

ANALYSE DATA-VISUALISATION

Transparence Santé



?



?



?



?



?

?

?

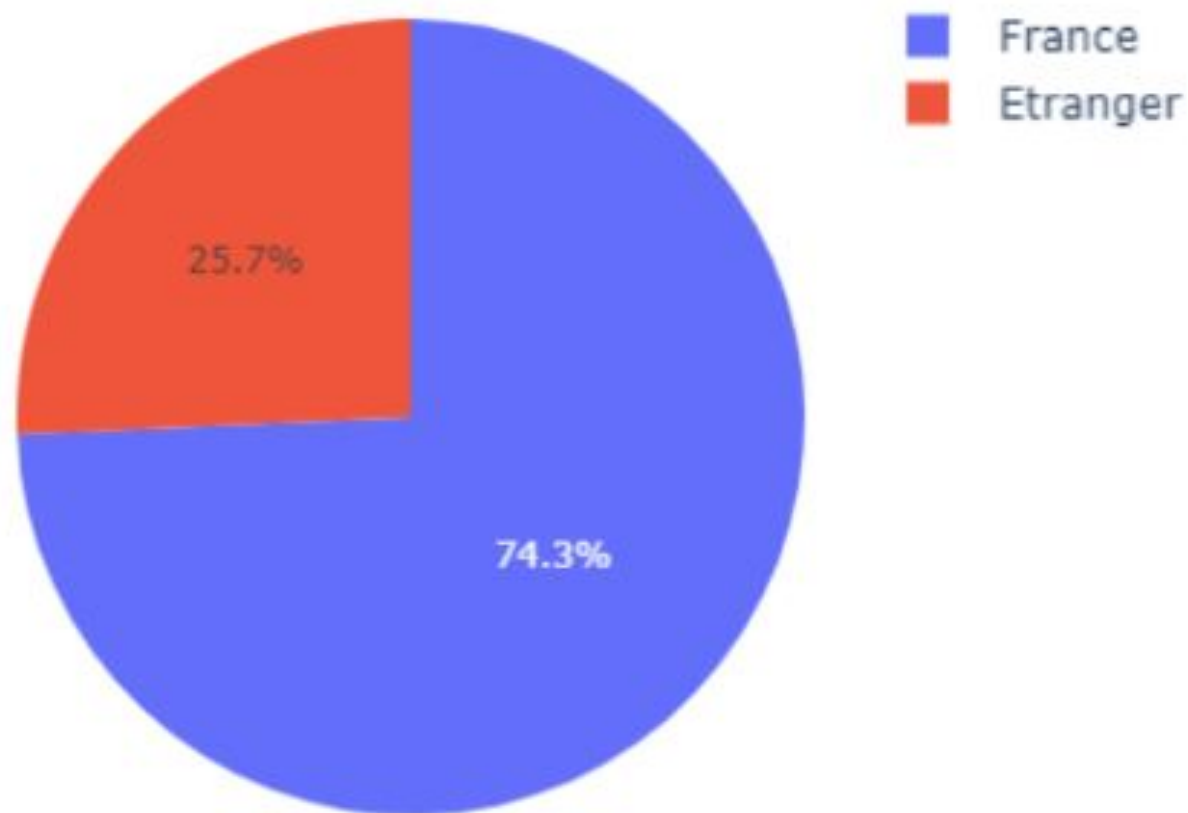
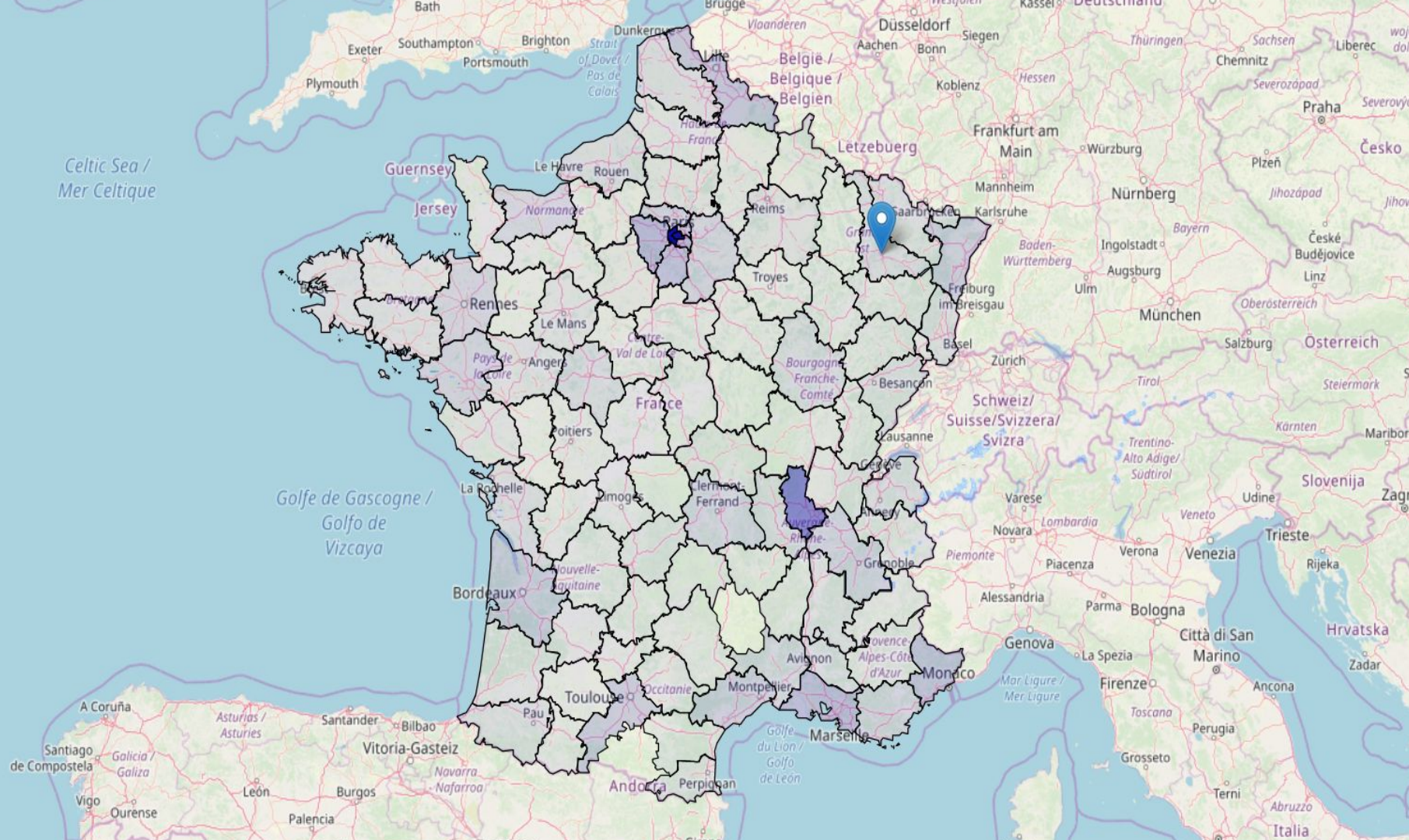


Figure 1 - Répartition des entreprises françaises et étrangères.



CHOIX DES LANGAGES



R

- Chargement du jeu de données en raison de la taille des BD
- Capacités de traiter des données volumineuses
- Commandes simplifiées dans une étude statistique

Python

- Langage le plus travaillé depuis le début de la formation
- Représentation des données

S

Q

I

- Étudier plus précisément les intérêts des langages R et Python.
- Observer la correspondance des résultats obtenus.

- Capacités à effectuer des nettoyages du jeu de données.
 1. La recherche et suppression des lignes pluri-présentes.
 2. Déceler l'existence de potentielles données aberrantes.
 3. Une étude indépendante des colonnes pour cerner leurs intérêts.
(rôle, catégories de valeurs)
 4. Possibilité de modifier des valeurs identiques saisies différemment.
Exemples: “ 54000” transformé en “54000”,
“101 avenue Anatole France” ayant différentes écritures)

ANALYSE DES DONNÉES

Statistique




```
> summary(avantage$avant_montant_ttc)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.0	23.0	40.0	143.9	60.0	3000000.0

```
> summary(convention$conv_montant_ttc)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
-16707	25	57	1880	299	33655638	4181071

```
> summary(remuneration$remu_montant_ttc)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10	120	600	5338	1800	4843939

Figure 1 - Répartition des quartiles sur les montants TTC.

```
> df_status(remuneration)
```

	variable	q_zeros	p_zeros	q_na	p_na	q_inf	p_inf	type	unique
1	<u>entreprise_identifiant</u>	0	0.00	0	0.00	0	0	factor	828
...									
8	benef_prenom	4	0.00	0	0.00	0	0	factor	9358
9	<u>benef_qualite_code</u>	0	0.00	0	0.00	0	0	factor	24
10	qualite	0	0.00	0	0.00	0	0	factor	24
11	<u>benef_adresse1</u>	9	0.00	6	0.00	0	0	factor	71105
12	<u>benef_adresse2</u>	262	0.06	4	0.00	0	0	factor	23489
13	<u>benef_adresse3</u>	398	0.09	0	0.00	0	0	factor	9521
14	<u>benef_adresse4</u>	292	0.06	0	0.00	0	0	factor	1446
...									
34	<u>remu_convention_liee</u>	0	0.00	0	0.00	0	0	factor	350244

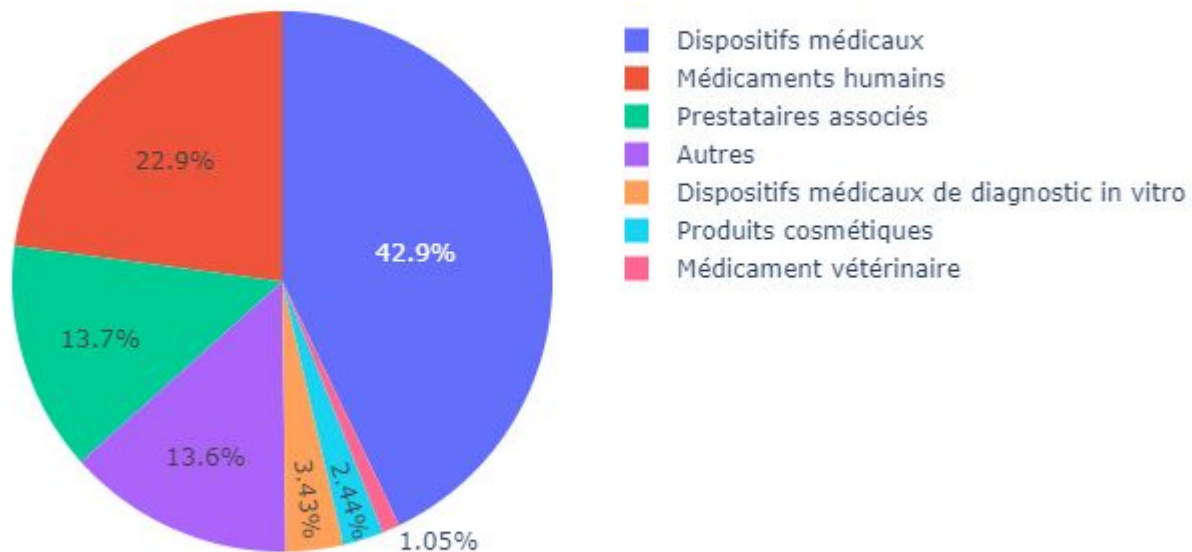
Figure 2 - Quantité des valeurs uniques prises par les variables considérées.

ANALYSE DES DONNÉES

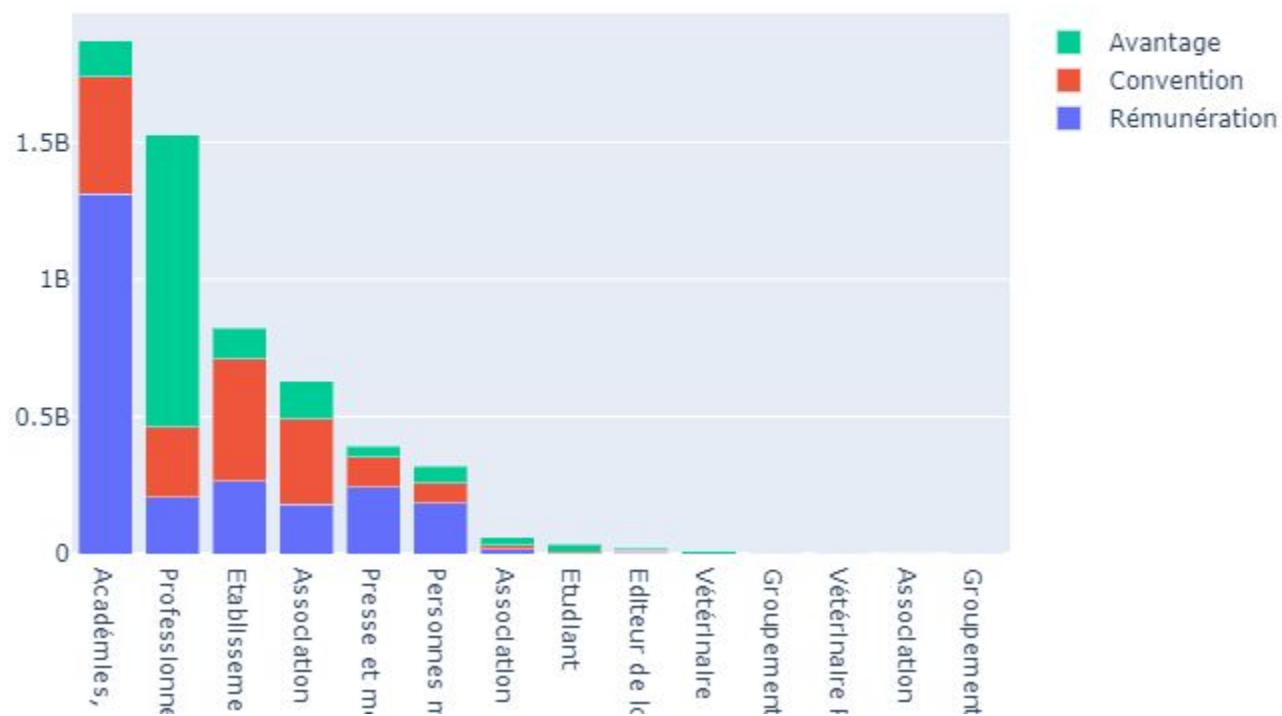
Associée aux entreprises



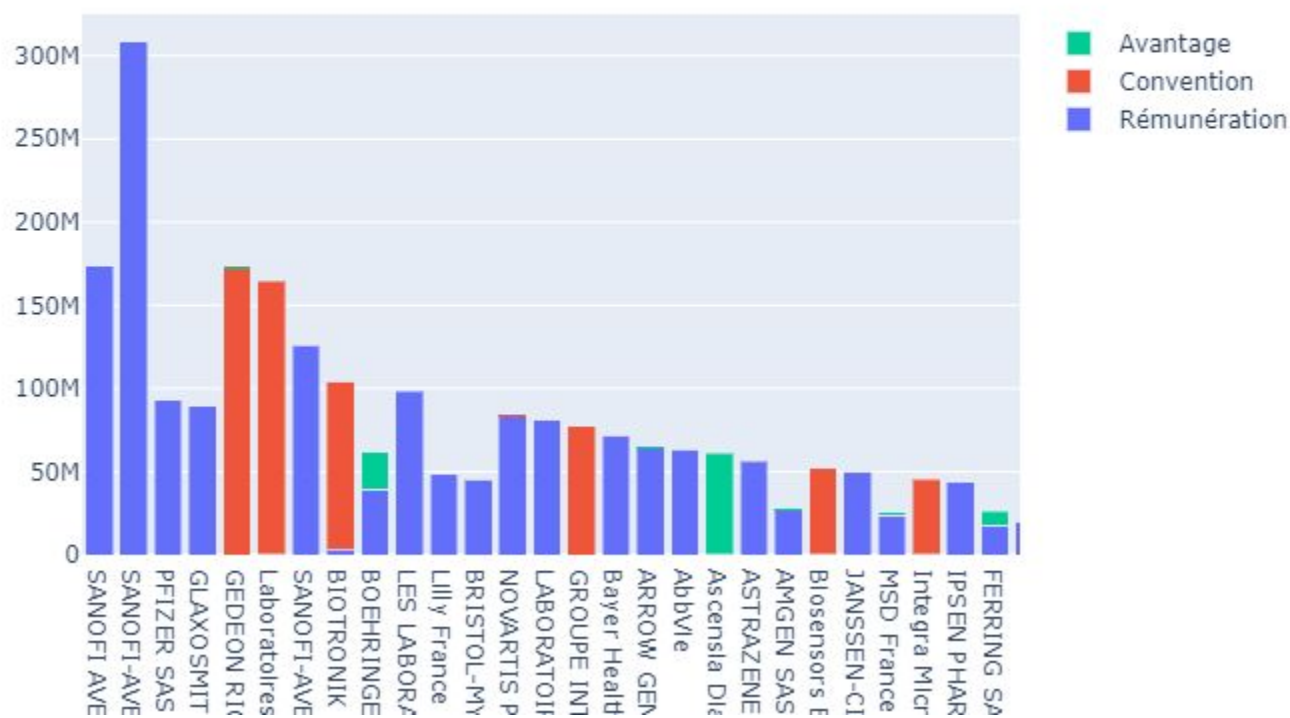
Nombre d'entreprise par secteurs



Montant investi par catégorie d'entreprise

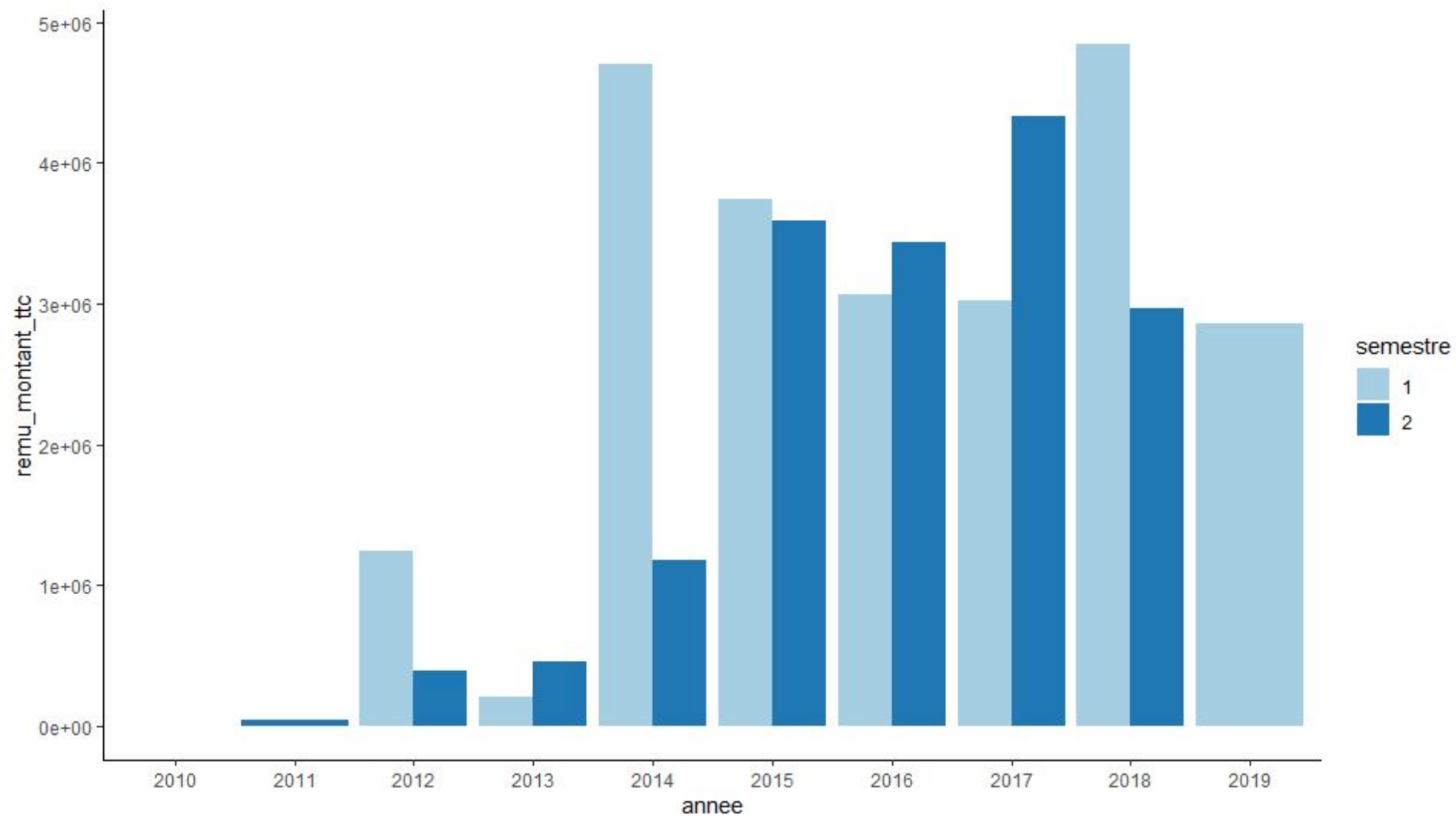


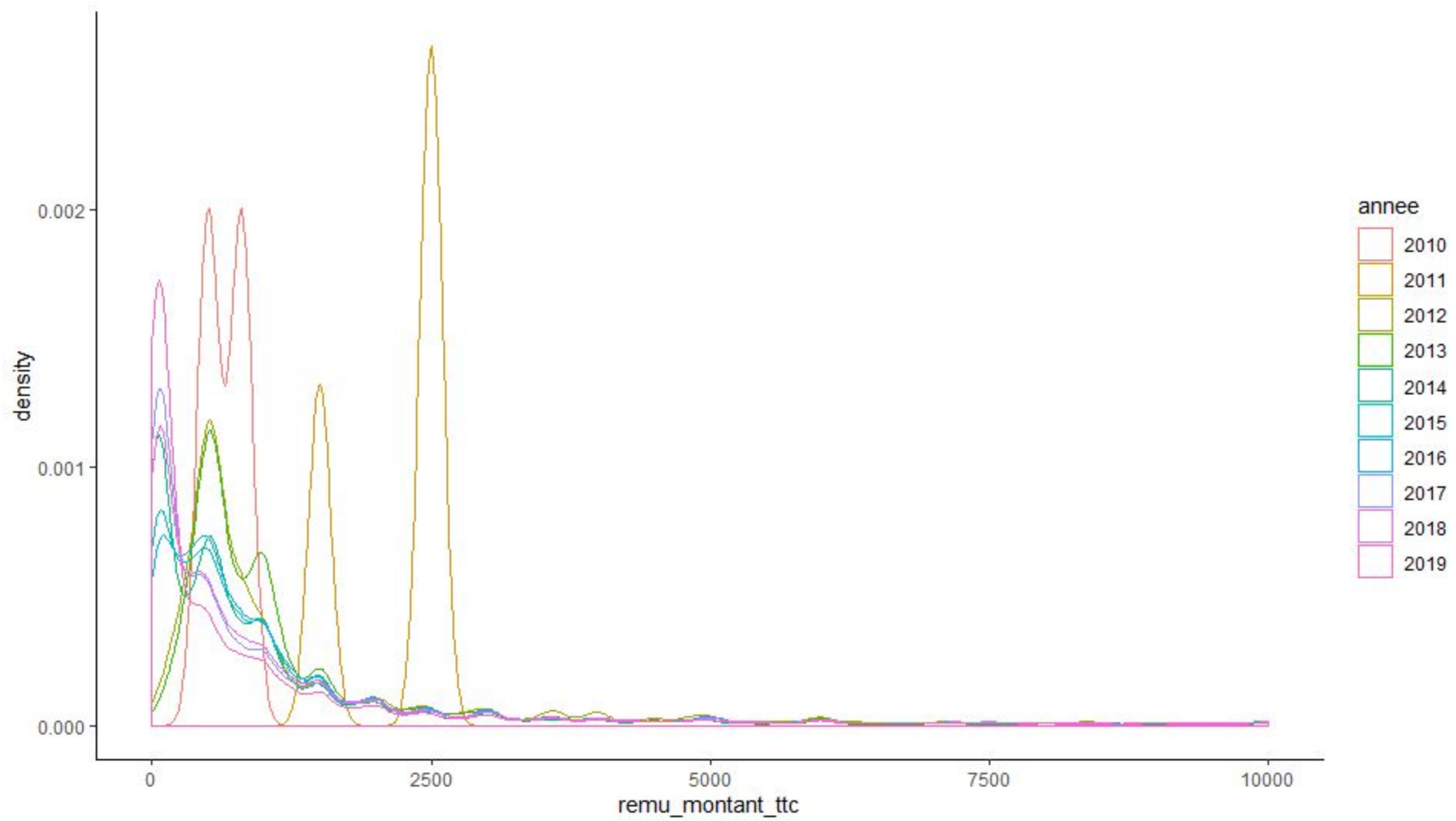
Investissement total par entreprise



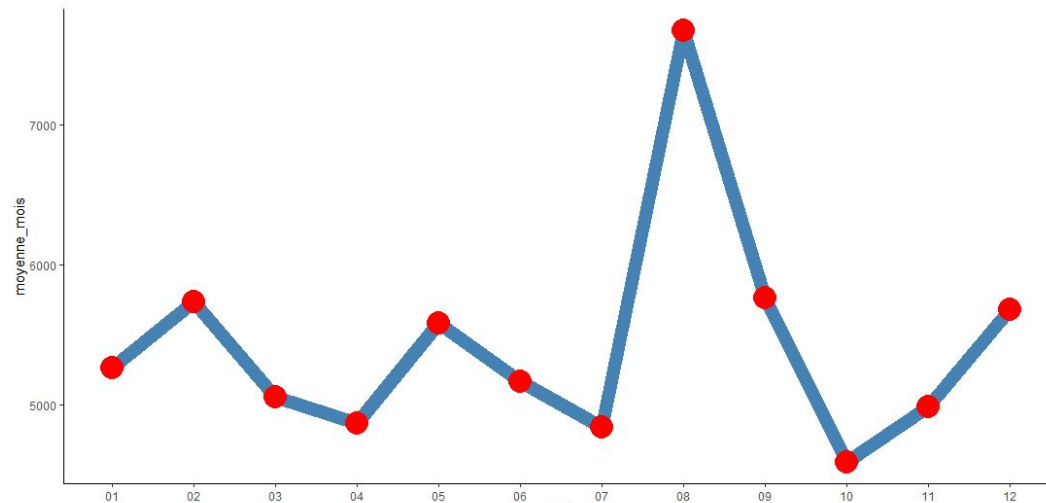
ANALYSE TEMPORELLE

de la rémunération

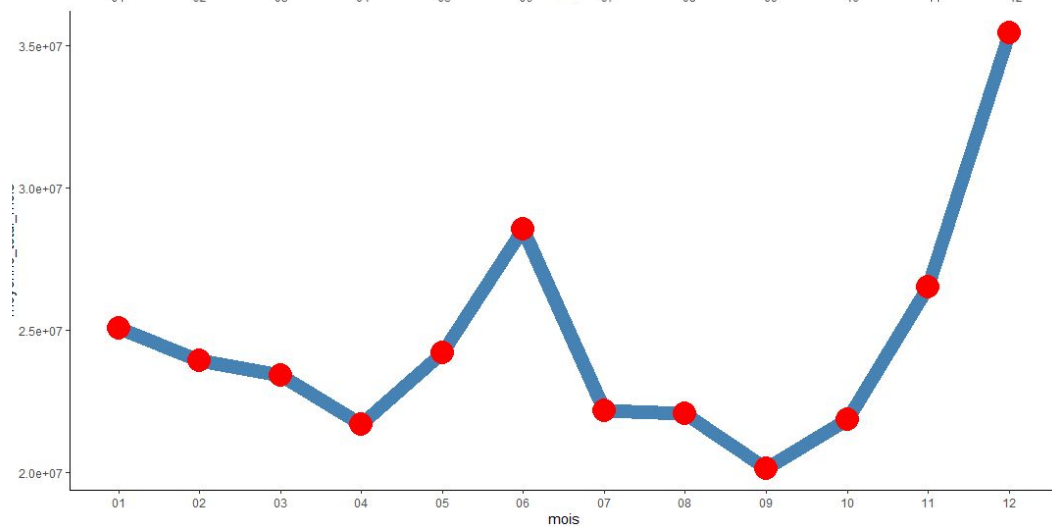




Moyenne des rémunérations
par mois

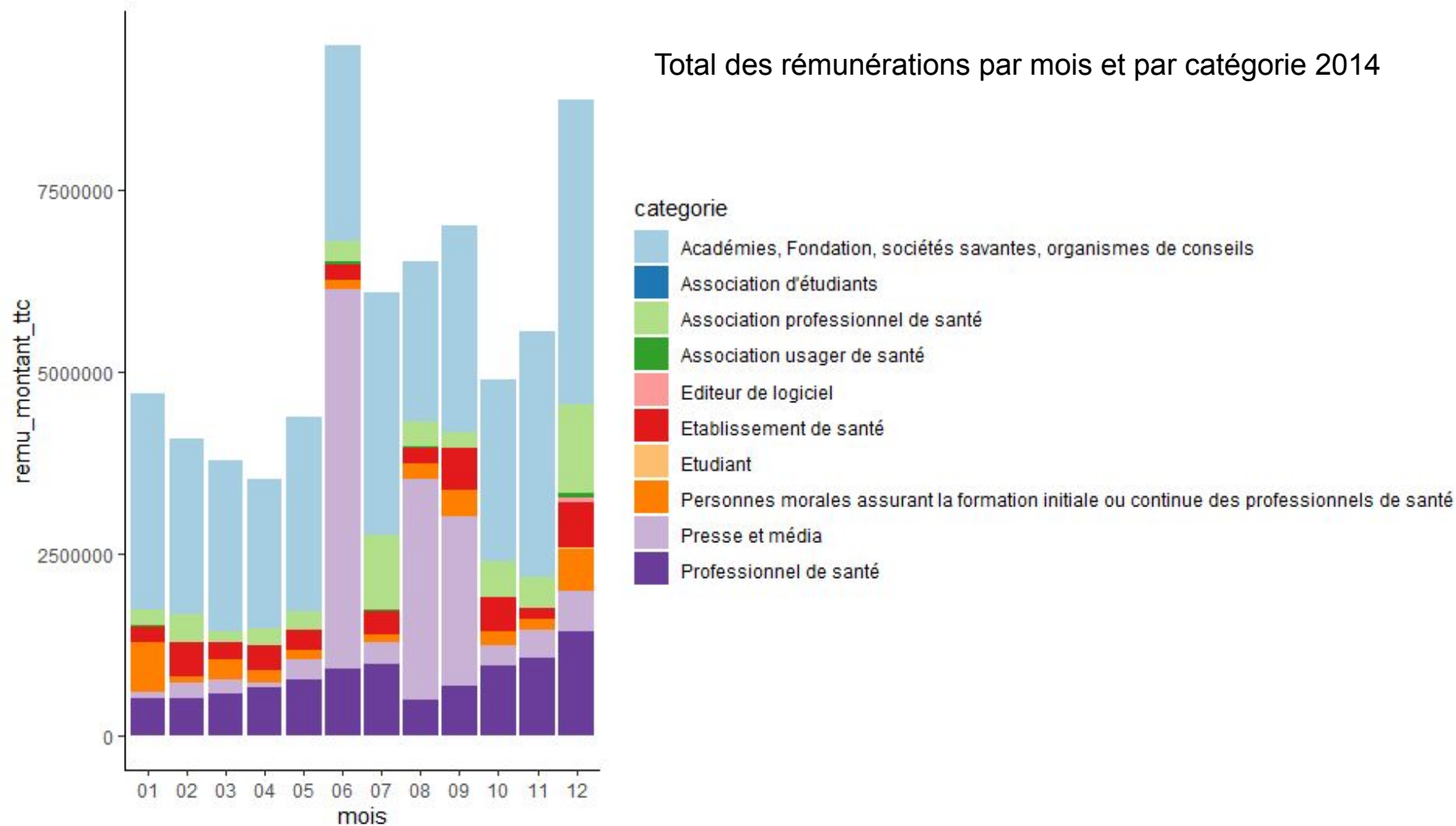


Moyenne des rémunérations
totales par mois

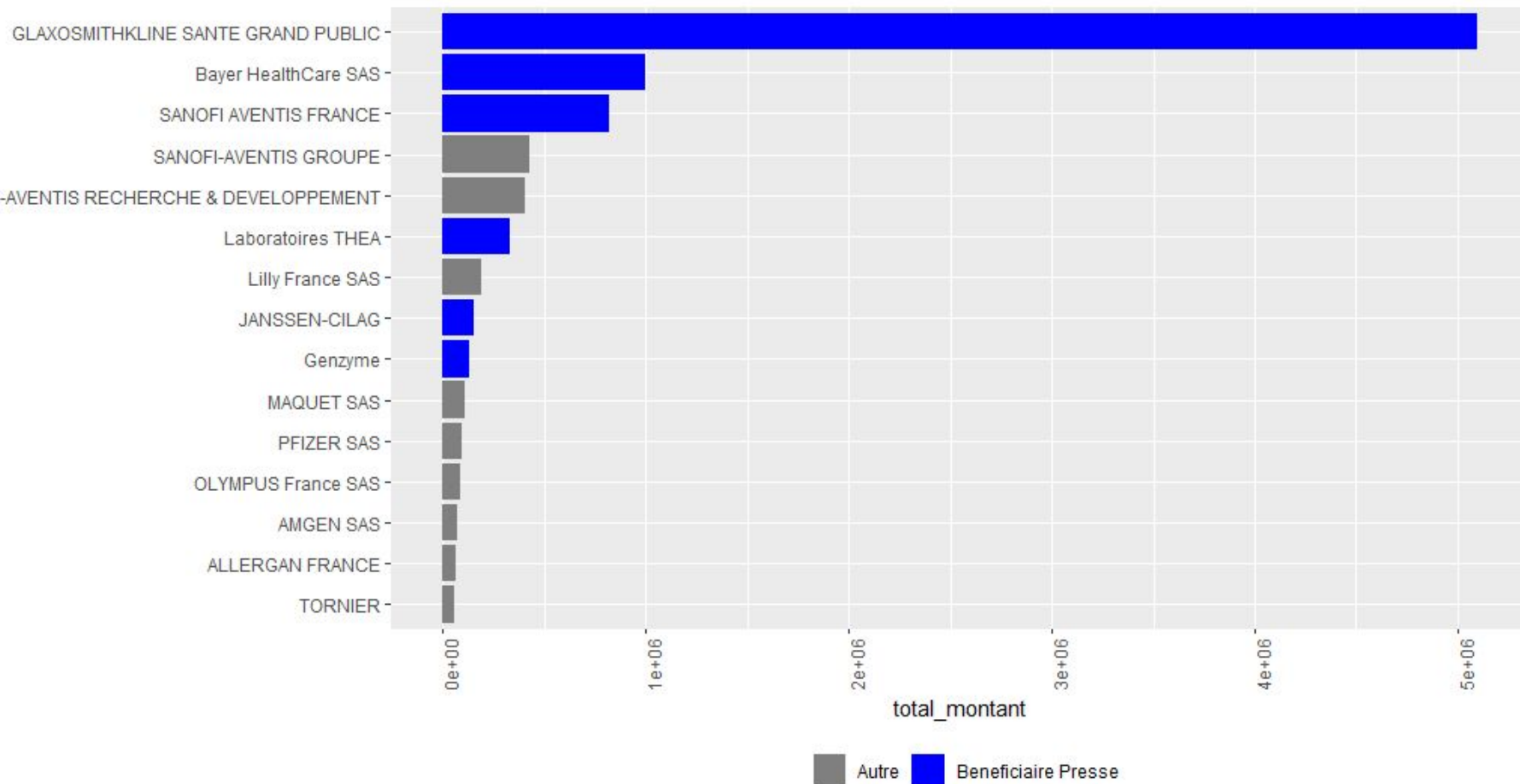


Cas GlaxoSmithKline

Total des rémunérations par mois et par catégorie 2014



reorder(denomination_sociale, total_montant)



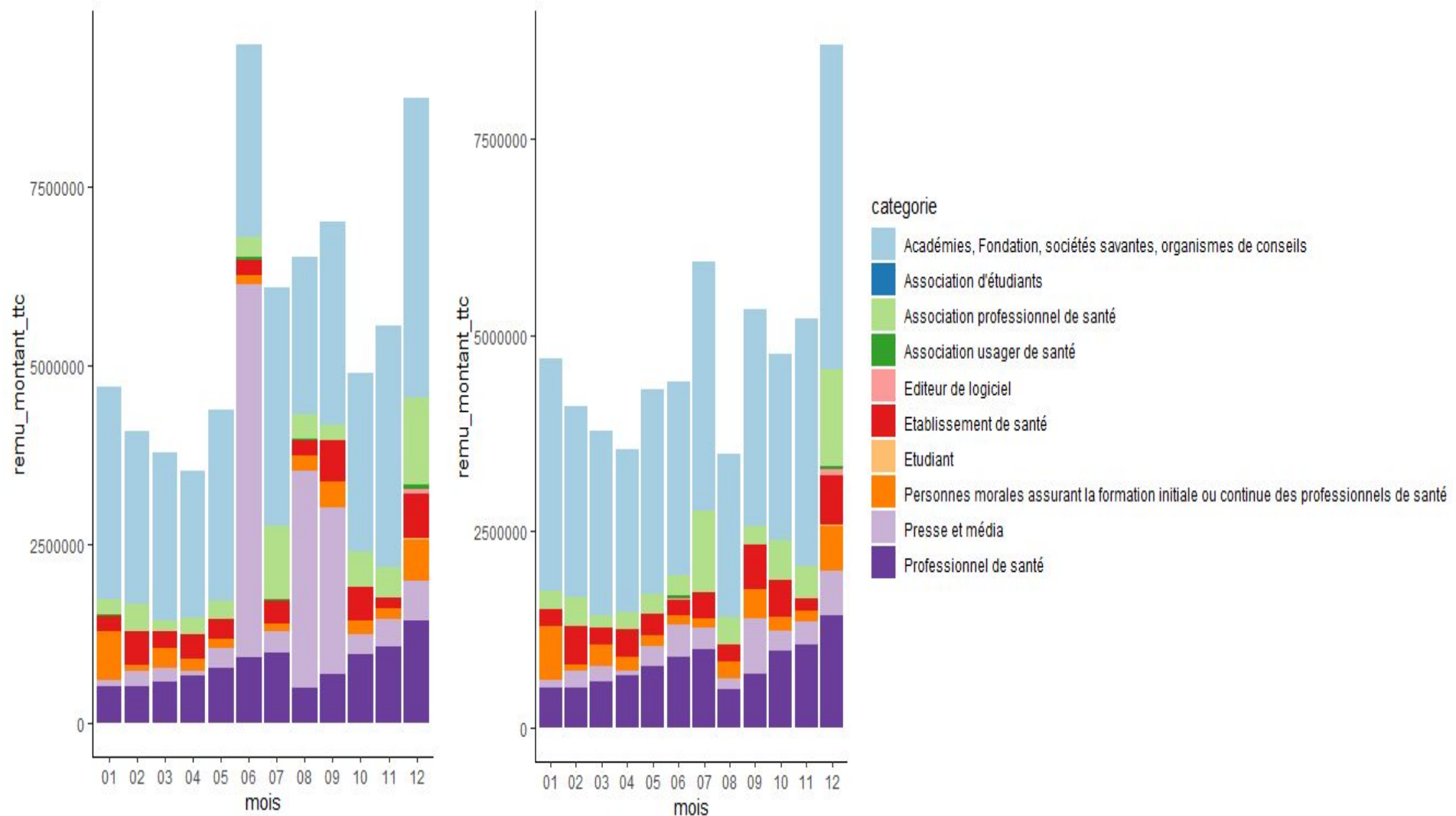


Diagramme de Gantt

