

Silvère Bonnabel : Mot introductif.

Voici le rapport du projet de Lilou Augeray et Thomas Clerte Barthélémy dans le cadre du cours de Mathématiques Appliquées A au semestre 5 de la licence de mathématiques. Ce cours porte sur les techniques d'analyse en composantes principales, et d'analyse factorielle des correspondances. J'ai transmis avec les accords nécessaires de tous des fichiers ayant été au préalable transformés par Benoit Martel pour les rendre compatibles avec des contraintes d'anonymat fortes, et ai obtenu un engagement de confidentialité de la part des étudiant.es.

L'analyse factorielle des correspondances a été mise en place par l'école française d'analyse des données dans les années soixante et soixante-dix, en lien avec le développement des ordinateurs. Il s'agit d'une méthode de statistique descriptive (ou inductive) visant à faire ressortir des liens entre variables catégorielles et permettant de les *visualiser*. Les travaux de cette école pilotée à l'université de Rennes par un mathématicien particulièrement brillant, Jean-Paul Benzecri, avec l'aide de Brigitte Cordier Escofier, ont connu un fort retentissement dans les sciences humaines et sociales. La théorie est simple et puissante à la fois, tout en étant très élégante.

Les motivations initiales de Jean-Paul Benzecri étaient l'analyse de textes, de façon inductive (c'est-à-dire qu'au lieu de se définir des règles, des axiomes, et d'en déduire des choses, on ne suppose rien et l'on cherche à faire apparaître des règles et des lois par inférence). Un exemple historique est celui traité par sa doctorante Brigitte Cordier Escofier, qui en se concentrant sur une quinzaine de mots et leur fréquence dans le vocabulaire, a permis de mettre en place un sociogramme des personnages de Phèdre de Racine. La fréquence d'utilisation de ces mots opposait en particulier les personnages au langage déférent à ceux au langage plus passionné et emporté. On sent donc bien ici l'intérêt de ces méthodes pour notre sujet.

Quelques avertissements au lecteur.

- Le premier est d'ordre mathématique. L'intérêt de la méthode est que sur les graphiques ci-dessous, la place qu'occupent les mots dans le graphique n'est pas simplement illustrative : elle est mathématique, en ce que le graphique est un reflet fidèle de propriétés mathématiques. En effet, la visualisation par analyse factorielle suit des lois mathématiques précises. En gros, les distances entre les mots et leur agencement dans le graphique correspondent à de véritables distances statistiques. A cet égard la méthode est bien plus précise que les graphes obtenus précédemment sur les co-occurrences de mots, qui n'étaient qu'illustratifs.
- La méthode retenue ignore les co-occurrences de mots. Le fait que le mot A ait toujours été énoncé avec le mot B par exemple, n'est pas pris en compte par la méthode. A cet égard l'analyse est complémentaire des analyses déjà faites sur les graphes de co-occurrence.
- Je n'ai pas contrôlé le travail des étudiant.es en termes de regroupements en concepts et analyse des fichiers. Je vous livre leur travail en l'état.
- Je donne des informations sur la méthode mais ne me lance dans aucune interprétation : ce n'est pas mon rapport.

Ce travail apparaît donc comme un premier défrichage fructueux, qui demande à être largement complété lors de stages à venir. J'insère quelques commentaires en bleu au fil du texte.

Analyse de données : La vision de la mine en fonction de plusieurs paramètres

Silvère Bonnabel : le but de ce travail est de regrouper des mots en concepts, puis d'étudier si l'occurrence de ces concepts dans diverses catégories de population varie statistiquement. A l'image des personnages de Phèdre, cela devrait permettre d'établir des profils de réaction à la mine en fonction de divers facteurs. Les concepts retenus par les étudiant.es sont les suivants

Développement dégradation environnementale échange commercial économie
nickel politique pollution relation politique ressource minière risque travail

I- Indépendance possible des données, test du chi carré

Silvère Bonnabel : Ce premier test permet de mettre en évidence une variation significative des concepts liés à la mine dans les différentes catégories de population. Elle s'avère très au-delà d'une simple fluctuation statistique qui serait liée au hasard d'échantillonnage.

Tout d'abord, les tests du chi carré effectués sur nos trois tableaux nous montrent que les catégories de mots utilisés sont liées à l'âge, aux lieux de vie, et au lien qu'entretient la personne interrogée avec la mine et les secteurs miniers.

```
|  
> summary(res1)  
  
Call:  
CA(X = data1)  
  
The chi square of independence between the two variables is equal to 60.41344 (p-value = 0.0008202038 ).
```

```
> summary(res2)
```

```
Call:
```

```
CA(X = data2)
```

```
The chi square of independence between the two variables is equal to 56.65921 (p-value = 2.311686e-05 ).
```

```
> summary(res3)
```

```
Call:
```

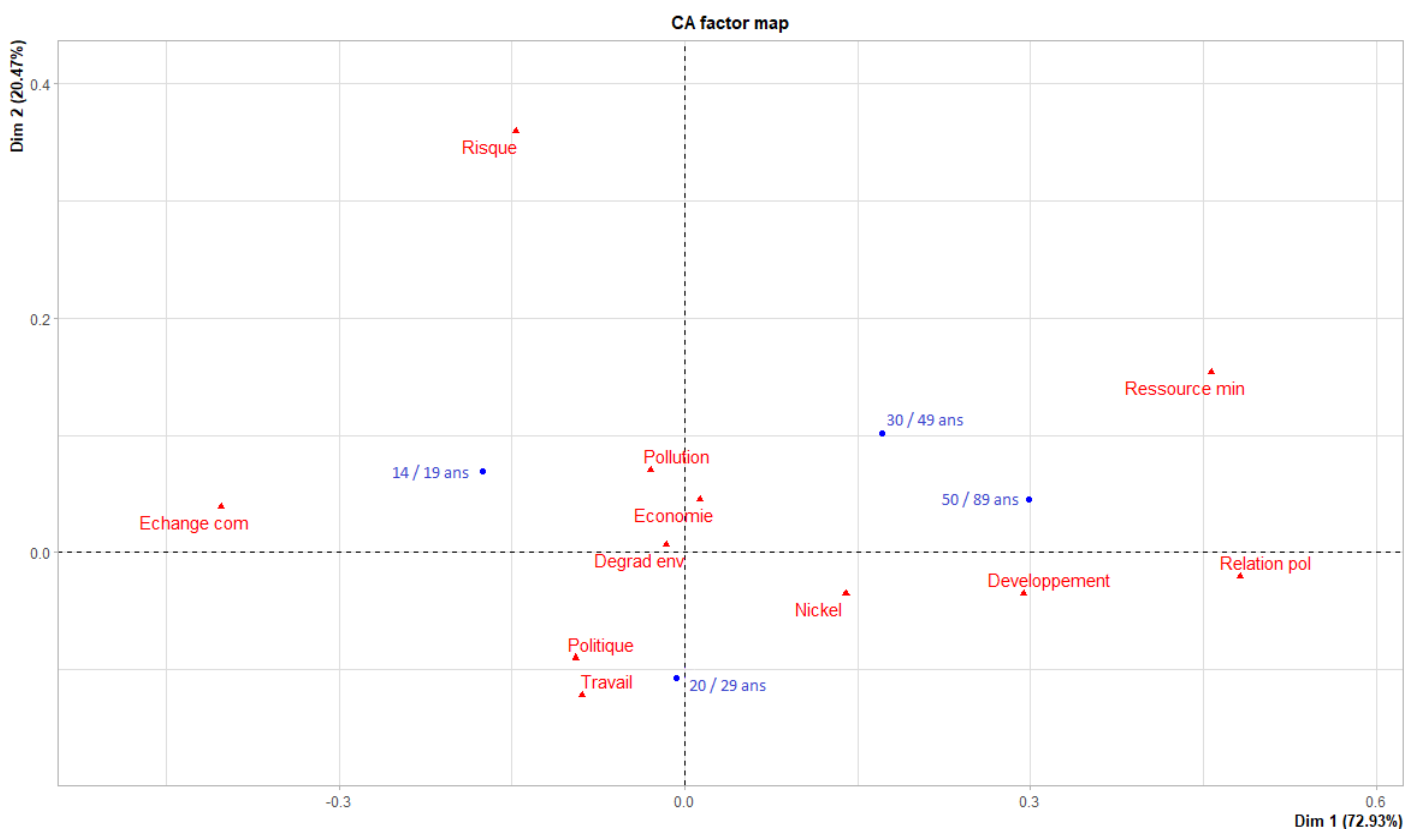
```
CA(X = data3)
```

```
The chi square of independence between the two variables is equal to 132.5268 (p-value = 1.306514e-18 ).
```

II- Analyse en composante factorielle

A) La vision de la mine de nos jours en fonction de l'âge des individus

Silvère Bonnabel : voici un graphique typique d'analyse des correspondances. Les axes



capturent des écarts à l'indépendance. Par exemple, le concept de « dégradation environnementale » est situé au centre, cela veut dire qu'il apparaît à peu près en même proportion dans les diverses catégories d'âge. Les concepts situés loin de l'origine sont plus sujets à des variations entre catégories. On ne parle pas de distance mais plutôt d'écart à l'indépendance. L'indépendance étant caractérisée par une fréquence d'occurrence dans chaque catégorie identique à l'occurrence dans la population totale.

Les axes retiennent les directions de plus grand écart à l'indépendance statistique. Ils permettent de faire des oppositions entre concepts, l'axe 1 capturant les plus grandes différences, et l'axe 2 capturant des différences mais dans une moindre mesure.

On peut faire la même analyse avec les catégories de population par concept et obtenir deux axes aussi.

Finalement il est à noter – et c'est ce qui est fascinant dans cet outil – que les concepts et les catégories de population apparaissent sur le même graphique. Il y a une propriété barycentrique qui veut que si une catégorie apparaît proche de concepts, alors ces concepts sont sur-représentés dans cette catégorie (avec une formule mathématique bien précise). Cela rend licite la représentation sur un même graphique, et cela la rend informative : l'axe 1 oppose des concepts et des âges (vieux versus jeunes). On voit donc que le concept « relation politique » est plus lié aux catégories âgées que jeunes.

On voit immédiatement ressortir sur le graphique des choses qui bien sûr sont présentes dans les données mais seraient peut-être plus difficiles à voir du fait de la multiplicité des concepts et des catégories sans avoir accès à ces graphiques.

Concepts versus catégories de population

A : Développement

E : dégradation environnementale

F : échange commercial

G : économie

N : nickel

Q : politique

R : pollution

S : relation politique

T : ressource minière

U : risque

X : travail

- 1 : 14 / 19 ans
- 2 : 20 / 29 ans
- 3 : 30 / 49 ans
- 4 : 50 / 89 ans

Après avoir fait tourner une AFC sur notre premier tableau on remarque que l'axe 1 oppose les échanges commerciaux avec les relations politiques et les ressources minières, et que l'axe 2 quant à lui oppose les risques avec le travail et la politique.

On constate que les 14/19 ans ont une vision de la mine assez basée sur les échanges commerciaux, l'économie, les risques, la pollution et la dégradation environnementale qu'entraîne la mine. Les 20/29 ans ont une vision plus politisée de la mine, le travail et la dégradation environnementale sont aussi des catégories qui caractérisent beaucoup la mine pour cette tranche d'âge.

On déduit de la fréquence des mots utilisés en rapport avec le développement que pour ces tranches d'âge la mine ne rime pas forcément avec le développement du pays (uniquement 7% pour les 14/19 ans et 13% pour les 20/29).

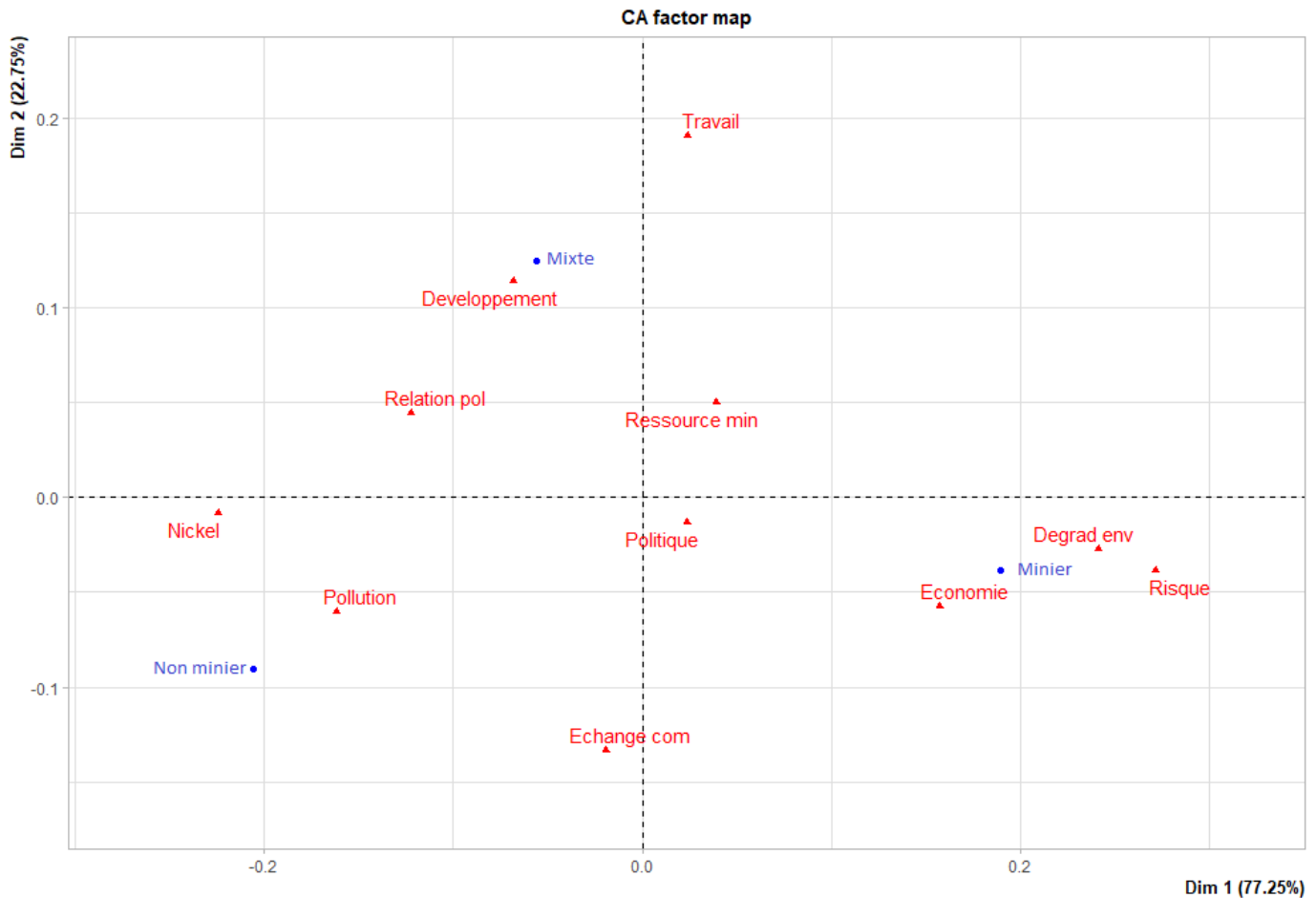
On remarque que les 30/49 ans ont une vision de la mine centrée autour de l'économie, de la ressource majeur de la mine qu'est le nickel et le développement du pays entraîné par celle-ci. C'est cette tranche d'âge qui accorde la plus grande valeur au nickel, en effet 21,4 % d'entre eux pense au mot richesse.

Enfin les 50/89 évoquent plutôt les aspects développement, relation politique et ressource minière. On peut penser ici que certains font le rapprochement entre la ressource et le développement, certainement parce que dans leurs jeunesse le nickel était considéré comme la richesse qui permettra de développer le pays.

Par ailleurs, ces catégories d'âge se préoccupe moins des échanges commerciaux que les autres catégories, plus jeunes.

B) La vision de la mine de nos jours en fonction du lieu de vie des individus

On observe que l'axe 1 oppose le nickel avec la pollution, tandis que l'axe 2 oppose le travail et les



Catégories versus concepts

1 : Minier

2 : Non minier

3 : Mixte

A : Développement

E : dégradation environnementale

F : échange commercial

G : économie

N : nickel

Q : politique

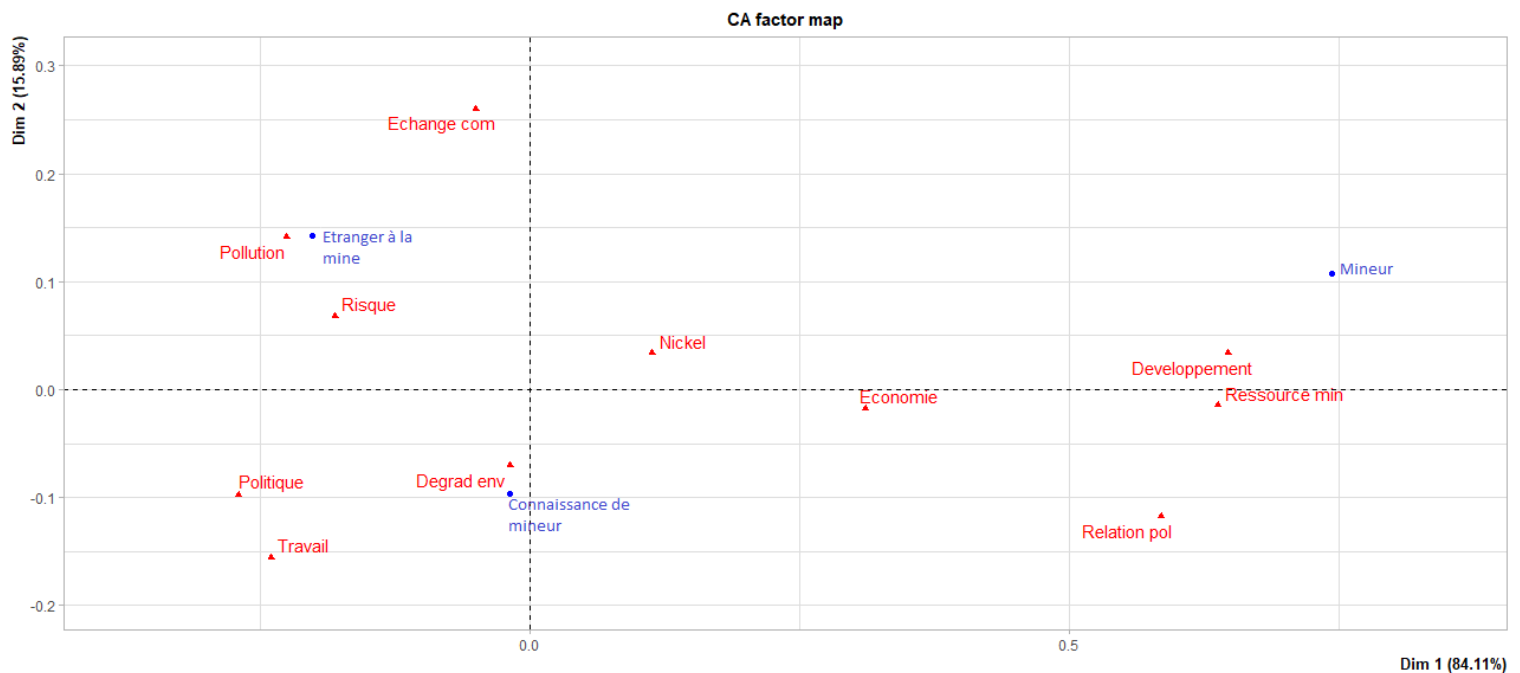
R : pollution
 S : relation politique
 T : ressource minière
 U : risque
 X : travail

Pour les gens habitant dans des lieux non-miniers, c'est le nickel (notamment la richesse due au nickel, 21,2% de ces personnes y font allusion) et la pollution qui leur viennent à l'esprit quand on parle de la mine de nos jours (même sans habiter près de secteur minier, les mines ne sont jamais représentées comme non polluante donc souvent c'est ce qui marque les esprits, surtout quand on n'y est pas confronté tout les jours).

Les personnes habitant dans des lieux mixtes, pensent plus au développement et au travail.

On constate que la politique et les relations politiques n'ont pas l'air d'être réellement influencées par les lieux de vies.

C) La vision de la mine de nos jours en fonction de l'implication des individus dans la mine



- 1 : Travaille dans la mine
- 2 : Ne travaille pas dans la mine mais connaît quelqu'un qui y travaille
- 3 : Ne travaille pas dans la mine et ne connaît personne qui y travail

A : Développement
E : dégradation environnementale
F : échange commercial
G : économie
N : nickel
Q : politique
R : pollution
S : relation politique
T : ressource minière
U : risque
X : travail

L'axe vertical oppose le travail avec les échanges commerciaux, et l'autre, le développement, les ressources minières avec la politique.

Les personnes qui travaillent dans le secteur de la mine ont une vision assez axée sur le développement, les ressources minières, et les relations politiques. Ce qui n'est pas étonnant car, que ce soit récemment ou depuis un moment, la mine s'est retrouvée à de nombreuses reprises dans des débats politiques. De plus ceux qui vivent de la mine ne se concentre pas forcément sur les côtés les plus négatifs mais sur ce qu'elle leur apporte de bon.

Les personnes ne travaillant pas dans la mine mais connaissant du monde qui y travail sont très proche des dégâts environnementaux mais aussi du travail et de la politique. Ce sont surtout des mauvais aspects de la mine, ça peut se comprendre car souvent ce que l'ont retient, ce qui marque le plus et le négatif, comme ils ne sont pas plongé dedans en y travaillant chaque jours, ils retiennent surement le négatif de ce que leur connaissance peuvent leur communiquer.

Enfin les gens n'ayant aucun rapport de près ou de loin avec la mine sont plus proches de la pollution, des risques et des échanges commerciaux. Comme pour le tableau précédent la pollution est ce qui frappe en premier ceux qui sont le plus éloignés de la mine. Les personnes qui sont dans la 3^{ème} catégorie, sont les plus susceptible de, n'avoir que les médias comme source d'info. Très souvent des informations sur la mine sont communiquées, soit quand il y a des fluctuations dans l'exportation, ou quand il y a des accidents ou des crises. Ainsi ceux qui ne sont pas dans le monde minier et qui ne le voit pas dans la vie courante auront tendance à ne voir que ces trois aspects de la mine.

En conclusion, en fonction du rapport que l'ont a avec la mine notre vision de celle-ci change. On constate que plus on y est impliqué, plus on aura tendance à percevoir les aspects positifs. Au contraire

les personnes moins concernées attribuent des valeurs plus souvent négatives à la mine. C'est sûrement une vision de la mine plus globale, opposée à une vision plus centrée sur la Nouvelle-Calédonie qui occasionne cette différence de pensée.