

# iaelyon

ÉCOLE UNIVERSITAIRE DE MANAGEMENT

UNIVERSITÉ  
**JEAN MOULIN**  
LYON 3



## I- INTRODUCTION

La ville de Toulouse est découpée en 6 secteurs et chaque secteur contient de 3 à 5 quartiers. La dénomination officielle de chaque quartier contient un ou plusieurs noms de lieux, afin de simplifier leur appellation nous utiliserons un seul de ces noms (choisi de façon arbitraire). Par exemple le troisième quartier du secteur 2 s'appelle Fontaine-Lestang, Bagatelle, Papus, il est numéroté 2.3 et on l'appellera Bagatelle. Pour chacun de ces quartiers nous avons à disposition le nombre d'écoles élémentaires, d'écoles maternelles, de galeries d'art, de gymnases, de musées et de théâtres (ces variables sont classées par ordre alphabétique).

La collecte des données a été effectuée grâce aux données disponibles sur <https://data.toulouse-metropole.fr> qui sont mises à disposition aux conditions de la Licence ODbL (Open Database License). Certaines données peuvent être téléchargées au format kmz qui en permettent une géolocalisation (dans google maps par exemple ou opengeoedata). Connaissant la délimitation des quartiers toulousains les données du fichier `donnees_quartiers_toulouse.xls` ont été obtenues en comptant les différents équipements au sein de chaque quartier.

Numer du quartier	Nom du quartier	Nom court	Ecole élémentaire	Ecole maternelle	Galerie d'art	Gymnase	Musée	Théâtre
1.1	CAPITOLE	Capitole	5	6	19	3	6	5
1.2	AMIDONNIERS - CAFFARELLI	Compans	2	2	2	2	0	0
1.3	CHALETS - BAYARD - BELFORT - SAINT AUBIN - DUPUY	Saint Aubin	3	3	4	2	0	3
2.1	SAINT-CYPRÉN	St Cyrien	3	4	5	2	4	3
2.2	CROIX DE PIERRE - ROUTE D'ESPAGNE	Croix de Pierre	3	3	0	1	0	1
2.3	FONTAINE LESTANG - BAGATELLE - PAPUS	Bagatelle	7	9	1	2	0	0
2.4	FONTAINE BAYONNE - CARTOUCHERIE	Cartoucherie	4	3	0	2	0	0
3.1	MINIMES - BARRIÈRE DE PARIS	Minimes	5	7	2	5	0	2
3.2	SEPT DENIERS - GINESTOUS	Sept Deniers	1	3	0	0	0	1
3.3	LALANDE - GRAND SELVE	Lalande	3	3	0	2	0	0
3.4	TROIS COCUS - BORDEROUGE - CROIX DAURADE - PALEFICAT	Borderouge	6	6	1	3	1	1
4.1	BONNEFOY - ROSERAIE - GRAMONT	La Roseraie	3	4	3	0	0	3
4.2	JOLIMONT - SOUPETARD - BONHOURE	Jolimont	5	5	2	4	0	1
4.3	CÔTE PAVÉE - CHÂTEAU DE L'HERS - LIMAYRAC	Côte Pavée	4	5	0	2	1	0
5.1	PONT DES DEMOISELLES - MONTAUDRAN - LA TERRASSE	Montaudran	5	6	0	2	0	0
5.2	RANGUEUIL - SAUZELONG - PECH DAVID - POUOURVILLE	Rangueil	5	6	1	2	0	1
5.3	SAINT MICHEL - LE BUSCA - EMPALOT - SAINT AGNE	St Michel	6	6	0	3	4	2
6.1	ARENES ROMAINES - SAINT MARTIN DU TOUCH	St Martin	3	3	0	1	0	0
6.2	LARDENNÉ - PRADETTE - BASSO CAMBO	Basso Cambo	4	5	0	3	0	0
6.3	MIRAIL - REYNERIE - BELLEFONTAINE	Miral	10	11	2	4	0	0
6.4	SAINT SIMON	St Simon	2	2	0	0	0	0
6.5	LAFOURGUETTE	La fourguette	1	1	0	1	0	0

**Notre objectif est de caractériser certains quartiers selon les variables que nous possédons.**

Pour se faire, nous utiliserons une ACP car nous n'avons que des variables quantitatives. Nous montrerons tout d'abord que l'ACP est pertinente puis nous effectuerons cette ACP. Nous dégagerons ensuite des facteurs principaux pour finalement interpréter la position des quartiers sur ces facteurs principaux.

## II- ACP

Nous allons tout d'abord récupérer les données stockées dans le fichier excel, que nous stockerons dans la variable *datas*.

A l'aide de la fonction *setwd()*, nous changeons le répertoire de travail de R afin qu'il puisse trouver notre fichier.

	Ecole_elementaires	Ecole_maternelles	Galleries_d_art	Gymnases	Musees	Theatres
Capitole	5	6	19	3	6	5
Compans	2	2	2	2	0	0
Saint Aubin	3	3	4	2	0	3
St Cyprien	3	4	5	2	4	3
Croix de Pierre	3	3	0	1	0	1
Bagatelle	7	9	1	2	0	0
Cartoucherie	4	3	0	2	0	0
Minimes	5	7	2	5	0	2
Sept Deniers	1	3	0	0	0	1
Lalande	3	3	0	2	0	0
Borderouge	6	6	1	3	1	1
La Roseraie	3	4	3	0	0	3
Jolimont	5	5	2	4	0	1
Cote Pavée	4	5	0	2	1	0
Montaudran	5	6	0	2	0	0
Rangueil	5	6	1	2	0	1
St Michel	6	6	0	3	4	2
St Martin	3	3	0	1	0	0
Basso Cambo	4	5	0	3	0	0
Miral	10	11	2	4	0	0
St Simon	2	2	0	0	0	0
La fourguette	1	1	0	1	0	0

Le tableau de données est chargé dans R. Afin de réduire la dimension pour faire l'ACP, il est nécessaire d'avoir des corrélations significatives. On vérifie cela à l'aide de la fonction *cor()*.

> cor=cor(datas)						
> cor	Ecole_elementaires	Ecole_maternelles	Galleries_d_art	Gymnases	Musees	Theatres
Ecole_elementaires	1.000000000	0.94530314	0.1194978	0.6843863	0.1456505	-0.001498905
Ecole_maternelles	0.945303143	1.00000000	0.1587734	0.6383556	0.1452210	0.061914377
Galleries_d_art	0.119497802	0.15877341	1.0000000	0.2160310	0.7310296	0.794450340
Gymnases	0.684386276	0.63835558	0.2160310	1.0000000	0.2087727	0.128173502
Musees	0.145650508	0.14522100	0.7310296	0.2087727	1.0000000	0.701013592
Theatres	-0.001498905	0.06191438	0.7944503	0.1281735	0.7010136	1.000000000

On constate que certaines variables sont très corrélées entre elles. La réalisation d'une ACP consiste à étudier ces corrélations. Pour pouvoir travailler sur notre ACP, nous allons charger les extensions *factoextra* et *FactoMineR*.

```
> library(factoextra)
Le chargement a nécessité le package : ggplot2
Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
> library(FactoMineR)
```

Nous allons maintenant effectuer l'ACP à l'aide de la fonction *PCA()*. À noter que cette fonction va normaliser nos données.

```
> ACP=PCA(datas,scale.unit=TRUE,ncp=6,graph=TRUE)
> ACP
**Results for the Principal Component Analysis (PCA)**
The analysis was performed on 22 individuals, described by 6 variables
*The results are available in the following objects:

      name                description
1  "$eig"            "eigenvalues"
2  "$var"             "results for the variables"
3  "$var$coord"       "coord. for the variables"
4  "$var$cor"          "correlations variables - dimensions"
5  "$var$cos2"         "cos2 for the variables"
6  "$var$contrib"      "contributions of the variables"
7  "$ind"              "results for the individuals"
8  "$ind$coord"        "coord. for the individuals"
9  "$ind$cos2"          "cos2 for the individuals"
10 "$ind$contrib"       "contributions of the individuals"
11 "$call"              "summary statistics"
12 "$call$centre"       "mean of the variables"
13 "$call$ecart.type"   "standard error of the variables"
14 "$call$row.w"         "weights for the individuals"
15 "$call$col.w"         "weights for the variables"
```

Le logiciel R nous retourne les résultats de l'ACP ainsi que deux graphiques que nous traiterons plus tard.

Dans une ACP centrée réduite la somme des valeurs propres est égale au nombre de variables et égale à l'inertie du nuage de points. Nous lisons les valeurs propres suivantes puis les additionnons entre elles.

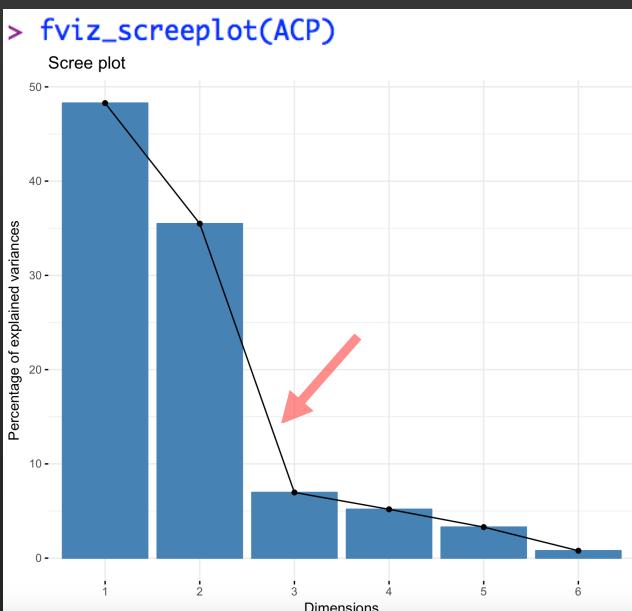
```
> get_eigenvalue(ACP)
      eigenvalue variance.percent cumulative.variance.percent
Dim.1 2.89676175      48.2793625           48.27936
Dim.2 2.12919449      35.4865749           83.76594
Dim.3 0.41815790      6.9692984           90.73524
Dim.4 0.31098030      5.1830050           95.91824
Dim.5 0.19756350      3.2927250           99.21097
Dim.6 0.04734206      0.7890343          100.00000

> sumvp=apply(get_eigenvalue(ACP),2,sum)
> sumvp
      eigenvalue      variance.percent cumulative.variance.percent
                           6.0000          100.0000                  517.9097
```

La somme des valeurs propres est bien égale à 6.

### III- CHOIX DU NOMBRE DE FACTEURS

**Critères des coudes de Catell** : La fonction `fviz_screeplot()` nous permet d'obtenir un éboulis des valeurs propres.



On peut voir un saut qui laisserait penser qu'il faudrait choisir deux axes.

**Critère de Kaiser** : La fonction `get_eigenvalue()` nous donne les valeurs propres suivantes :

```
> get_eigenvalue(ACP)
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.89676175	48.2793625	48.27936
Dim.2	2.12919449	35.4865749	83.76594
Dim.3	0.41815790	6.9692984	90.73524
Dim.4	0.31098030	5.1830050	95.91824
Dim.5	0.19756350	3.2927250	99.21097
Dim.6	0.04734206	0.7890343	100.00000

On constate que seules les deux premières valeurs propres sont  $>1$ . On choisira donc deux axes selon le critère de Kaiser.

**3ème critère du pourcentage d'inertie conservée** : La fonction `get_eigenvalue()` nous donne également la variance expliquée par chaque valeur propre dans la deuxième colonne, ainsi que ces variances cumulées dans la troisième.

On constate que les deux premières valeurs propres expliquent à elles seules 83,77% de la variance, soit plus de 70%. D'après le critère de Kaiser, on gardera donc deux axes.

**En conclusion, les critères de sélection du nombre de facteurs nous poussent à choisir de travailler sur deux axes.**

## IV- ANALYSE DES VARIABLES ET DES INDIVIDUS

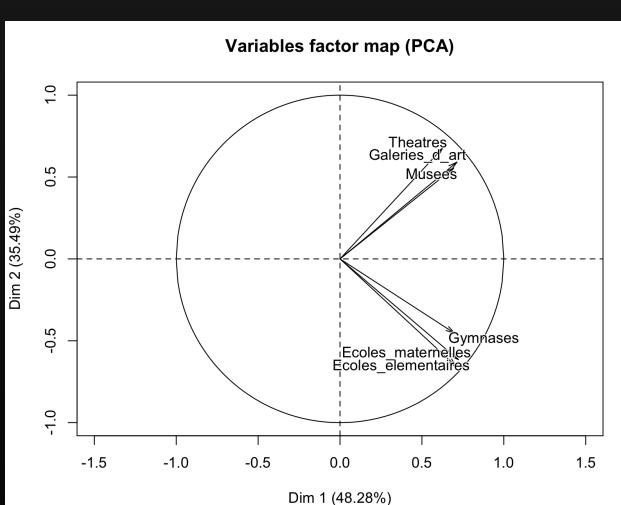
Tout d'abord, nous devons nous demander si les variables sont bien représentées. Pour cela, nous allons commencer par lire les cosinus carré avec \$var\$cos2 puis nous analyserons le cercle des corrélations que nous avons obtenu en lançant l'ACP.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Ecole_elementaires	0.5042318	0.4318341	3.835698e-02	0.0004925386	3.354687e-06	2.508124e-02
Ecole_maternelles	0.5219872	0.3786503	7.217061e-02	0.0050609806	8.945960e-04	2.123631e-02
Galerie_d_art	0.5098104	0.3493369	3.463430e-04	0.0282603353	1.122409e-01	5.112863e-06
Gymnases	0.4889586	0.2060720	3.045544e-01	0.0000610155	9.809739e-05	2.558550e-04
Musees	0.4804956	0.3070665	2.699547e-03	0.2072330717	2.232259e-03	2.730786e-04
Theatres	0.3912781	0.4562347	3.000775e-05	0.0698723558	8.209432e-02	4.904699e-04

Nous allons maintenant calculer les cosinus carré cumulés des deux premiers facteurs.

```
> cos2F=ACP$var$cos2[1:6,1:2]
> cos2F
      Dim.1   Dim.2
Ecole_elementaires 0.5042318 0.4318341
Ecole_maternelles 0.5219872 0.3786503
Galerie_d_art     0.5098104 0.3493369
Gymnases          0.4889586 0.2060720
Musees            0.4804956 0.3070665
Theatres          0.3912781 0.4562347
> cos2cum2F=apply(cos2F,1,sum)
> t(cos2cum2F)
      [,1]
Ecole_elementaires 0.9360659
Ecole_maternelles 0.9006375
Galerie_d_art     0.8591473
Gymnases          0.6950306
Musees            0.7875620
Theatres          0.8475128
```

On constate que les cosinus carrés cumulés de nos six variables sont tous supérieurs à 60% sur les deux facteurs. Les 6 variables sont donc bien représentées.



Le cercle des corrélations obtenu au lancement de l'ACP est le suivant :

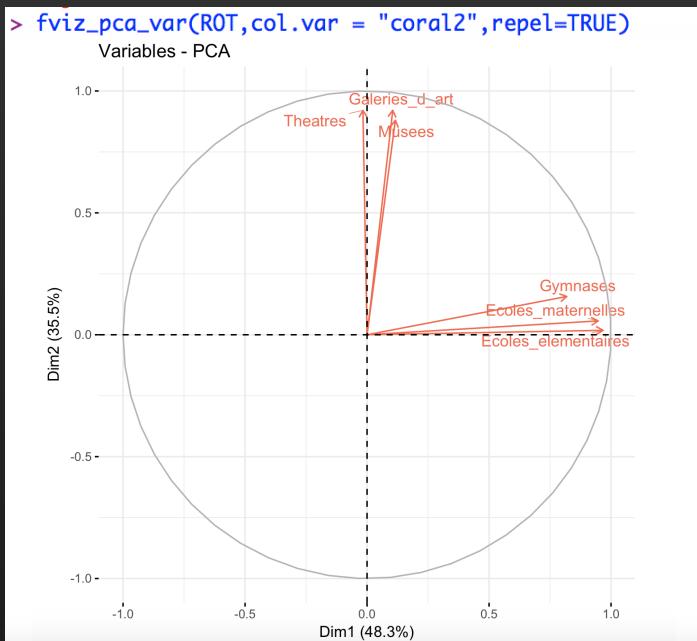
On constate que les variables sont bien représentées car elles sont proches du bord du cercle.

L'examen du cercle des corrélations montre un effet taille. En effet, toutes les variables de départ sont positivement corrélées au premier facteur. On remarque aussi qu'une rotation serait pertinente afin de faire coïncider les groupes de variables avec les facteurs.

Après avoir déplacer le fichier rotationRB.R dans notre répertoire de travail et l'avoir charger dans R, nous lançons la rotation à l'aide de la fonction fournie.

```
> ROT=rotation(ACP)
> ROT
**Results for the Principal Component Analysis (PCA)**
The analysis was performed on 22 individuals, described by 6 variables
*The results are available in the following objects:
```

On demande à R de réaliser un nouveau cercle de corrélations après rotation.



On remarque un premier groupe de variables (*Gymnases*, *Ecoles\_maternelles*, *Ecole\_elementaires*) qui sont fortement corrélées entre elles et corrélées au premier axe.

De même, on remarque un second groupe de variables (*Theatres*, *Galeries\_d\_art*, *Musees*) qui sont fortement corrélées entre elles et corrélées au deuxième axe.

Pour confirmer cela, nous allons regarder la contribution des variables aux différents axes avec `$var$contrib`.

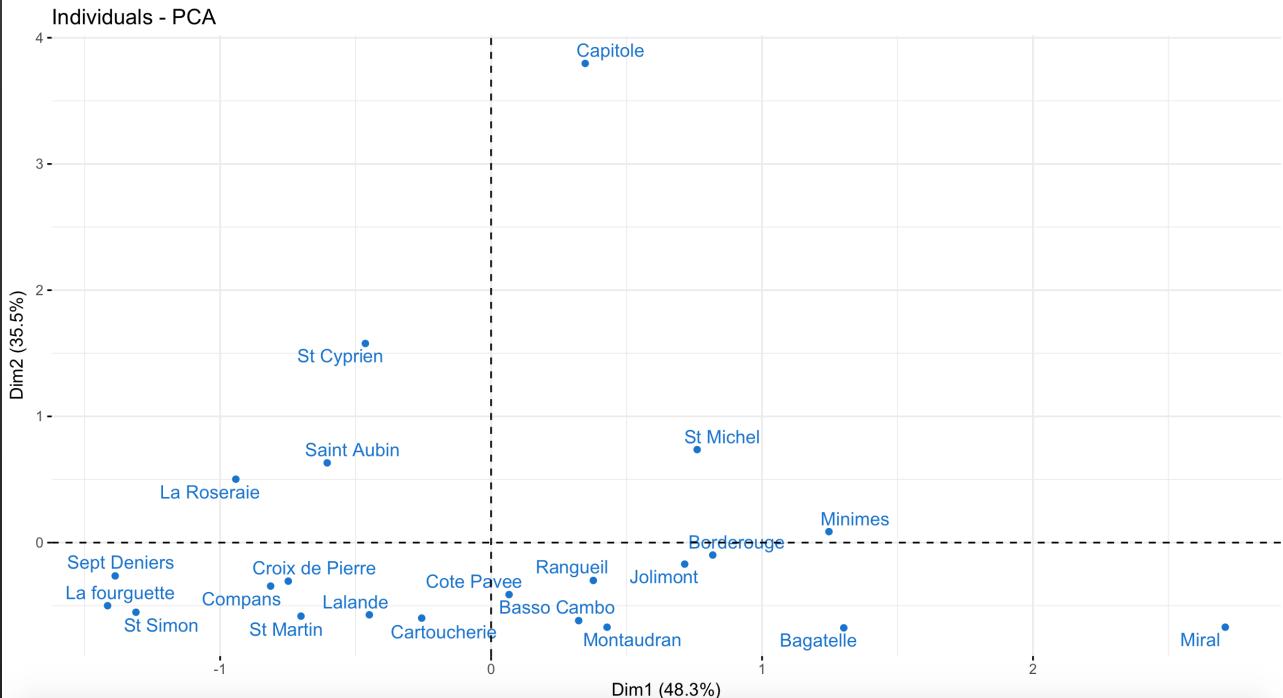
```
> ROT$var$contrib
[,1]      [,2]
Ecole_elementaires 37.01027932 0.01312954
Ecole_maternelles 35.49422885 0.12932774
Galeries_d_art    0.43920572 33.95380394
Gymnases          26.51341596 0.98840860
Musees             0.53156242 30.99419195
Theatres           0.01130773 33.92113824
```

Il apparaît que ce sont les 3 variables (*Gymnases*, *Ecole\_maternelles*, *Ecole\_elementaires*) qui contribuent le plus au premier axe. Ce premier axe correspond donc à ce premier groupe de variables et peut être interprété comme un axe de *Points d'accès à l'éducation*.

Les 3 autres variables (*Theatres*, *Galeries\_d\_art*, *Musees*) sont les variables qui contribuent le plus au deuxième axe. Cet axe peut être interprété comme un axe de *Point d'accès à la culture*.

On va maintenant analyser nos 22 individus. On demande à R de réaliser une représentation de nos individus sur les deux axes avec la fonction `fviz_pca_ind()`.

```
> fviz_pca_ind(ROT,col.ind = "dodgerblue3",repel=TRUE)
```



Nous allons étudier la qualité de représentation de ces individus grâce à leur cosinus carré cumulés ainsi que leur contribution sur les deux axes.

```
> cos2cum2Find=apply(ROT$ind$cos2[1:22,1:2],1,sum)
```

	Compans	Saint Aubin	St Cyprien	Croix de Pierre
Capitole	0.6839071	0.4868431	0.8765705	0.9234190
Bagatelle	0.7608956	0.5691837	0.8898870	0.8305828
Borderouge	0.4659698	0.4772540	0.4754027	0.8834772
Rangueil	0.5014588	0.5101458	0.5014588	0.5101458
Sept Deniers	0.9269849	0.9390102	0.9390102	0.9390102
St Simon	0.9444682	0.9444682	0.9444682	0.9444682

```
> ACP$ind$contrib
```

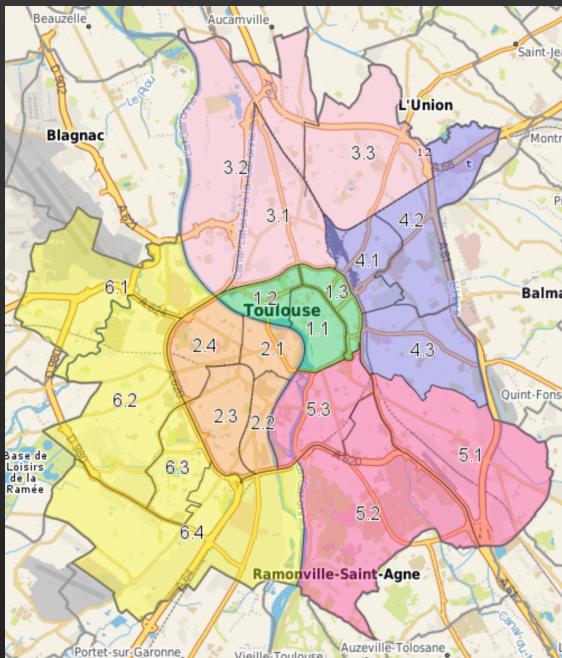
	Dim.1	Dim.2
Capitole	3.772131e+01	2.834155e+01
Compans	3.096971e+00	4.515440e-01
Saint Aubin	1.252476e-05	3.478582e+00
St Cyprien	2.614987e+00	9.678305e+00
Croix de Pierre	2.565760e+00	4.048070e-01
Bagatelle	1.007302e+00	8.767608e+00
Cartoucherie	1.631318e+00	2.930738e-01
Minimes	4.183196e+00	2.918000e+00
Sept Deniers	6.362747e+00	2.702076e+00
Lalande	2.358944e+00	4.652884e-02
Borderouge	1.237483e+00	1.848707e+00
La Roseraie	4.985398e-01	4.682787e+00
Jolimont	7.170899e-01	1.735332e+00
Cote Pavée	2.563339e-01	5.344515e-01
Montaudran	1.102330e-01	2.764135e+00
Rangueil	1.882358e-02	1.037222e+00
St Michel	5.097418e+00	9.256230e-05
St Martin	3.760595e+00	1.974990e-02
Basso Cambo	1.732076e-01	2.039094e+00
Miral	1.008560e+01	2.532520e+01
St Simon	8.009321e+00	1.182905e+00
La fourguette	8.492812e+00	1.748252e+00

Les quartiers ayant un cosinus carré cumulé supérieur à 70% peuvent être interprétés. Entre 50% et 70%, on les interprète avec des réserves. En dessous de 50%, on ne peut pas les interpréter. Ainsi, les quartiers La Roseraie, Jolimont et Cote Pavée ne seront pas interpréter.

Les contributions des individus sont excessives si elles dépassent  $4/22=18,18\%$ . La contribution de Capitole au premier axe est de 37,72%, et de 21,34% au second axe. On considère Capitole comme une valeur exceptionnelle. Il faudrait refaire une ACP sans cet individu et le réintégrer au biplot, mais nous poursuivrons ici avec.

## V. ANALYSE ET CONCLUSIONS

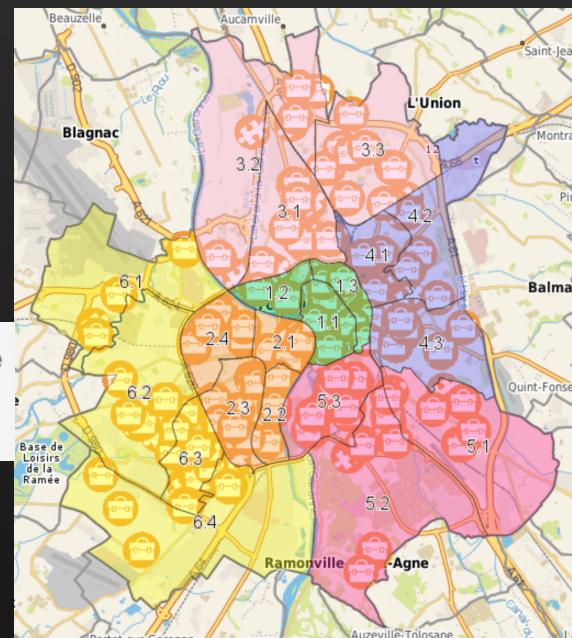
Nous allons tout d'abord utiliser la carte numérique mis à disposition sur le site <https://www.plan.toulouse.fr> pour profiter d'une vision géographique de la situation.



- Ecole Elementaire
- Ecole Maternelle

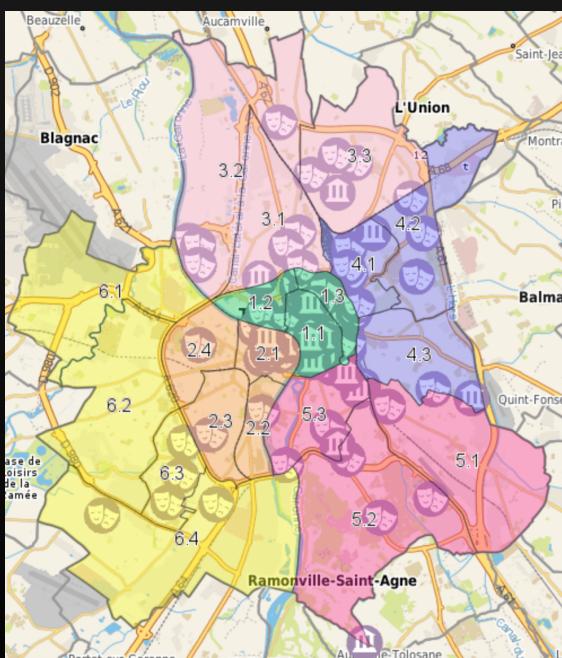
On remarque que chaque secteur bénéficie de points d'accès à l'éducation. Sur Toulouse en général, ces points d'accès semblent plutôt centrer que proche banlieue.

À l'aide des outils fournis sur le site, nous allons visualiser la position géographique des écoles maternelles et des écoles primaire dans les 6 secteurs de Toulouse, puis la position des musées et théâtres toulousains.



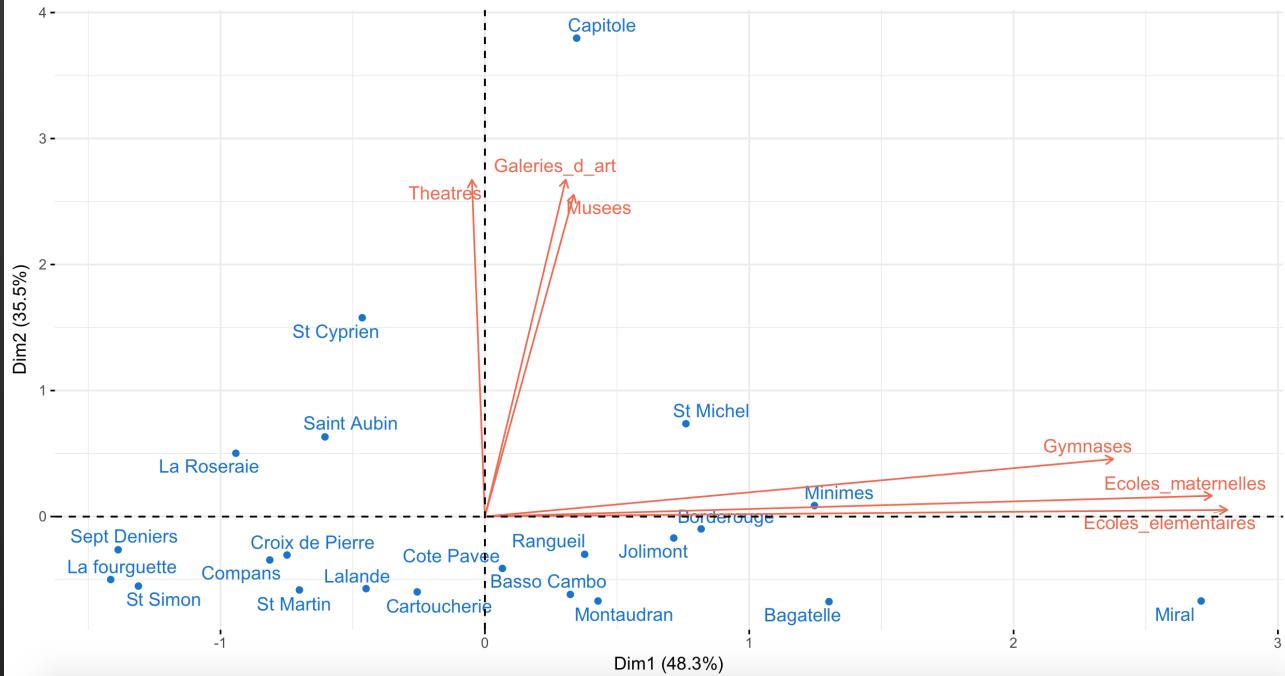
- Musées
- Théâtres

Les points d'accès à la culture sont fortement concentrées dans le centre de Toulouse. Si tous les secteurs peuvent profiter de théâtres, la majorité des musées se situent dans ou autour du premier secteur ; le centre-ville.



Pour l'analyse, nous allons réaliser un biplot des deux représentations (respectivement des variables et des individus) avec la fonction `fviz_pca_biplot()`.

```
> fviz_pca_biplot(ROT,col.var = "coral2",col.ind = "dodgerblue3",repel=TRUE)
PCA - Biplot
```



Les quartiers La Roseraie, Jolimont et Cote Pavée ne sont pas interprétables d'après l'analyse des individus.

On remarque que le quartier qui est le plus actif est le Capitole. En plein centre-ville, celui-ci présente de nombreux points d'accès à la culture et de nombreux points d'accès à l'éducation.

Le quartier Miral est celui qui bénéficie du plus de points d'accès à l'éducation.

Les quartiers St-Simon et La Fourquette sont des quartiers ayant très peu de points d'accès à la culture et à l'éducation. Il s'agit probablement de quartiers résidentiels.

