

Analyse des données RéFérens III

Antoine Blanchard

24/02/2021

Import du référentiel métier RéFérens

RéFérens (REFérentiel des Emplois-types de la recherche et de l'Enseignement Supérieur) version III est fourni par le Ministère de l'enseignement supérieur, de la recherche et de l'innovation, sur leur portail open data. On l'importe et on liste l'ensemble de ses variables (colonnes) :

```
data <- read.csv2("fr-esr_referentiel_metier_referens_3.csv") %>%  
  mutate(Id = row_number())  
names(data) #noms de colonnes
```

```
## [1] "referens_bap_id"  
## [2] "referens_bap"  
## [3] "referens_fap"  
## [4] "referens_fap_rattachement"  
## [5] "referens_cs_id"  
## [6] "referens_cs"  
## [7] "referens_cs_lib"  
## [8] "referens_domaine_formation"  
## [9] "referens_metiers"  
## [10] "referens_mission"  
## [11] "referens_fap_reme"  
## [12] "referens_activites_principales"  
## [13] "referens_conditions"  
## [14] "referens_connaissances"  
## [15] "referens_competences_operationnelles"  
## [16] "referens_competences_comportementales"  
## [17] "referens_prerequis"  
## [18] "referens_facteurs_d_evolution_moyen_terme"  
## [19] "referens_impacts"  
## [20] "referens_id"  
## [21] "referens_intitule"  
## [22] "referens_id_precedent"  
## [23] "referens_intitule_precedent"  
## [24] "tricarto"  
## [25] "referens_fap_id"  
## [26] "referens_experience"  
## [27] "trifap"  
## [28] "referens_categorie_emploi"  
## [29] "referens_type_id_precedent"  
## [30] "referens_type_precedent"  
## [31] "referens_programme_du_concours_externe_fichier"  
## [32] "referens_programme_du_concours_externe"  
## [33] "referens_programme_du_concours_externe_classe_normale"
```

```
## [34] "referens_programme_du_concours_externe_classe_superieure"
## [35] "fiche_url"
## [36] "fiche_pdf"
## [37] "referens_connaissances_definitions"
## [38] "referens_competences_operationnelles_definitions"
## [39] "referens_competences_comportementales_definitions"
## [40] "competences_id"
## [41] "bap"
## [42] "fap"
## [43] "Id"
```

```
variables <- data %>%
  select(-Id) %>%
  colnames()
```

Il comporte 242 lignes ou fiches métiers, conformément à ce qui est annoncé sur le portail data.gouv.fr :

Ce jeu de données reprend les informations et nomenclatures concernant les 242 métiers des ingénieurs et personnels techniques de recherche et de formation (ITRF) et des ingénieurs et personnels techniques de la recherche (ITA) décrits dans le répertoire des branches d'activités professionnelles et des emplois-types, dénommé RéFérens (REférentiel des Emplois-types de la recherche et de l'Enseignement Supérieur).

Transformation des données

On définit les chaînes de caractère qui seront comptées dans le document comme des occurrences du terme “données” :

```
donnees <- "Donnees|donnees|Données|données|Donnee|donnee|Donnée|donnée"
```

On dénombre, pour chaque cellule, les occurrences du terme “données” et on les ajoute aux données brutes pour former le tableau **data_filtered** :

```
denombrement <- data %>%
  column_to_rownames("Id") %>%
  #select(Nom des colonnes, séparées par une virgule) %>%
  mutate_all(str_count, donnees) %>%
  mutate(somme = rowSums(.)) %>%
  rownames_to_column("Id") %>%
  filter(somme > 0) %>%
  arrange(desc(somme))

data_filtered <- data %>%
  filter(Id %in% denombrement$Id) %>%
  mutate(Id = as.character(Id)) %>%
  left_join(denombrement %>% select(Id, somme), by = "Id")
```

Analyse des données

Fréquence de l'expression “données” par champ

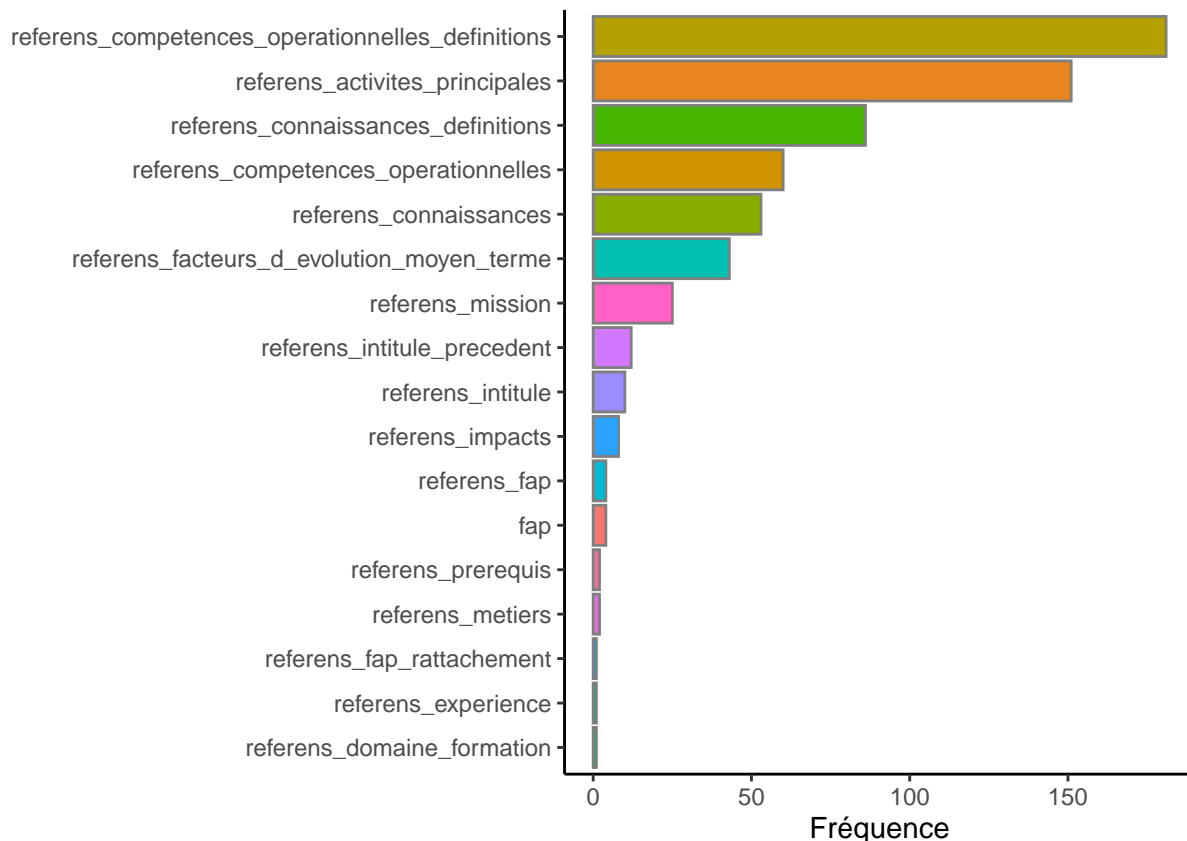
Dans quelles colonnes du jeu de données trouve-t-on le plus le terme “données” ?

```
denombrement %>%
  select(-somme) %>%
  column_to_rownames("Id") %>%
  summarise_all(sum) %>%
```

```

rownames_to_column("Id") %>%
gather("variable", "somme", -Id) %>%
filter(somme > 0) %>%
ggplot(aes(x=fct_reorder(variable, somme), y = somme)) +
geom_col(aes(fill = variable), color = "grey50") +
labs(x = "", y = "Fréquence")+
coord_flip()+
theme_classic() +
theme(legend.position = "none")

```



Comme on pouvait s’y attendre, le terme données apparaît le plus dans les champs des fiches métiers qui décrivent les compétences opérationnelles (rang 1) et les activités principales (rang 2). Il apparaît assez peu dans l’intitulé des fiches métiers (voir plus loin pour une analyse détaillée), qui est au 9e rang, précédé du champ “intitulé précédent” qui mérite d’être creusé pour comprendre pourquoi certaines occurrences ont disparu. Au 6e rang, donc assez haut, on trouve le champ “Factures d’évolution à moyen terme” du métier qui vise, selon la documentation de RéFérens, à identifier les facteurs clés et en déduire leur impact qualitatif sur le métier.

Fréquence de l’expression “données” par BAP

Proportion des fiches métiers qui contiennent l’expression “données” par BAP :

```

BAP_filtered <- data_filtered %>%
  count(bap) %>%
  arrange(bap)

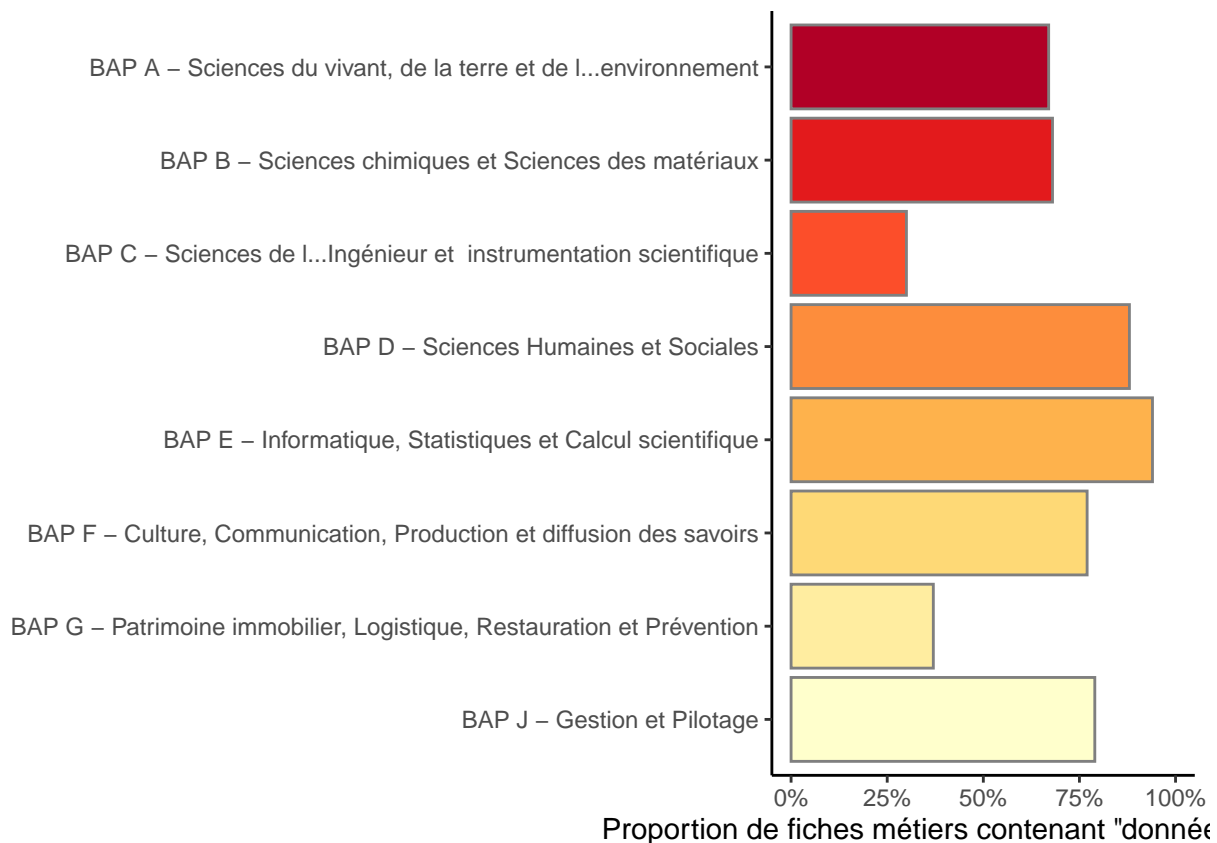
BAP <- data %>%

```

```
count(bap)

BAP_filtered_pourcentage <- BAP_filtered %>%
  left_join(BAP, BAP_filtered, by = "bap") %>%
  mutate(pourcentage_bap=round(n.x/n.y,2),
         bap=fct_reorder(bap, bap, .desc = TRUE))

BAP_filtered_pourcentage %>%
  ggplot(aes(x = bap, y = pourcentage_bap)) +
  geom_col(aes(fill = bap), color = "grey50") +
  labs(x = "", y = "Proportion de fiches métiers contenant 'données' par BAP") +
  coord_flip() +
  theme_classic() +
  theme(legend.position = "none") +
  scale_fill_manual(values = brewer.pal(8, "YlOrRd")) +
  scale_y_continuous(labels = scales::percent, limits = c(0,1))
```



On constate que les BAP E et D possèdent en proportion la plus grande part de métiers liés aux données : c'est intéressant de voir le soutien à la recherche en SHS si haut. Les BAP A, B et C, qui concernent directement le soutien à la recherche, sont moins concernés par les données que les BAP F et J qui sont liés à l'administration de l'université et se retrouvent plutôt dans l'administration centrale.

Fréquence de l'expression "données" par corps

Proportion des fiches métiers qui contiennent l'expression "données" par corps :

```

# Classer les corps dans l'ordre hiérarchique
ordre_corps <- c("ATRF", "TECH", "AI", "IE", "IR")

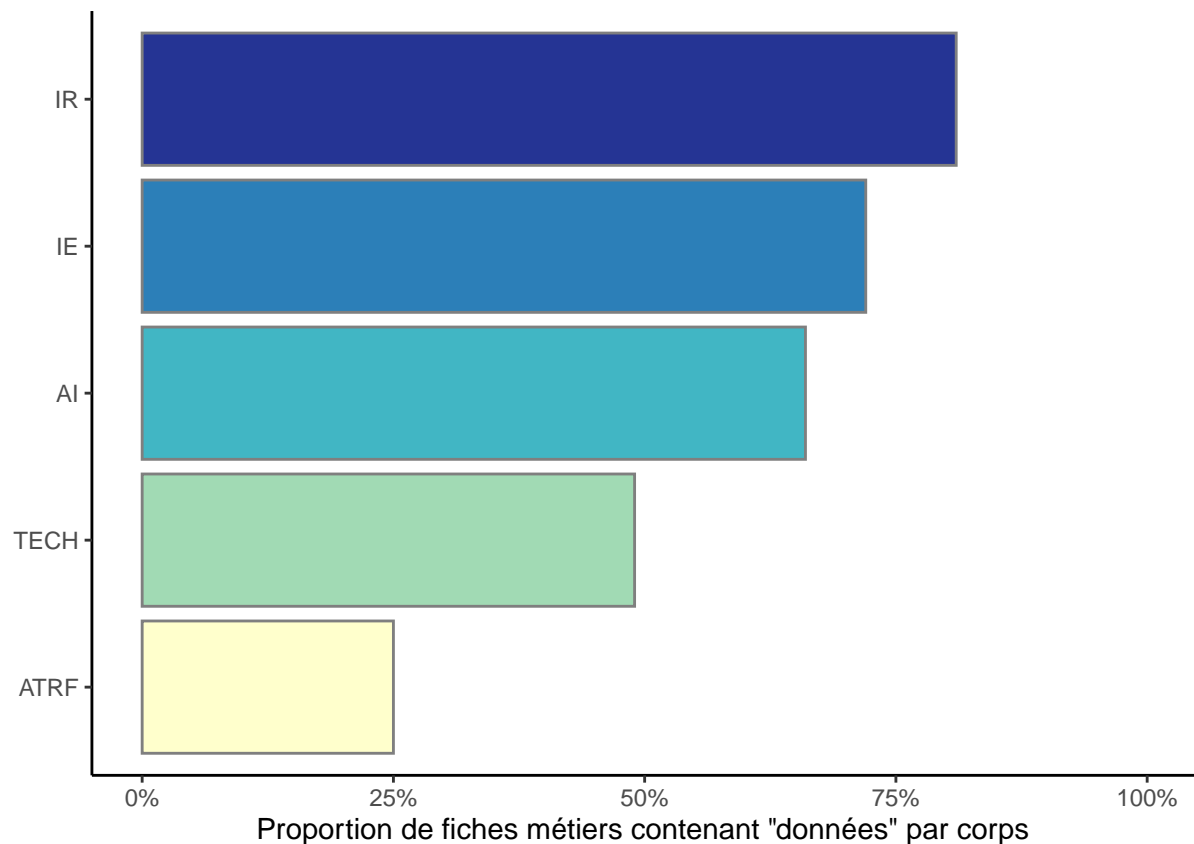
corps_filtered <- data_filtered %>%
  count(referens_cs) %>%
  arrange(desc(n))

corps <- data %>%
  count(referens_cs)

corps_filtered_pourcentage <- corps_filtered %>%
  left_join(corps, corps_filtered, by = "referens_cs") %>%
  mutate(pourcentage_corps=round(n.x/n.y,2),
         referens_cs=factor(referens_cs, levels = ordre_corps))

corps_filtered_pourcentage %>%
  ggplot(aes(x = referens_cs, y = pourcentage_corps)) +
  geom_col(aes(fill = referens_cs), color = "grey50") +
  labs(x = "", y = "Proportion de fiches métiers contenant 'données' par corps") +
  coord_flip() +
  theme_classic() +
  theme(legend.position = "none") +
  scale_fill_manual(values = brewer.pal(5, "YlGnBu")) +
  scale_y_continuous(labels = scales::percent, limits = c(0,1))

```



On constate que la proportion de métiers liés aux données varie en même temps que l'ordre hiérarchique : elle est plus grande pour les corps ayant le plus de missions d'encadrement.

Liste des 11 métiers dont l'intitulé contient l'expression "données"

```
intitule_TRUE <- denombrement %>%
  filter(denombrement$referens_intitule >0 | denombrement$referens_metiers >0)

intitule_filtered <- data_filtered %>%
  filter(Id %in% intitule_TRUE$Id) %>%
  mutate(Id = as.character(Id),
         referens_cs=factor(referens_cs, levels = ordre_corps)) %>%
  left_join(data_filtered %>% select(Id, somme), by = "Id") %>%
  select(referens_intitule, referens_metiers, bap, referens_fap, referens_cs) %>%
  group_by(bap) %>%
  arrange(referens_cs, .by_group = TRUE)
```

Table 1: Fiches métiers faisant apparaître le terme 'données' dans leur intitulé ou équivalent

referens_intitule	referens_metiers	bap	referens_fap	referens_cs
Assistant-e ingénieur-e biologiste en traitement de données		BAP A - Sciences du vivant, de la terre et de l'environnement	Biologie et santé, Sciences de la vie et de la terre	AI
Ingénieur-e biologiste en traitement de données	épidémiologiste	BAP A - Sciences du vivant, de la terre et de l'environnement	Biologie et santé, Sciences de la vie et de la terre	IE
Ingénieur-e biologiste en analyse de données	informaticien	BAP A - Sciences du vivant, de la terre et de l'environnement	Biologie et santé, Sciences de la vie et de la terre	IR
Technicien-ne en production, traitement de données et enquêtes		BAP D - Sciences Humaines et Sociales	Production, traitement et analyse des données	TECH
Assistant-e ingénieur-e en production, traitement de données et enquêtes		BAP D - Sciences Humaines et Sociales	Production, traitement et analyse des données	AI
Ingénieur-e d'études en production, traitement, analyse de données et enquêtes		BAP D - Sciences Humaines et Sociales	Production, traitement et analyse des données	IE
Ingénieur-e de recherche en production, traitement et analyse de données		BAP D - Sciences Humaines et Sociales	Production, traitement et analyse des données	IR
Administrateur-trice de bases de données / Intégrateur d'applications	administrateur de bases de données (dba)	BAP E - Informatique, Statistiques et Calcul scientifique	Ingénierie technique et de production	IE
Chargé-e de l'édition de corpus numériques	éditeur de données scientifiques	BAP F - Culture, Communication, Production et diffusion des savoirs	Édition et graphisme	IE
Chargé-e du traitement des données scientifiques		BAP F - Culture, Communication, Production et diffusion des savoirs	Information scientifique et technique, documentation et collections patrimoniales	IE

referens_intitule	referens_metiers bap	referens_fap	referens_cs
Gestionnaire de données et indicateurs patrimoniaux	BAP G - Patrimoine immobilier, Logistique, Restauration et Prévention	Patrimoine immobilier	AI

TO DO

Regarder les métiers qui concentrent le plus le terme “données”

Regarder les métiers où “données” apparaît dans facteur_d_evolution_moyen_terme

Regarder les métiers où le terme données apparaît comme “intitulé précédent” et pas dans l’intitulé actuel

Afficher le terme “données” en kwic (keyword-in-context) et représenter son voisinage lexical