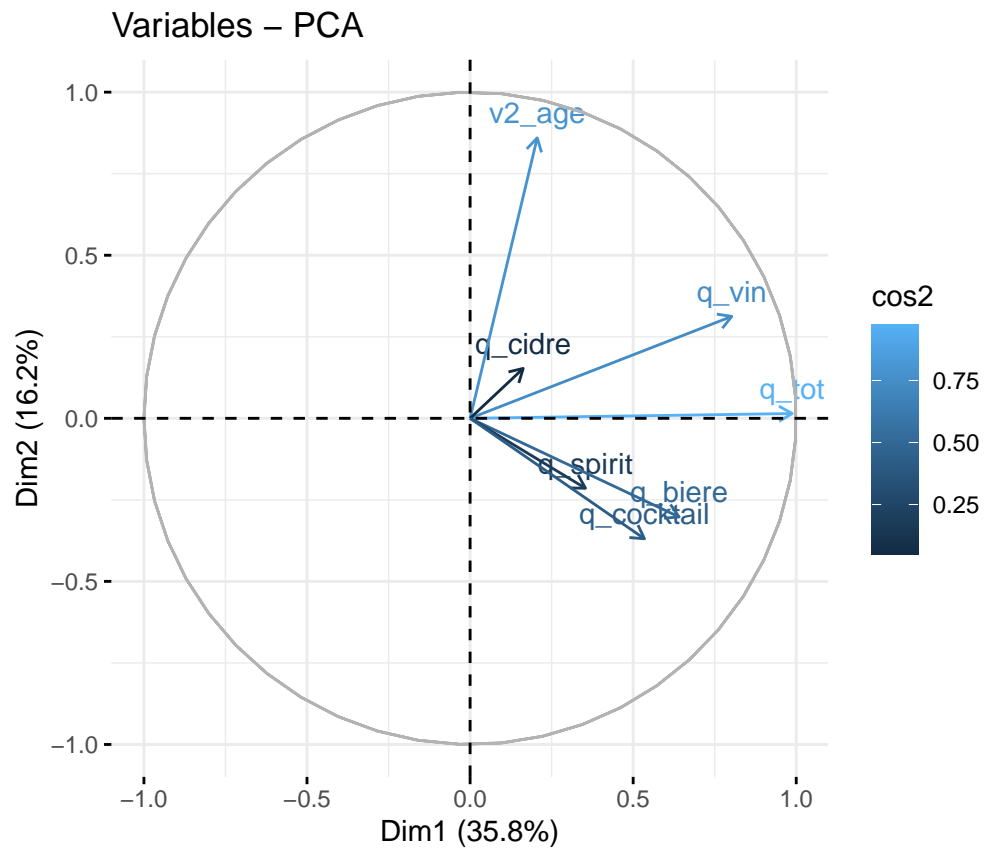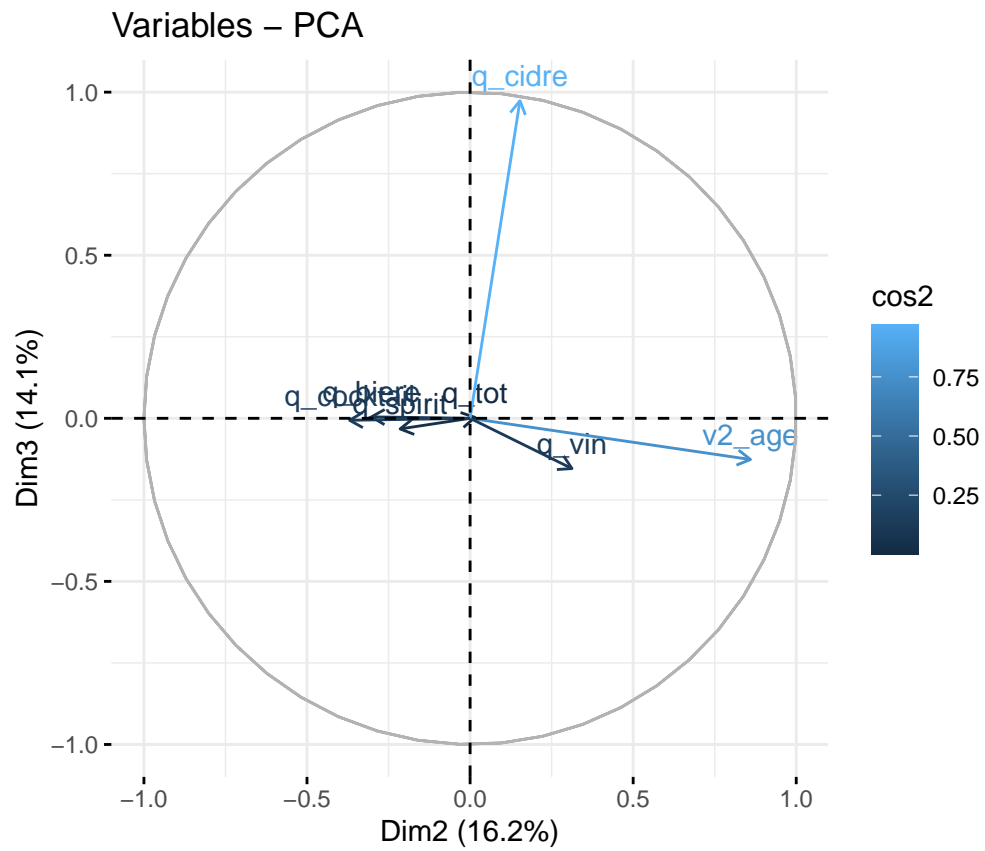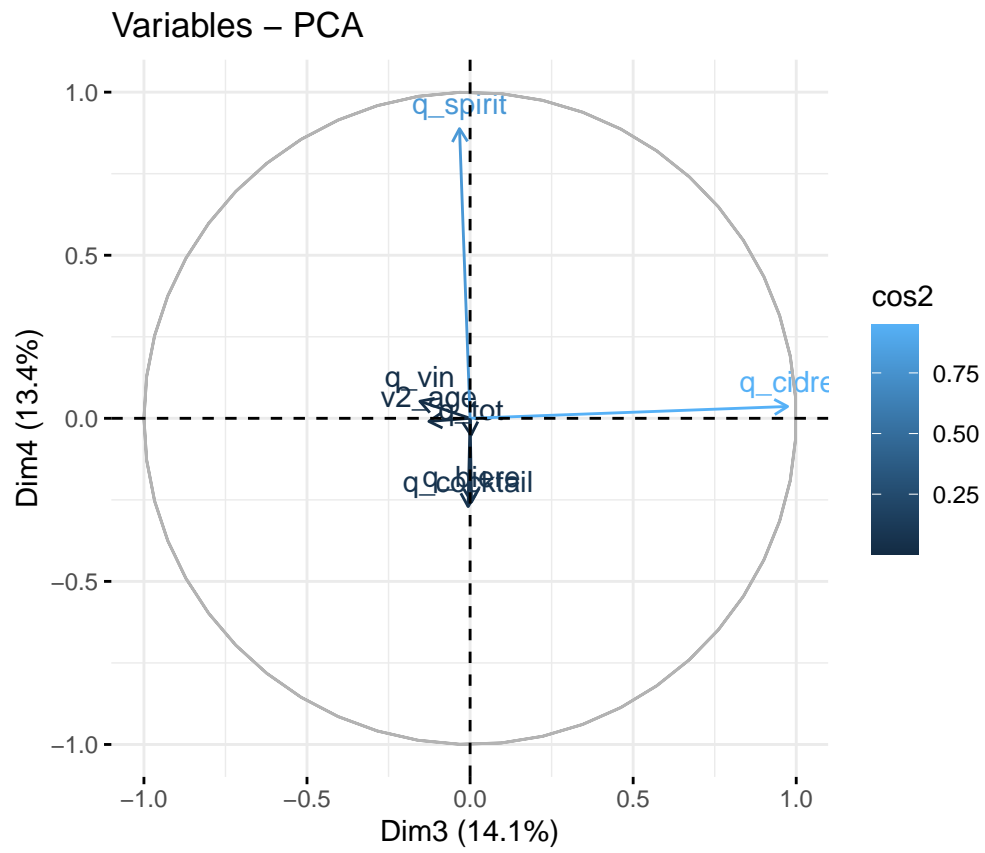
Sous bloc 2 : Les jours de la semaine
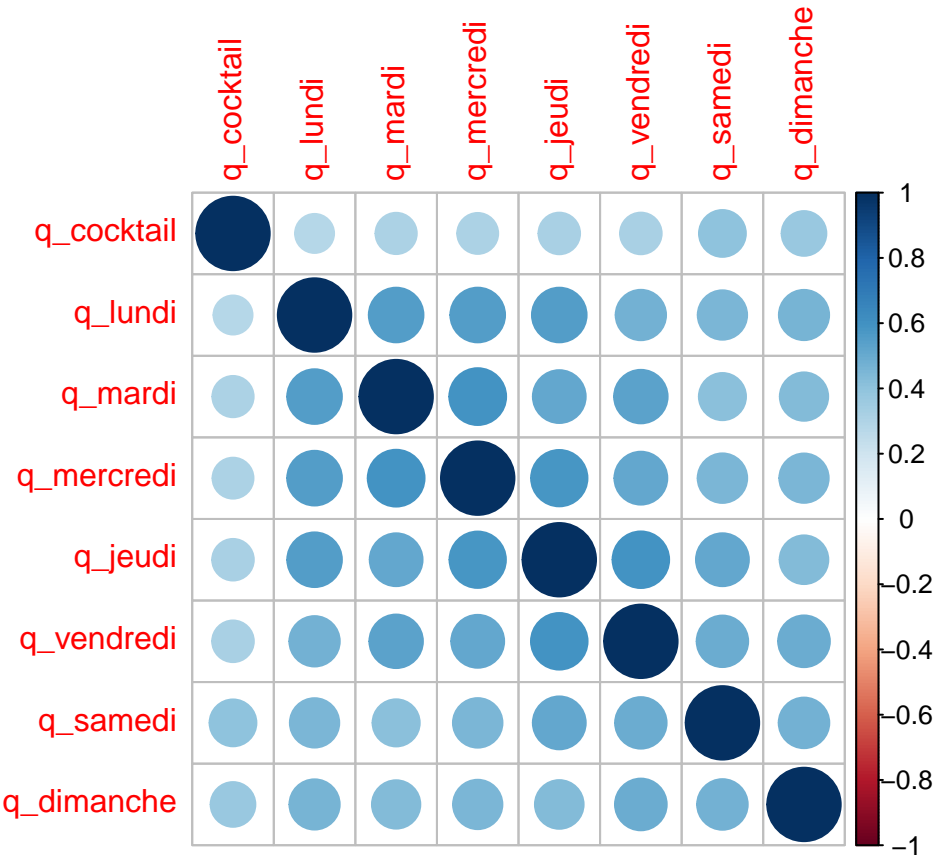
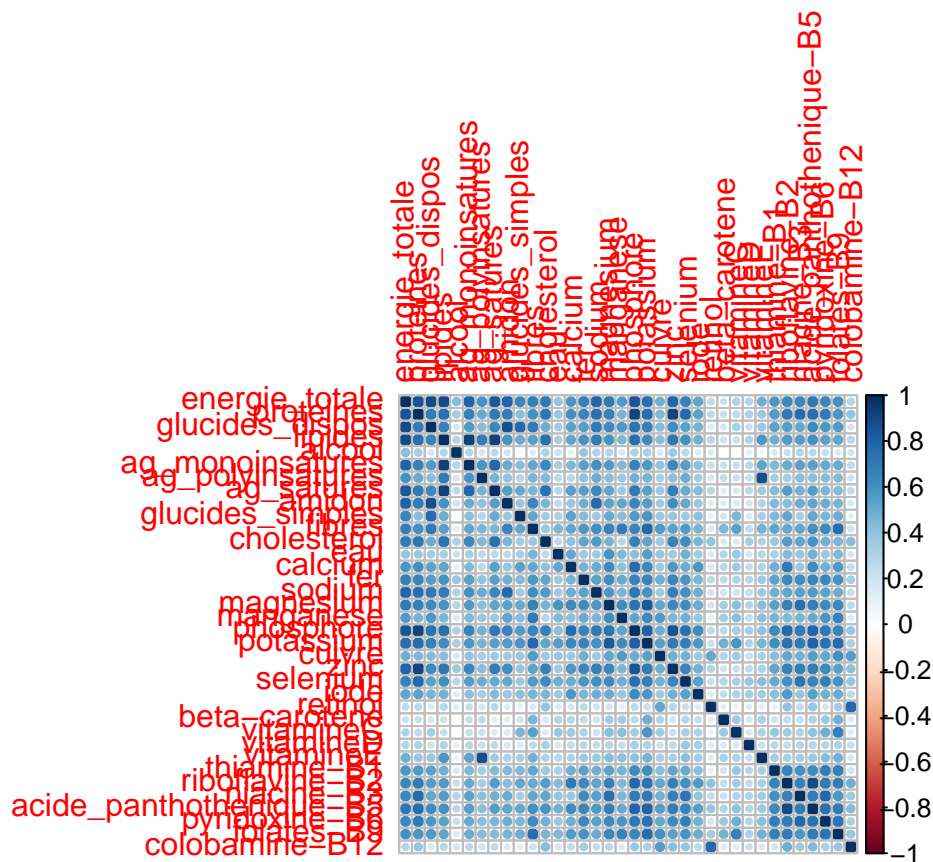|	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	4.2612399	53.265498	53.26550
Dim.2	0.8542927	10.678659	63.94416
Dim.3	0.6049862	7.562327	71.50648
Dim.4	0.5643232	7.054040	78.56052
Dim.5	0.5164864	6.456079	85.01660
Dim.6	0.4380387	5.475484	90.49209
Dim.7	0.4201987	5.252484	95.74457
Dim.8	0.3404343	4.255429	100.00000

Scree plot

Variables – PCA

Variables – PCA

Variables – PCA

q_dimanche

q_lundi
q_mardi
q_mercredi
q_cocktail
q_vendredi
q_samedi
q_jeudi

Dim3 (7.6%)

Dim4 (7.1%)

cos2
0.4
0.3
0.2
0.1
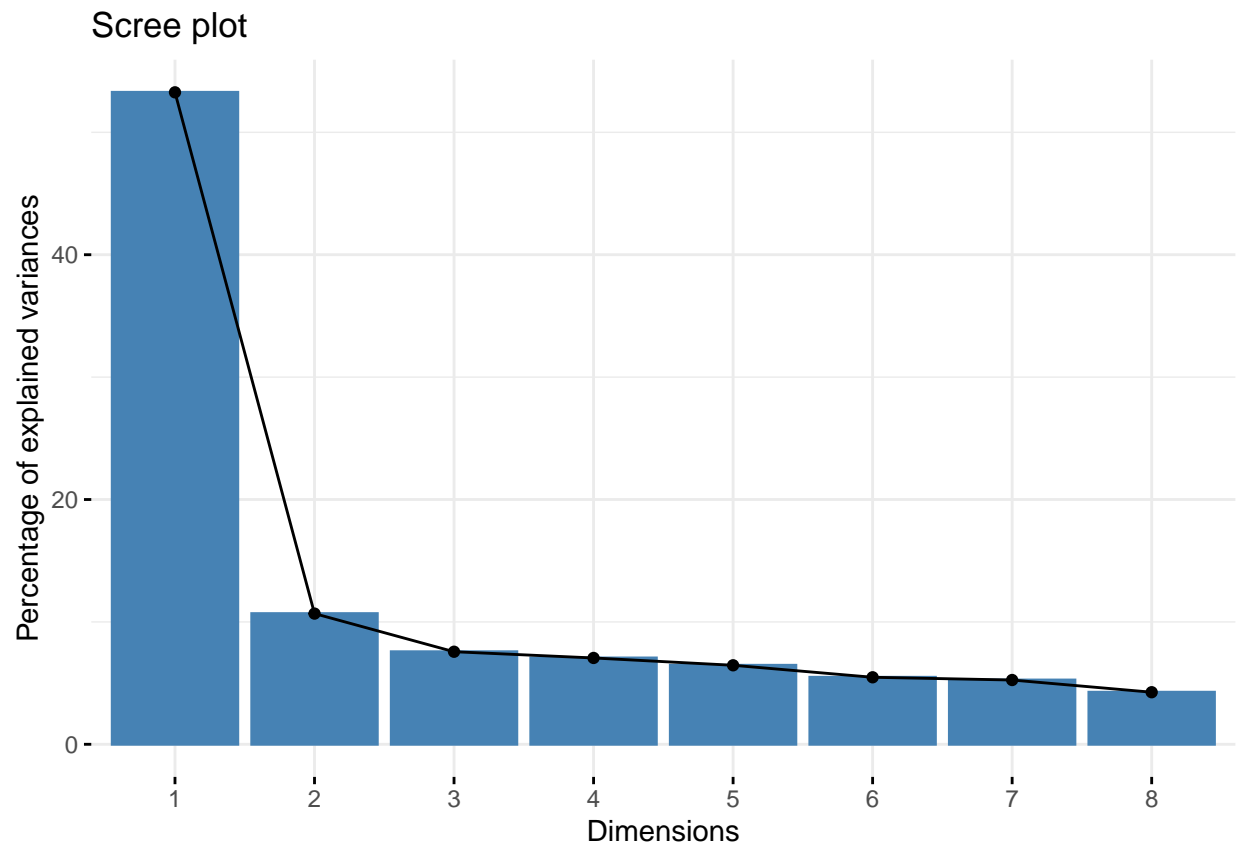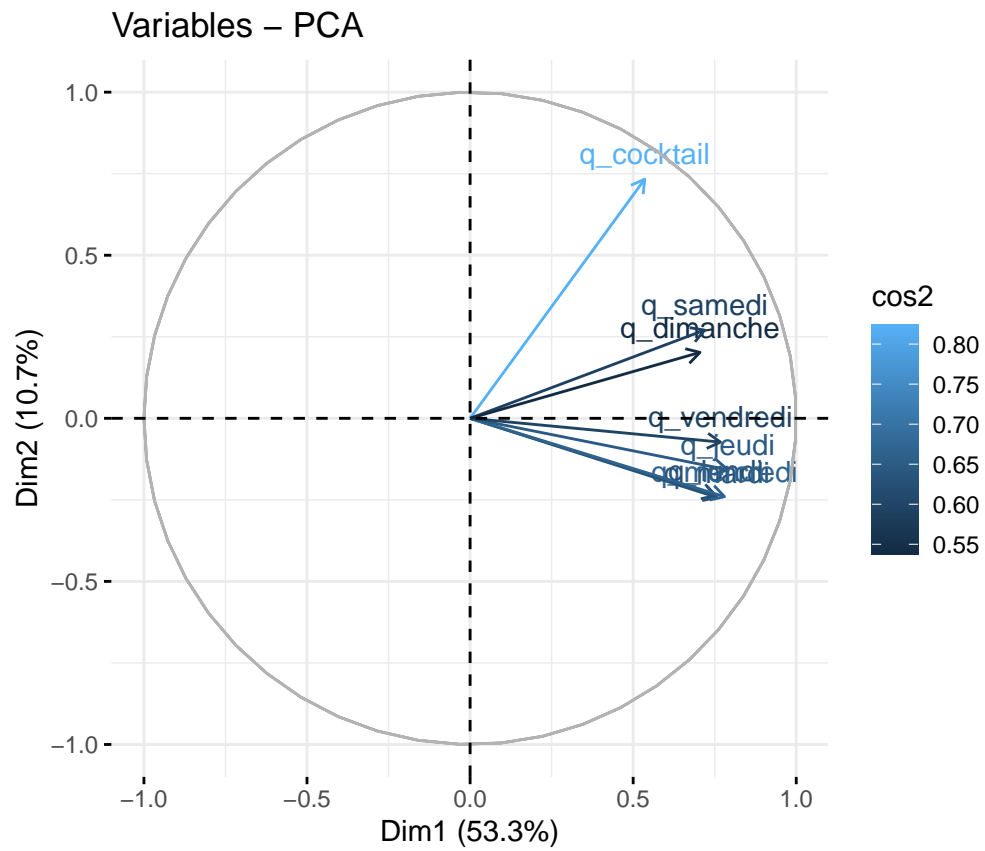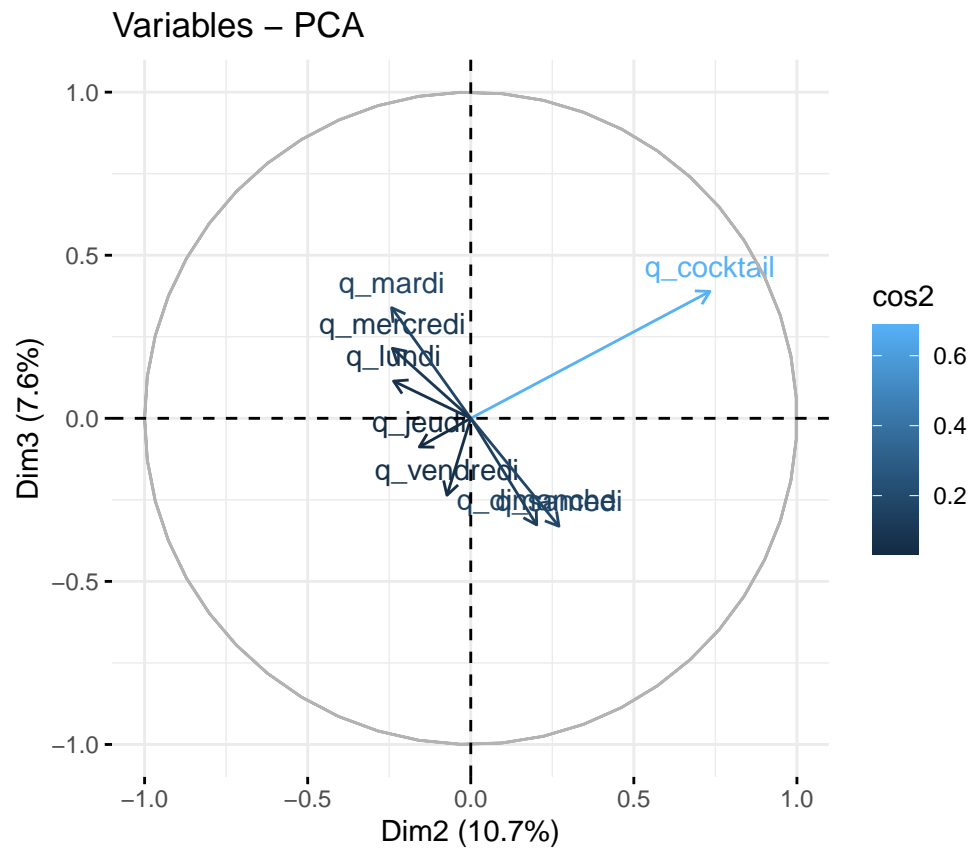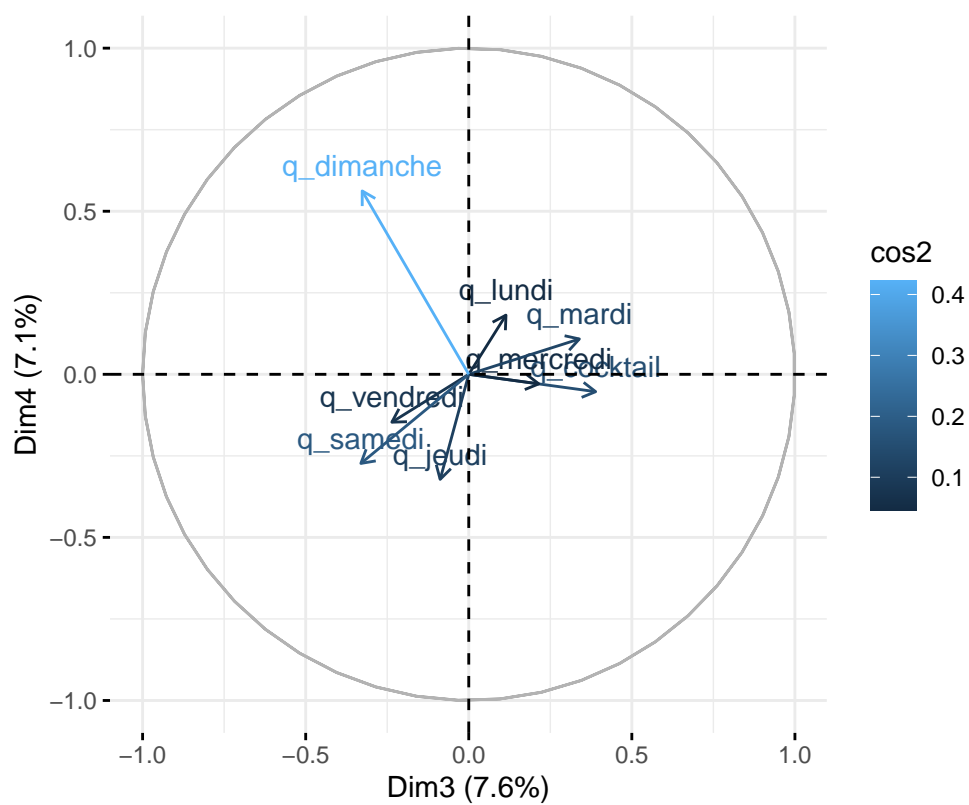
Sous bloc 3 : Par jour

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	18.1856505	50.5156957	50.51570
Dim.2	2.4154407	6.7095576	57.22525
Dim.3	2.0620554	5.7279317	62.95318
Dim.4	1.5984793	4.4402202	67.39341
Dim.5	1.2425182	3.4514395	70.84484
Dim.6	1.1431351	3.1753753	74.02022
Dim.7	0.9721495	2.7004153	76.72064
Dim.8	0.9160863	2.5446842	79.26532
Dim.9	0.8181952	2.2727644	81.53808
Dim.10	0.7523230	2.0897861	83.62787
Dim.11	0.6295460	1.7487389	85.37661
Dim.12	0.5836158	1.6211551	86.99776
Dim.13	0.5332328	1.4812021	88.47897
Dim.14	0.5067351	1.4075974	89.88656
Dim.15	0.4314593	1.1984980	91.08506
Dim.16	0.3869920	1.0749778	92.16004
Dim.17	0.3549910	0.9860862	93.14613
Dim.18	0.3347196	0.9297766	94.07590
Dim.19	0.2744095	0.7622487	94.83815
Dim.20	0.2527190	0.7019973	95.54015
Dim.21	0.2371643	0.6587896	96.19894
Dim.22	0.2243979	0.6233275	96.82227
Dim.23	0.1858817	0.5163381	97.33860

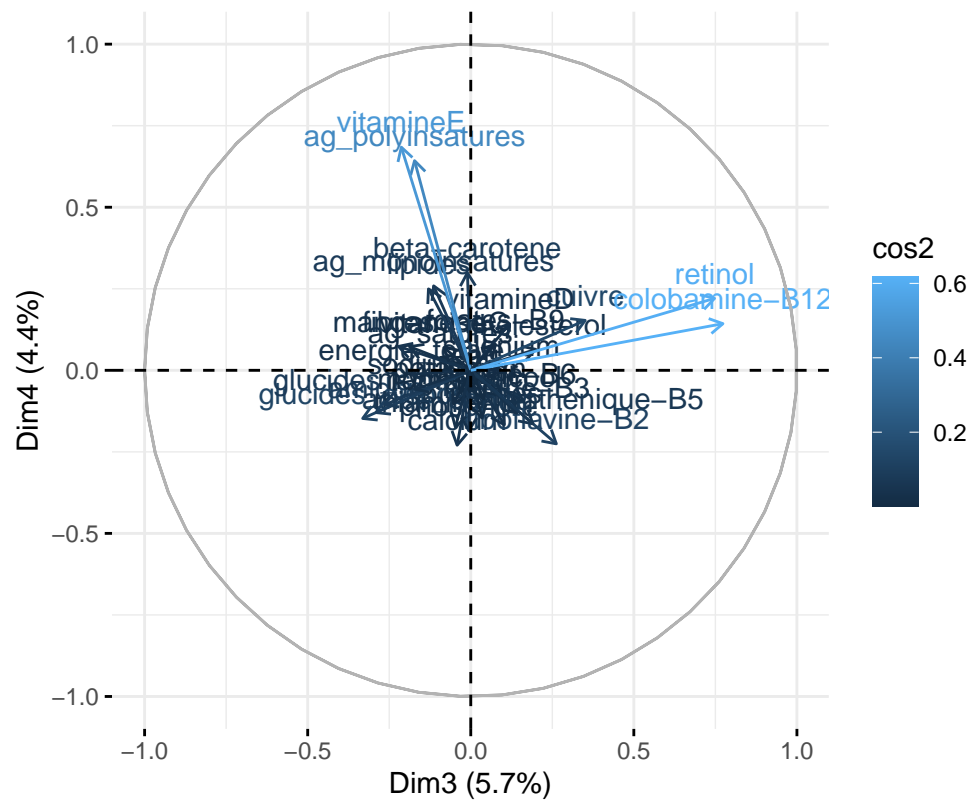|	eigenvalue	variance.percent	cumulative.variance.percent
Dim.24	0.1749078	0.4858549	97.82446
Dim.25	0.1637204	0.4547789	98.27924
Dim.26	0.1297530	0.3604250	98.63966
Dim.27	0.1142207	0.3172796	98.95694
Dim.28	0.0924240	0.2567334	99.21367
Dim.29	0.0841125	0.2336457	99.44732
Dim.30	0.0674378	0.1873273	99.63465
Dim.31	0.0523870	0.1455195	99.78017
Dim.32	0.0429028	0.1191744	99.89934
Dim.33	0.0326686	0.0907462	99.99009
Dim.34	0.0022090	0.0061360	99.99622
Dim.35	0.0013241	0.0036780	99.99990
Dim.36	0.0000353	0.0000980	100.00000

Scree plot
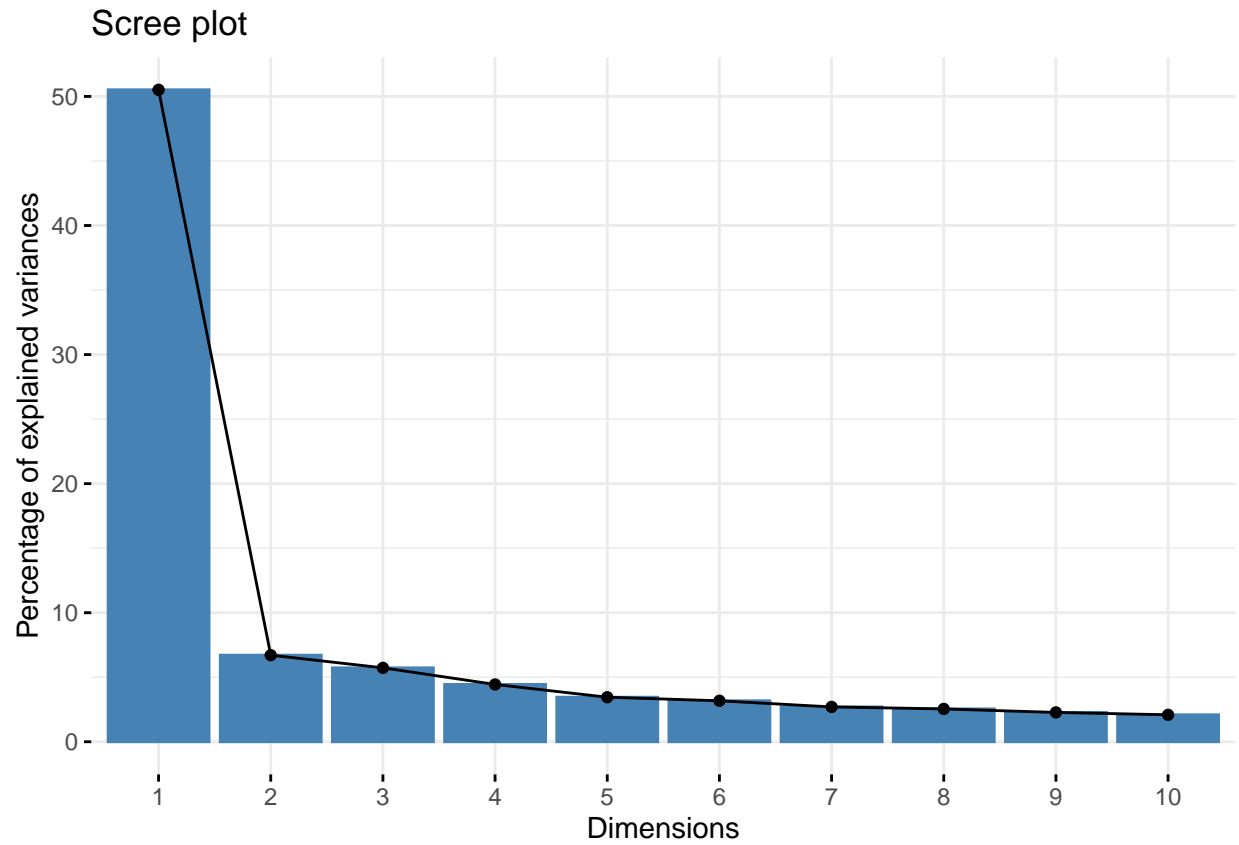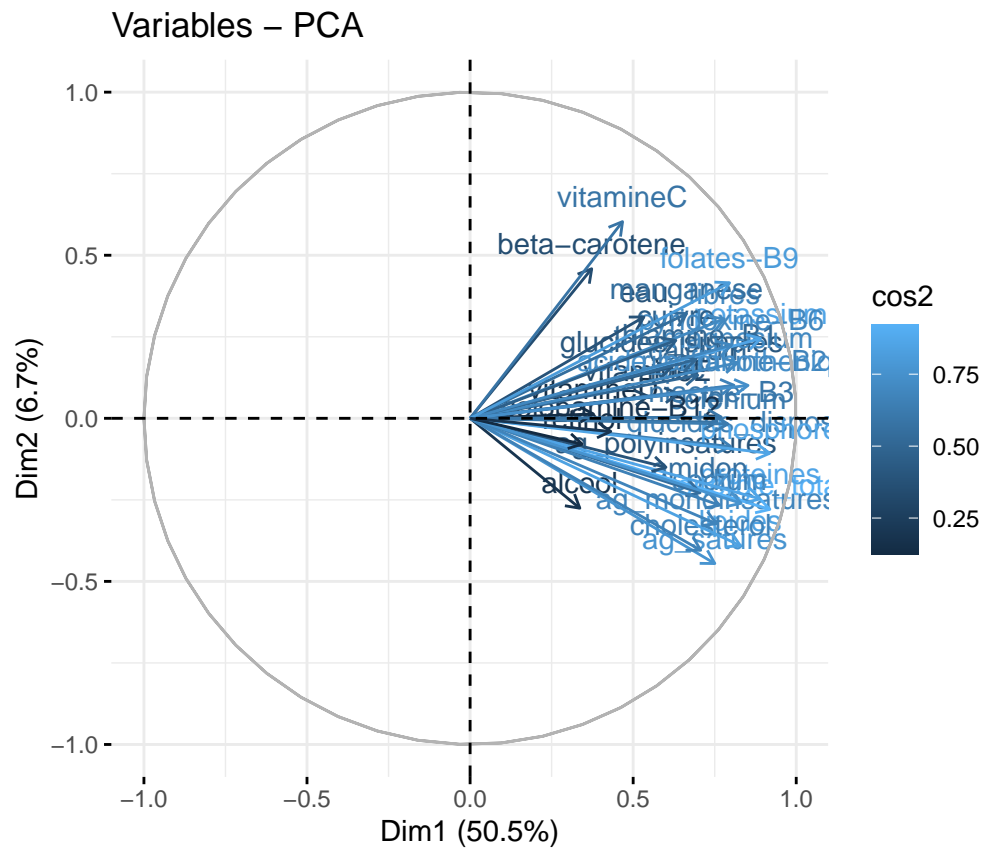
Variables – PCA
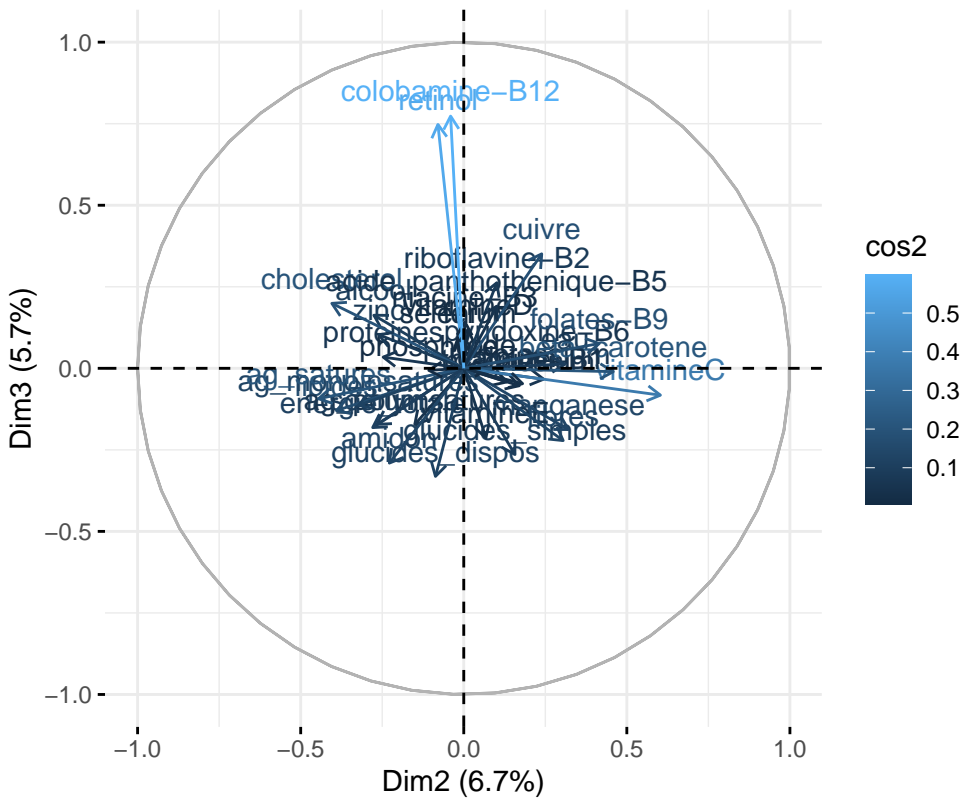
Variables – PCA

Variables – PCA

References