

Université de Reims Champagne-Ardenne

UFR Sciences Exactes et Naturelles

Discipline : Intelligence Artificielle



Réalisation d'une analyse de données via les techniques de visualisation

Réalisé par :

MOALI Sarah

LAFDHAL Ahmed

1. Introduction

La visualisation de données prend de plus en plus d'importance dans tous les secteurs d'activité. C'est un outil très puissant qui consiste à traduire toute donnée ou information dans un contexte visuel, tel qu'un graphique, un diagramme, une carte afin de faciliter la compréhension des données représentées par le cerveau humain et d'en tirer des enseignements.

Dans notre projet, nous allons effectuer une analyse de données via les techniques de visualisation sous RStudio en R.

2. Problématique

Cet ensemble de données est constitué de données sur le placement des étudiants du campus d'une université. Il comprend le pourcentage et la spécialisation de l'enseignement secondaire et secondaire supérieur. Il comprend également la spécialisation du diplôme, le type et l'expérience de travail et les offres de salaire pour les étudiants placés.

Dans le but de déterminer les critères de recrutements, nous allons effectuer une analyse de données via les techniques de visualisation sur le dataset «Placement_Data_Full_Class».

3. Présentation des données :

Pour réaliser cette étude, nous nous sommes appuyés sur un ensemble de données de 215 individus, chacun représentant un étudiant.

De plus chaque individu est caractérisé par 14 variables :

- gender : désigne le genre de l'étudiant prend en valeur M ou F
- ssc_p: désigne le pourcentage dans l'enseignement secondaire - 10e année
- ssc_b= désigne Conseil dans l'éducation. prend en valeur Central ou Others.
- hsc_p=désigne le Pourcentage dans l'enseignement secondaire supérieur - 12e année.
- hsc_b= désigne le conseil de l'éducation. prend en valeur Central ou Others.
- hsc_s=Spécialisation dans l'enseignement secondaire supérieur exemple (commerce , science ..etc)
- degree_p: désigne le pourcentage de diplômés
- degree_t= désigne le domaine d'études du diplôme de premier cycle exemple(Sci&Tech , Comm&Mgmt ..Etc)
- workex=désigne si l'étudiant a de l'expérience ou pas
- salary = désigne le salaire si l'etudiant est recruté
- etest_p =désigne le test d' employabilité
- mba_p = désigne le pourcentage de l'etudiant dans le MBA

4. Diagrammes et interprétation :

Affichage de la base de données dans la console de RStudio :

sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.00	Mkt&HR	58.80	Placed	270000
2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.50	Mkt&Fin	66.28	Placed	200000
3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.00	Mkt&Fin	57.80	Placed	250000
4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.00	Mkt&HR	59.43	Not Placed	NA
5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.80	Mkt&Fin	55.50	Placed	425000
6	M	55.00	Others	49.80	Others	Science	67.25	Sci&Tech	Yes	55.00	Mkt&Fin	51.58	Not Placed	NA
7	F	46.00	Others	49.20	Others	Commerce	79.00	Comm&Mgmt	No	74.28	Mkt&Fin	53.29	Not Placed	NA
8	M	82.00	Central	64.00	Central	Science	66.00	Sci&Tech	Yes	67.00	Mkt&Fin	62.14	Placed	252000
9	M	73.00	Central	79.00	Central	Commerce	72.00	Comm&Mgmt	No	91.34	Mkt&Fin	61.29	Placed	231000
10	M	58.00	Central	70.00	Central	Commerce	61.00	Comm&Mgmt	No	54.00	Mkt&Fin	52.21	Not Placed	NA
11	M	58.00	Central	61.00	Central	Commerce	60.00	Comm&Mgmt	Yes	62.00	Mkt&HR	60.85	Placed	260000
12	M	69.60	Central	68.40	Central	Commerce	78.30	Comm&Mgmt	Yes	60.00	Mkt&Fin	63.70	Placed	250000
13	F	47.00	Central	55.00	Others	Science	65.00	Comm&Mgmt	No	62.00	Mkt&HR	65.04	Not Placed	NA
14	F	77.00	Central	87.00	Central	Commerce	59.00	Comm&Mgmt	No	68.00	Mkt&Fin	68.63	Placed	218000
15	M	62.00	Central	47.00	Central	Commerce	50.00	Comm&Mgmt	No	76.00	Mkt&HR	54.96	Not Placed	NA
16	F	65.00	Central	75.00	Central	Commerce	69.00	Comm&Mgmt	Yes	72.00	Mkt&Fin	64.66	Placed	200000
17	M	63.00	Central	66.20	Central	Commerce	65.60	Comm&Mgmt	Yes	60.00	Mkt&Fin	62.54	Placed	300000
18	F	55.00	Central	67.00	Central	Commerce	64.00	Comm&Mgmt	No	60.00	Mkt&Fin	67.28	Not Placed	NA
19	F	63.00	Central	66.00	Central	Commerce	64.00	Comm&Mgmt	No	68.00	Mkt&HR	64.08	Not Placed	NA
20	M	60.00	Others	67.00	Others	Arts	70.00	Comm&Mgmt	Yes	50.48	Mkt&Fin	77.89	Placed	236000
21	M	62.00	Others	65.00	Others	Commerce	66.00	Comm&Mgmt	No	50.00	Mkt&HR	56.70	Placed	265000

Figure 1 : la base de données « Placement_Data_Full_Class.csv »

Installer et charger les packages "ggplot" et "patchwork", afin de tracer plusieurs plots au même temps.

a. Le genre joue-t-il un rôle dans les recrutements ?

```
plot_gender<-ggplot(data=data, aes(x=gender, y=factor(1),fill = gender)) +
  geom_bar(stat="identity")+facet_wrap(~status)+ylab("effectif")

plot_status<-ggplot(data=data, aes(x=status, y=factor(1),fill = status)) +
  geom_bar(stat="identity")+facet_wrap(~gender)
plot_gender
```

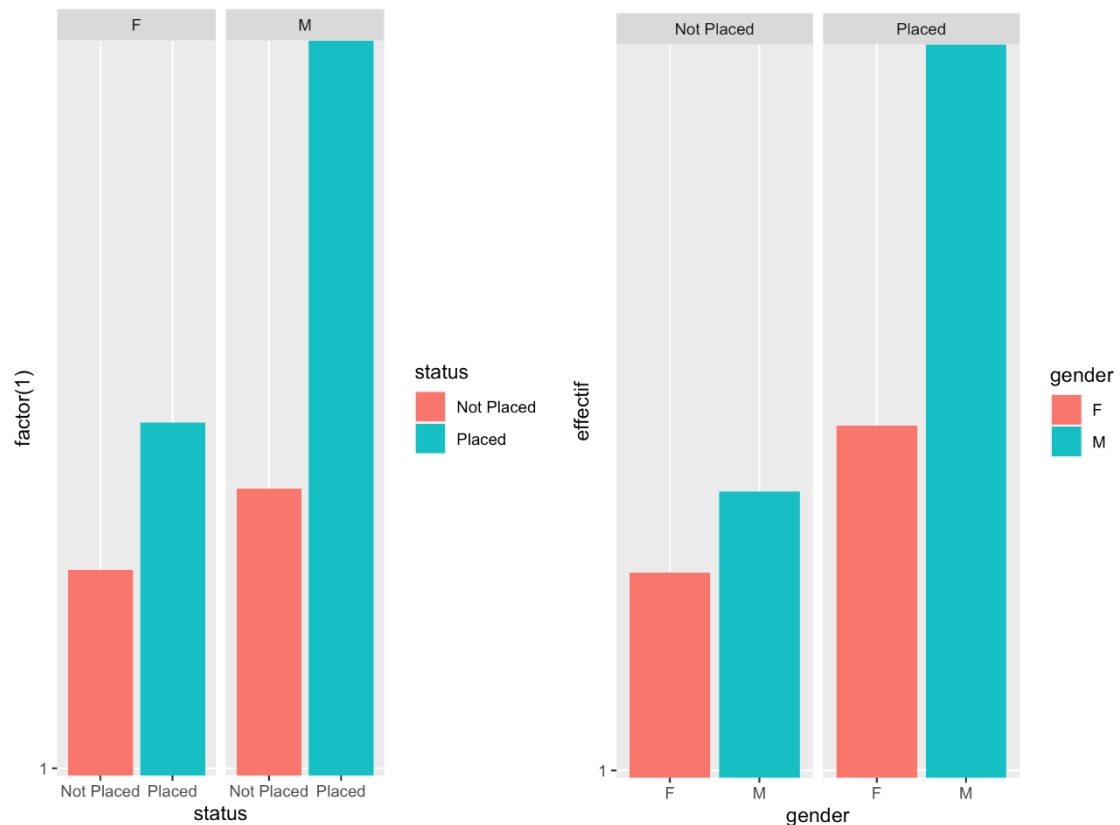


Figure 2 : Le graphe du gauche représente les recrutés et les non recrutés par rapport au sexe et celui de la droite représente le sexe par rapport au recrutement.

D'après les 2 graphes on peut dire que le genre n'est pas un facteur déterminant dans le recrutement d'un étudiant. Le pourcentage d'hommes placés est un peu plus élevé que celui des femmes, mais il n'y a pas de différence radicale entre les deux.

b. Les résultats scolaires des étudiants sont-ils importants pour le recrutement ?

```
plot_ssc_p<-ggplot(data=data) +
  stat_summary(aes(x=gender, y=ssc_p, fill = gender), geom = "col",
    fun.data = mean_se)
+facet_wrap(~status)+ylab("effectif")+ggtitle("Pourcentage 10ème")

plot_hsc_p<-ggplot(data=data) +
  stat_summary(aes(x=gender, y=hsc_p, fill = gender), geom = "col",
    fun.data = mean_se)+facet_wrap(~status)+ylab("effectif")+
  ggtitle("Pourcentage 12ème")

plot_degree_p<-ggplot(data=data) +
  stat_summary(aes(x=gender, y=degree_p, fill = gender), geom = "col",
    fun.data = mean_se)+facet_wrap(~status)+ylab("effectif")+
  ggtitle("Pourcentage Diplômé")

plot_mba_p<-ggplot(data=data) +
  stat_summary(aes(x=gender, y=mba_p, fill = gender), geom = "col",
    fun.data = mean_se)+facet_wrap(~status)+ylab("effectif")+
  ggtitle("Pourcentage mba")

plot_ssc_p+plot_hsc_p+plot_degree_p+plot_mba_p
...
```

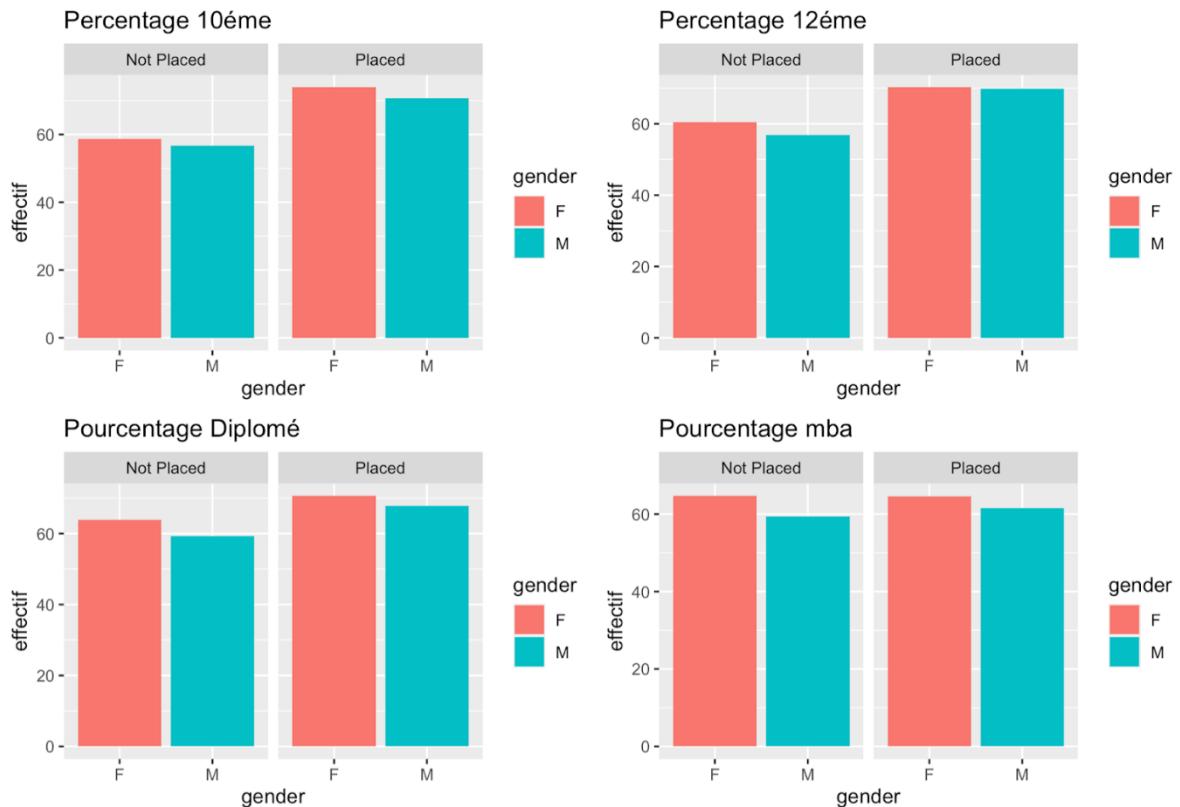


Figure 2 : diagrammes représentant les étudiants par leurs genres par rapport à leur statut

En regardant les diagrammes à barres, il est clair que les étudiants qui ont été placés ont eu de meilleures performances en 10ème, 12ème et premier cycle universitaire.

Mais la performance des étudiants placés et non placés dans leur MBA est presque similaire et donc pas un facteur décisif pour être placé.

En outre, les femmes ont obtenu de meilleurs résultats scolaires que les hommes aux quatre niveaux d'éducation.

D'après ce graphe on constate que les étudiants qui ont été recrutés ont eu de meilleure performance en 10ème et 12ème et premier cycle universitaire.

Cependant la performance des étudiants recrutés et non recrutés dans leur MBA est similaire donc le MBA n'est pas un facteur décisif dans le recrutement.

c. L'expérience affecte-elle le recrutement ?

Pour cela on visualise les étudiants qui ont de l'expérience par rapport à s'ils sont recrutés ou pas.

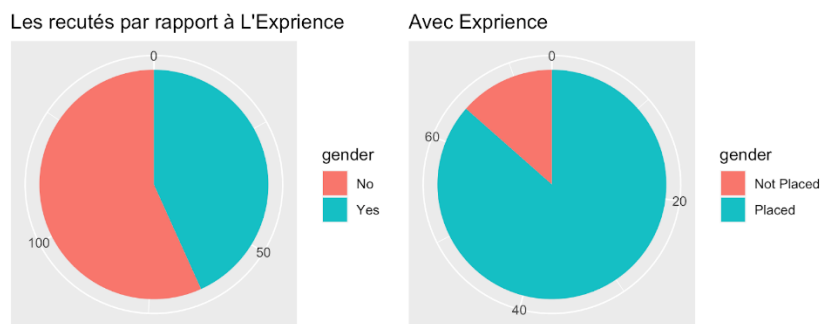
```

cp <- coord_polar(theta = "y")
cp$axis_free <- function() TRUE
worknoexp<-ggplot(data[data$status=="Placed",]) +
  geom_bar(aes(x= factor(1),fill =workex )) +
  cp +
  theme(axis.text.y = element_blank(),
        axis.title.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.x = element_blank()) +
  labs(fill = "gender")+ggtitle("Les recrutés par rapport à L'Exprience")

workexp<-ggplot(data[data$workex=="Yes",]) +
  geom_bar(aes(x= factor(1),fill =status )) +
  cp +
  theme(axis.text.y = element_blank(),
        axis.title.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.x = element_blank()) +
  labs(fill = "gender")+ggtitle("Avec Exprience")

worknoexp+workexp

```



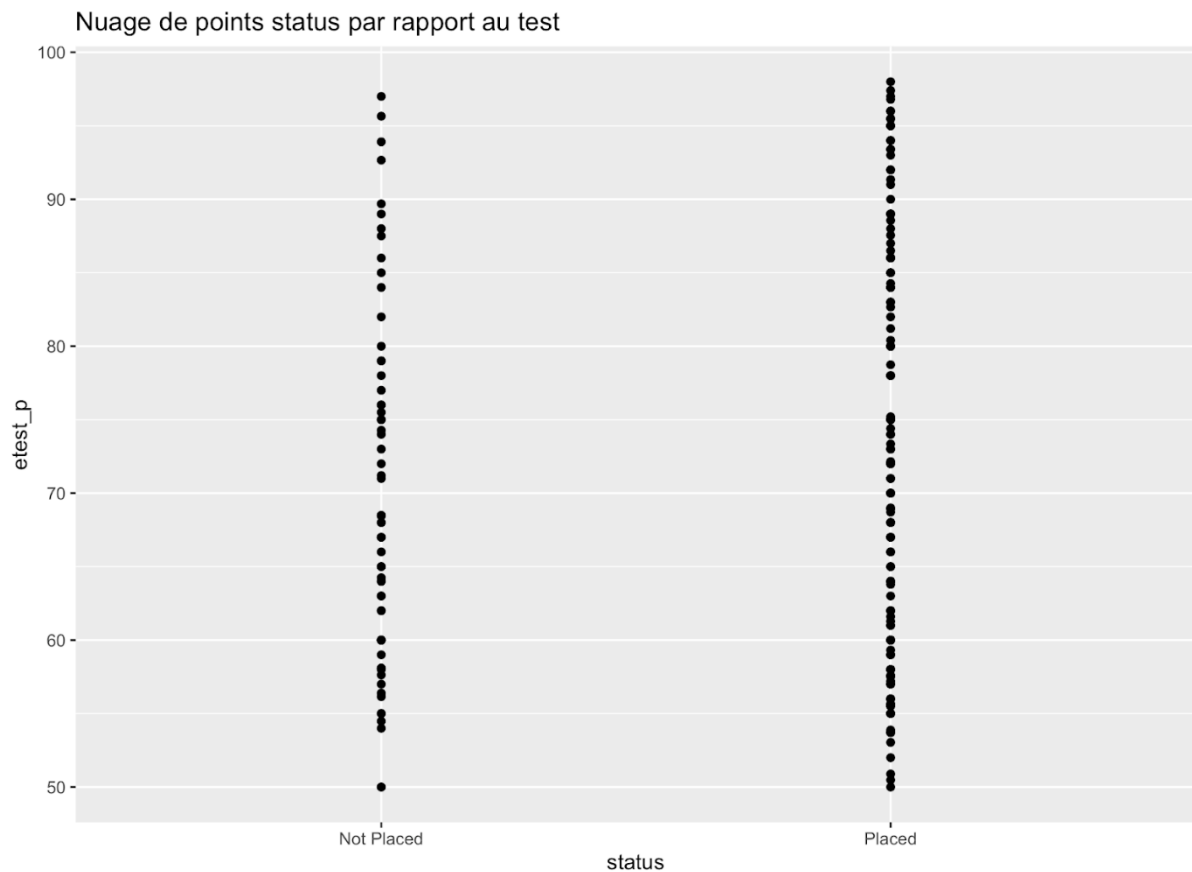
D'après le premier piechart on constate que l'expérience professionnelle n'est pas un facteur principal de décider si des personnes seront recrutés mais d'après le 2eme on constate que si on a de l'expérience professionnel on a une grande chance d'être recruté car on a un petit pourcentage des étudiants non recrutés dans le piechart de ceux qui ont de l'expérience.

d. Le test employabilité joue-il un effet sur le recrutement

```

ggplot(data, aes(x=status, y=etest_p)) + geom_point()+
  ggtitle("Nuage de points status par rapport au test ")

```



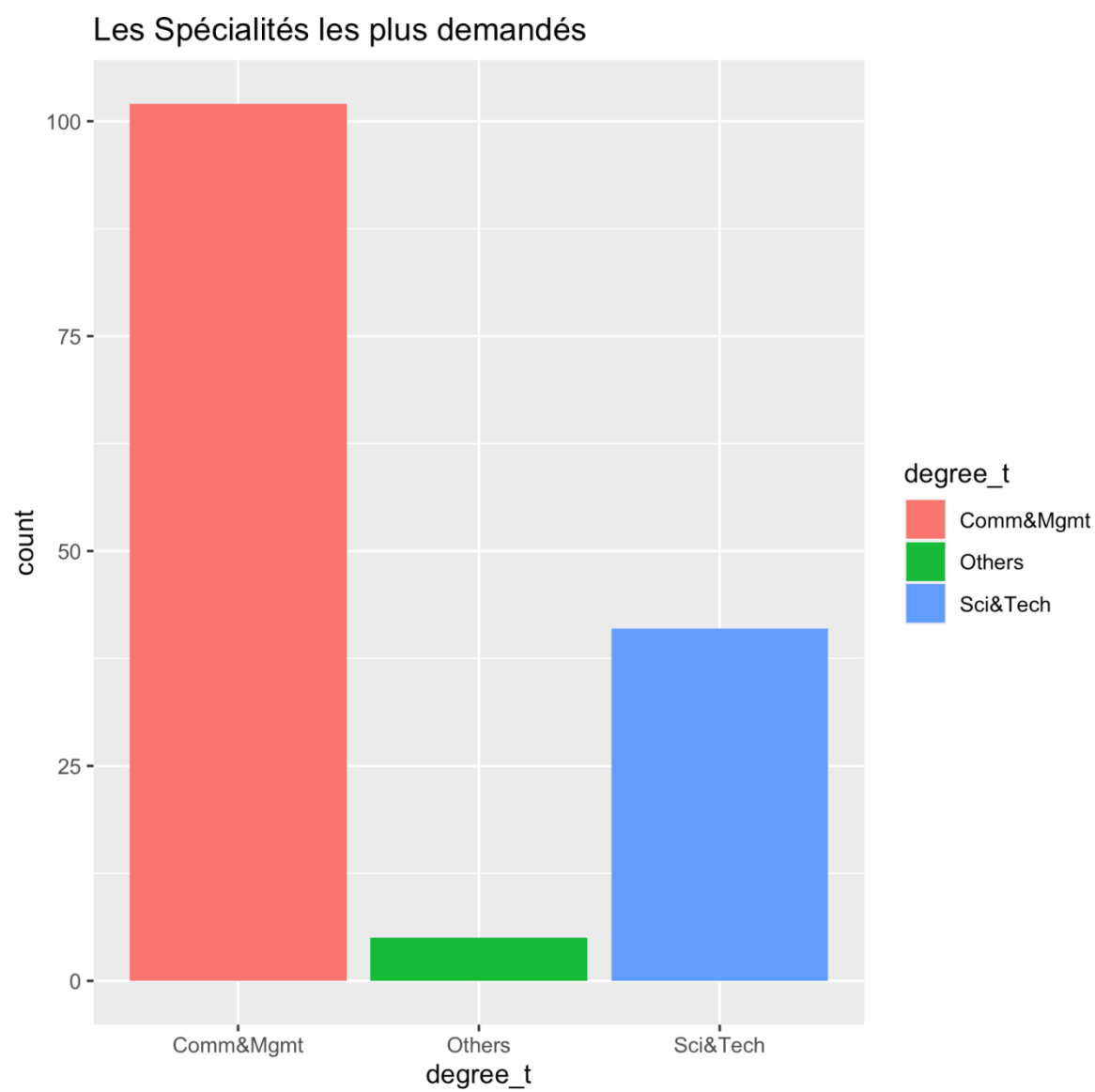
En regardant le diagramme de dispersion, il est évident que le pourcentage du test d'employabilité n'est pas très important car les pourcentages sont uniformément distribués pour les candidats placés et non placés donc ne joue pas un rôle sur le recrutement.

e. Quelles sont les spécialités les plus demandées ?

```

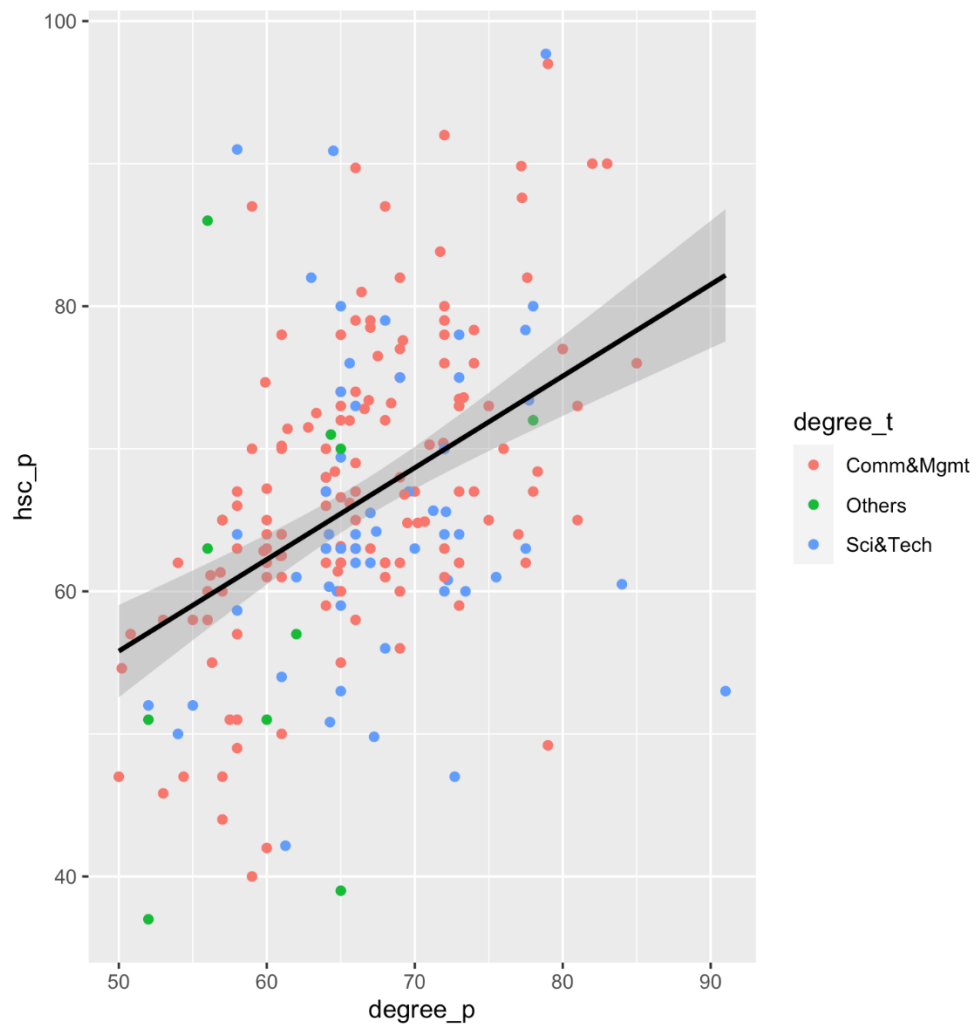
|
g <- ggplot(data[data$status=="Placed",], aes(degree_t,fill=degree_t))
g + geom_bar(aes(weight = sum(1)))

```

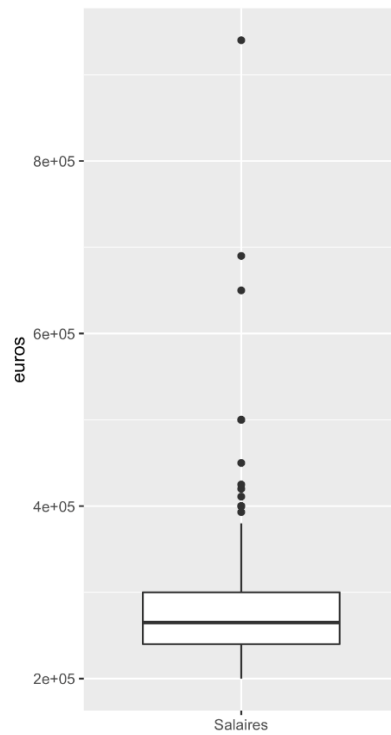


D'après le graphe les spécialités les plus demandées sont communication et management (Comm&Mgmt) et Science/technologie (Sci&Tech) ensuite les autres avec un grand écart.

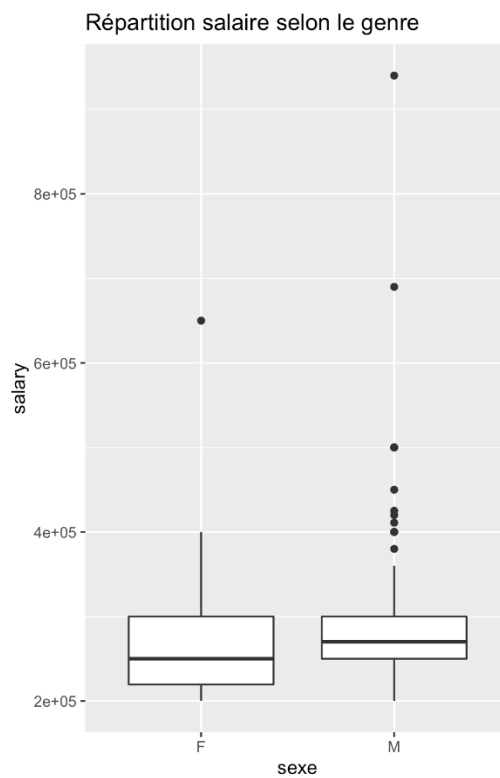
5) Quelques graphes utiles :



Ce graphe nous montre la corrélation entre les notes lycée avec les notes des études supérieures avec la droite de moindres carrés.



On constate que la majorité des étudiants ont le salaire entre 2000 et 4000 euros



On constate que les Hommes sont un peu plus payés que les femmes.

Conclusion :

Après la visualisation du jeu de données et l'analyse de ces diagrammes, nous pouvons déduire que :

- ✓ Plus de candidats masculins ont été placés par rapport aux candidates féminines (car plus de candidats masculins se sont présentés aux placements).
- ✓ Les candidats masculins ont obtenu des CTC plus élevés que les candidates féminines.
- ✓ Le type de conseil d'administration choisi n'a pas d'effet sur les placements, nous pouvons donc le laisser tomber dans les étapes de prétraitement.
- ✓ Les candidats ayant des pourcentages plus élevés dans leur 10e, 12e et undergrad ont de meilleures chances de placement.
- ✓ Le pourcentage de MBA des étudiants n'est pas un facteur décisif pour le placement car la plupart d'entre eux ont un pourcentage similaire.
- ✓ La plupart des candidats n'ont pas d'expérience professionnelle lorsqu'ils se présentent pour un placement.
- ✓ Le fait d'avoir une expérience professionnelle augmente les chances d'être placé.
- ✓ Le pourcentage du test d'employabilité n'a pas d'effet sur les placements.
- ✓ La majeure Mkt&Fin a des salaires plus élevés et plus de chance de placement par rapport à Mkt&HR.
- ✓ La plupart des étudiants ont obtenu leur diplôme de premier cycle dans le domaine des sciences ou de la technologie.
- ✓ Les étudiants issus de ces deux filières ont de meilleures chances de trouver un emploi.