

# Typologie des compétences numériques des Français

Suppression des ind qui ont répondu 'nsp', 'rdr' ou 'comprend.pas' 3 fois ou +

*Laurent Mell, Nicolas Deporte Éco*

*17 novembre 2017*

L'enquête Capacity visait à alimenter les travaux entrepris dans le cadre du projet de recherche ANR « Capacity », portant sur les réalités de l'empowerment par les usages numériques en France, ainsi que les réflexions de l'Agence du numérique, partie prenante de cette enquête. Le questionnaire intègre également les « questions communes » à chaque pays membre du WIP (World Internet Project) auquel Marsouin participe en tant que représentant de la France. L'enquête (réalisée par Kantar TNS) s'est déroulée en face-à-face, au domicile des répondants entre novembre et décembre 2016 et a permis d'obtenir les réponses auprès d'un échantillon de 2036 personnes représentatif de la population française métropolitaine des 18 ans et plus. La représentativité de l'échantillon est assurée par la méthode des quotas sur le sexe croisé avec l'âge, la profession, la catégorie d'agglomération et la région géographique du répondant. Cette enquête a été financée par l'Agence nationale de la recherche et l'Agence du Numérique. Pour en savoir plus sur cette enquête : <https://www.marsouin.org/mot281.html>

## Introduction

A travers ce travail, nous proposons une nouvelle exploration des données de l'enquête Capacity en focalisant notre attention sur ce qui, dans le questionnaire, a été identifié comme des compétences numériques et regroupé en différents groupes de compétences. Quelques mots de description sur les compétences, regroupées en bloc/type de compétences.

## Quelques stat descriptives (sans doute dans Intro)

Présenter dans cette partie, un état des lieux rapide sur les compétences numériques des Français

Voir pe pour un graphique en barres empilé et prop à 100% pour les différents blocs de compétences -> comparaison de l'accessibilité de chaque usage.

% de gens qui se sentent compétents par bloc pour 80% des usages du bloc

## Méthodologie

Explication de la préparation de la base, notamment la suppression de ceux qui ont répondu trop souvent 'nsp', 'rdr' ou 'comprend.pas'

Présentation de l'analyse : AFM+HCPC qui se prête particulièrement bien à ces données structurées en blocs

# Résultats de la typologie

Description des 5 classes

1 graph avec les 5 classes projetées sur le 1er plan factoriel

## Annexes

Présentation plus détaillée des résultats + graph de l'AFM + HCPC

On a un problème avec les individus qui ont répondu 'nsp', 'rdr' ou 'comprend.pas'. C'est un comportement "inhabituel" par rapport à la norme des réponses, qui va générer la création d'axes (dans les 1ers axes) dédiés à 'nsp', 'rdr' ou 'comprend.pas' et ensuite générer des classes autour de ces réponses et brouiller la classification.

Il faut donc traiter ces réponses préalablement à l'AFM + HCPC.

2 solutions :

- soit on supprime simplement tous les individus qui ont répondu au moins une fois 'nsp', 'rdr' ou 'comprend.pas'. On enlève alors 402 répondants.
- soit on supprime seulement les individus qui ont répondu trop souvent 'nsp', 'rdr' ou 'comprend.pas', par exemple, tous ceux qui ont répondu 3 fois ou plus 'nsp', 'rdr' ou 'comprend.pas'. On enlève alors seulement 146 répondants. (Si on enlève tous ceux qui ont répondu 2 fois ou plus 'nsp', 'rdr' ou 'comprend.pas', on enlève 204 répondants.)

Ici on réalise la solution 2.

La préparation est faite !

## AFM

```
#load(file="skill_affine.Rdata")
resMFA <- MFA(skill, group=c(15,1,4,5,4,4,4), type=c("n","c","n","n","n","n","n"), ncp=43, name.group=c
# ou ncp=43 -> 70% de l'inertie initiale est conservée (pour la CAH)
#summary(resMFA)

#save(resMFA, file="resMFA_affine.Rdata")
```

## Histogramme des valeurs propres

```
load(file="resMFA_affine.Rdata")
barplot(resMFA$eig[1:45,3], main="Histogramme des valeurs propres",
        xlab="Dimension", ylab="Pourcentage d'inertie (%)", cex.axis=0.7, cex.names=0.7,
        col=c(rep("orange",43),rep("gray",2)), names.arg=1:45, ylim=c(0,80))
abline(h=70, lty=2)
text(x=0,y=72,labels="70 %",cex=0.8)

fviz_eig(resMFA, choice="variance", ncp=15, addlabels=TRUE, ylim=c(0,13), xlim=c(1,15.2))
fviz_eig(resMFA, choice="eigenvalue", ncp=20, addlabels=TRUE, ylim=c(0,4.2))
```

On garde les 43 premières composantes principales, ce qui permet de conserver 70% de l'inertie initiale du jeu de données.

Si on veut étudier les premiers axes de l'AFM, on peut regarder principalement les 4-5 premiers axes.

## Etude de l'AFM

```
# Groupes
#-----
# Représentation des groupes sur le 1er plan factoriel
plot(resMFA, choix=c("group"), shadowtext=TRUE, cex=0.8, ylim=c(0,0.6))
#fviz_mfa_var(resMFA, "group")

# Contribution des groupes aux 6 premières dimensions
p1=fviz_contrib(resMFA, "group", axes=1) + ggtitle("Dim 1")
p2=fviz_contrib(resMFA, "group", axes=2) + ggtitle("Dim 2")
p3=fviz_contrib(resMFA, "group", axes=3) + ggtitle("Dim 3")
p4=fviz_contrib(resMFA, "group", axes=4) + ggtitle("Dim 4")
p5=fviz_contrib(resMFA, "group", axes=5) + ggtitle("Dim 5")
p6=fviz_contrib(resMFA, "group", axes=6) + ggtitle("Dim 6")
grid.arrange(p1,p2,p3,p4,p5,p6, ncol=3, nrow=2)

# Axes partiels
#-----
# Représentation des axes partiels sur les axes de l'AFM
plot(resMFA, choix=c("axes"), cex=0.8, select="contrib 10", autoLab="auto", shadowtext=TRUE, legend=TRUE)
#fviz_mfa_axes(resMFA, select.axes=list(contrib=10), repel=TRUE) + theme(legend.position='none')

# # Qualité de représentation (cos2) des modalités des variables sur les axes factoriels
# var=get_mfa_var(resMFA, "quali.var")
# library(corrplot)
# corrplot(var$cos2, is.corr=FALSE)
# a=corrplot(var$contrib, is.corr=FALSE)
# corrplot(t(a))
```

La projection des groupes dans le premier plan factoriel montre une proximité importante entre les groupes “Compétences opérationnelles”, “informationnelles”, “sociales” et “mobiles”. Les compétences créatives quant à elles apparaissent relativement éloignées des 4 groupes précédemment cités. Cet éloignement traduit un rôle spécifique joué par les compétences créatives dans les profils de compétences des répondants. A l'inverse, la proximité constatée entre les 4 autres groupes traduit des profils/réponses assez proches d'un type de compétences à l'autre.

Ces 4 groupes de compétence ont des coordonnées positives élevées sur le plan factoriel 1-2 (1er quadrant) et contribuent fortement à la construction des axes 1, 2 et 3. Les compétences créatives ont une contribution moins forte sur ces axes. En revanche, elles vont très fortement contribuer aux axes 4, 5 et 6 et définir quasiment à elles seules les axes 4 et 6 (les compétences mobiles sont les principales contributrices de l'axe 5).

La projection des axes partiels sur le premier plan de l'AFM montre que les dimensions 1 (resp. 2) des groupes “Compétences opérationnelles”, “mobiles”, “sociales” et “informationnelles” sont très corrélées positivement à la première (resp. deuxième) dimension de l'AFM. Tandis que la dimension 1 des compétences créatives est assez fortement corrélée négativement à la dimension 1 de l'AFM et que la dimension 2 de ces compétences est moyennement corrélée positivement à la dimension 2 de l'AFM. Ce qui confirme le rôle particulier joué par ce type de compétences dans la compétence numérique globale.

```
# Individus
#-----
```

```
# Représentation des individus
plot(resMFA, choix=c("ind"), invisible=("quali"), lab.ind=FALSE)
# Représentation des individus colorés en fonction de leur cos2
fviz_mfa_ind(resMFA, col.ind="cos2", gradient.cols=c("#00AFBB", "#E7B800", "#FC4E07"), geom="point", label=
```

La projection des individus et des variables illustratives sur le premier plan factoriel laisse entrevoir une opposition le long de l'axe 1 entre les “jeunes, étudiants, très à l'aise avec Internet” situés sur la gauche du graphique et les “personnes âgées, retraités, pas à l'aise avec Internet voire non-internautes” situés sur la droite du graphique.

L'axe 1 présente également un gradient de compétences, des individus les plus compétents à gauche du graphique vers les individus les moins compétents à droite.

L'axe 2 oppose les individus extrêmes (très compétents ou très peu compétents) à gauche aux individus moyens (plutôt ou plutôt pas compétents) à droite.

L'axe 3 oppose les plutôt compétents (à gauche) aux plutôt pas compétents (à droite).

L'axe 4 oppose les très compétents créatifs (en haut et en particulier pour les compétences “savoir créer un site web” et “savoir programmer dans un langage informatique”) aux moins compétents (en bas).

```
# Modalités
#-----
# Représentation des modalités
fviz_mfa_var(resMFA, "quali.var", axes=c(1,2), col.var="contrib", gradient.cols=c("#abd7d8", "#E7B800", "#FC4E07"), geom="point", label="contrib")
fviz_mfa_var(resMFA, "quali.var", axes=c(2,3), col.var="contrib", gradient.cols=c("#abd7d8", "#E7B800", "#FC4E07"), geom="point", label="contrib")
fviz_mfa_var(resMFA, "quali.var", axes=c(3,4), col.var="contrib", gradient.cols=c("#abd7d8", "#E7B800", "#FC4E07"), geom="point", label="contrib")

# Contributions des modalités aux 4 premières dimensions
p1=fviz_contrib(resMFA, choice="quali.var", axes=1, top=20, palette="jco") + ggtitle("Dim 1")
p2=fviz_contrib(resMFA, choice="quali.var", axes=2, top=20, palette="jco") + ggtitle("Dim 2")
p3=fviz_contrib(resMFA, choice="quali.var", axes=3, top=20, palette="jco") + ggtitle("Dim 3")
p4=fviz_contrib(resMFA, choice="quali.var", axes=4, top=20, palette="jco") + ggtitle("Dim 4")
grid.arrange(p1,p2,p3,p4, ncol=2, nrow=2)

# Cos2 des modalités sur les 4 premières dimensions
p1=fviz_cos2(resMFA, choice="quali.var", axes=1, top=20, palette="jco") + ggtitle("Dim 1")
p2=fviz_cos2(resMFA, choice="quali.var", axes=2, top=20, palette="jco") + ggtitle("Dim 2")
p3=fviz_cos2(resMFA, choice="quali.var", axes=3, top=20, palette="jco") + ggtitle("Dim 3")
p4=fviz_cos2(resMFA, choice="quali.var", axes=4, top=20, palette="jco") + ggtitle("Dim 4")
grid.arrange(p1,p2,p3,p4, ncol=2, nrow=2)
```

Une analyse plus détaillée des contributions des modalités des variables actives de l'AFM à la construction des premiers axes montre que :

- Axe 1
  - contribution positive : les modalités “Cela ne vous ressemble pas du tout” de nombreuses variables parmi lesquelles les 4 compétences opérationnelles, crea3, les 4 sociales, les 4 informationnelles, mob3, mob4, mob1, etc.
  - contribution négative : les modalités “Cela vous ressemble beaucoup” des variables crea3, soc3, mob1, etc.
- Axe 2
  - contribution positive : les modalités “Cela vous ressemble plutôt bien” de nombreuses variables parmi lesquelles les 4 compétences opérationnelles, les 4 mobiles, les 4 sociales, inf4, inf1, inf2, etc.
  - contribution négative : les modalités “Cela vous ressemble beaucoup” ou “Cela ne vous ressemble pas du tout” des variables crea3, crea1, soc1, inf1, etc.
- Axe 3

- contribution positive : les modalités “Cela vous ressemble assez peu” de nombreuses variables parmi lesquelles les 4 compétences opérationnelles, soc2, inf1, inf2, soc3, mob3, etc.
- contribution négative : les modalités “Cela vous ressemble plutôt bien” des variables mob2, crea3, mob1, crea1, soc3, etc.
- Axe 4
  - contribution positive : les modalités “Cela vous ressemble beaucoup” des variables de compétences créatives.
  - contribution négative : les modalités “Cela vous ressemble plutôt bien” et “Cela vous ressemble assez peu” des variables de compétences créatives, soc1, ope3, ope1, mob2, mob1, etc.

## Classification ascendante hiérarchique sur le résultat de l’AFM

```
#resHCPC=HCPC(resMFA)
nbClasse=5
# partition en 5 ok un peu global -> sans doute la plus optimale / claire
# 6 ok (distingue les crea++ en cl1, notamment crea4++ coder),
# 7 ok commence à être fin (distingue les - et les -- chez les internautes),
# 8 trop fin (sépare les crea4++ et +), 9 trop fin (même séparation que pour 8 au niveau des crea)
resHCPC=HCPC(resMFA, nb.clust=nbClasse, graph=FALSE)
# resHCPC=HCPC(resMFA,nb.clust=9, graph=FALSE)

# Dendrogramme
plot(resHCPC, choice="tree", labels=FALSE)
# Nb de classes
plot(resHCPC, choice="bar")

skillClust=resHCPC$data.clust
skillClust$poids=as.numeric(as.character(skillClust$poids))
# Répartition par classe
freq(wtd.table(skillClust$clust, weights=skillClust$poids))
#catdes(skillClust,38)

## Description des classes
#
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`1`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`2`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`3`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`4`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`5`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`6`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`7`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`8`,2),45)
# head(round(catdes(skillClust,38,proba=0.01,row.w=skillClust$poids)$category$`9`,2),45)
```

Le dendrogramme laisse deviner 5 classes naturelles. Ce choix est confirmé par le saut d’inertie présent entre le passage de 4 à 5 classes et le passage de 5 à 6 classes.

Les classes obtenues présentent des effectifs allant de 215 répondants (soit 11% de la population) pour la plus petite à 670 répondants (soit 36%) pour la plus grande.

```
# Représentation des classes sur les premiers plans factoriels

# plot(resHCPC, choice="map", ind.names=FALSE, draw.tree=FALSE, axes=c(1,2))
```

```

# plot(resHCPC, choice="map", ind.names=FALSE, draw.tree=FALSE, axes=c(2,3))
# plot(resHCPC, choice="map", ind.names=FALSE, draw.tree=FALSE, axes=c(3,4))
# plot(resHCPC, choice="map", ind.names=FALSE, draw.tree=FALSE, axes=c(4,5))

fviz_cluster(resHCPC, repel=FALSE, shape=16, ellipse=TRUE, geom="point",
  ggtheme=theme_minimal(), main="Factor map") +
  geom_vline(xintercept=0, linetype="dotted") + geom_hline(yintercept=0, linetype="dotted")
fviz_cluster(resHCPC, repel=FALSE, shape=16, ellipse=TRUE, geom="point",
  ggtheme=theme_minimal(), main="Factor map", axes=c(2,3)) +
  geom_vline(xintercept=0, linetype="dotted") + geom_hline(yintercept=0, linetype="dotted")
fviz_cluster(resHCPC, repel=FALSE, shape=16, ellipse=TRUE, geom="point",
  ggtheme=theme_minimal(), main="Factor map", axes=c(3,4)) +
  geom_vline(xintercept=0, linetype="dotted") + geom_hline(yintercept=0, linetype="dotted")

# library(tibble)
# a1=rownames_to_column(skillClust)
# a2=rownames_to_column(data[,c("WIP.outil_4", "WIP.outil_5")])
# a=left_join(a1,a2,by="rowname")
# round(prop.table(table(a$clust,a$WIP.outil_5),1)*100)

```

La projection des classes sur le plan 1-2 laisse apparaître des classes plutôt bien séparées, laissant augurer des profils de compétences distincts.

## Description des classes (signif à 1%)

### Classe 1

La classe 1 est la deuxième classe la plus importante en taille et contient environ 20% des répondants. Elle regroupe les individus se déclarant très à l'aise avec l'ensemble des compétences abordées, quel que soit le type de compétences, même créatives. En effet, 90% d'entre eux déclarent savoir comment créer quelque chose de nouveau à partir d'images, de musiques ou de vidéos, et 96% savent mettre en ligne du contenu. Cette classe rassemble 97% des répondants ayant déclaré savoir créer un site web et 93% de ceux déclarant savoir coder/programmer dans un langage informatique. Ces personnes sont internautes (100%). Parmi elles, les jeunes (18-24 ans 26%, 25-34 ans 26%) sont sur-représentés, tout comme les étudiants, les cadres et les artisans, les hommes, les personnes ayant fait des études supérieures courtes et longues, les célibataires, les CSP+, les personnes déclarant avoir une vie très confortable, vivant à 3 ou 4 personnes par foyer, et 5 ou plus, écrit et actu fréquent, ancienneté sur Internet  $\geq 17$  ans.

### Classe 2

La classe 2 est la classe la plus grande avec 36% des répondants. La classe 2 regroupe les individus déclarant également de très bonnes compétences opérationnelles, informationnelles, sociales et mobiles, mais contrairement aux individus de la classe 1 ils sont nettement moins à l'aise avec les compétences créatives et notamment pour ce qui a trait à la création de site et à la programmation. En effet, 82% (resp. 90%) d'entre eux disent que savoir créer un site web (resp. programmer) ne leur ressemble pas du tout. Prof intermédiaire, employés, supérieur court et long, 25-34, et dans une moindre mesure 18-24 et 35-49 ans, 3-4 personnes dans le foyer, CSP+ (et -), revenu  $>4000\text{€}$ , ancienneté Internet  $\geq 17$  ans et 12-16 ans, très à l'aise avec Internet, internaute (100%), pacsé, célibataire, concubinage, écrit assez fréquent.

### **Classe 3**

La classe 3 est la plus petite avec environ 11% des répondants. C'est la classe des individus ayant de plutôt bonnes compétences dans tous les types abordés, même s'ils sont globalement moins à l'aise dans les compétences créatives. 35-49 ans, employé, à l'aise avec Internet, ancienneté 9-12 ans, internautes (99%), écrit assez fréquent, étude niveau bac.

### **Classe 4**

La classe 4 contient environ 16% de la population. C'est la classe des individus ayant des compétences faibles ou très faibles dans la plupart des compétences abordées, et notamment mobiles et créatives. Ils se sentent néanmoins assez à l'aise sur quelques points comme la recherche d'informations en ligne, l'utilisation de bons mots-clés, rechercher des itinéraires en ligne, ouvrir un fichier téléchargé, envoyer un email et y joindre un document, se connecter à un réseau wifi. Bref, quelques usages assez basiques. 95% de ces individus sont internautes. 50-64 ans et  $\geq 65$  ans, retraités, assez peu à l'aise avec Internet, CAP-BEP, mariés, 2 personnes dans le foyer, ancienneté  $< 9$  ans

### **Classe 5**

La classe 5 contient environ 17% des répondants. Ces individus ont des compétences très faibles sur l'ensemble des types de compétences (plus encore que la classe 4). 75% ne sont pas internautes (grosse différence par rapport à la classe 4).  $\geq 65$  ans, 50-64 ans, retraités, veuf, séparé, 1p dans le foyer ou 2, revenu  $< 1400\text{€}$ , vie difficile et très difficile, ancienneté  $< 9$  ans, pas du tout à l'aise avec Internet, agriculteurs, étude type collègue, CAP-BEP, écrit très rare