

exercice 22

Nous utiliserons des données sont extraites d'un recueil de données issu d'une enquête portant sur une population d'enseignants de collèges.

Table 1: Extrait des données

Sexe	Age	EtatCivil	Nbenfant	Diplome	Anciennete	Salaire	Satisfaction	Stress	EstimeSoi	AvisReforme
Homme	37	Célibataire	0	Bac+3	11	1600	14.45	15.70	16.15	Défavorable
Homme	38	Célibataire	2	Bac+3	14	1670	17.57	18.88	17.56	Défavorable
Femme	29	Célibataire	0	Bac+3	1	1600	4.05	21.38	4.31	Défavorable
Homme	53	Marié(e)	2	Bac+3	28	1896	32.55	13.88	34.56	Défavorable
Homme	30	Marié(e)	1	Bac+3	7	1996	10.50	17.90	10.05	Défavorable
Homme	44	Marié(e)	2	Bac+3	18	1960	22.16	18.76	22.62	Défavorable

Le tableau 1 est un extrait des données que nous utiliserons. Il y a un total de 11 variables, pour 168 individus. La plupart des variables sont explicites. Le salaire est exprimé en euros, l'âge et l'ancienneté en années. Le stress, l'estime de soi et la satisfaction au travail sont mesurés sur des échelles allant de 0 à 50 suivant des techniques appropriées.

Table 2: Résumé statistique du salaire

Salaire
Min. :1200
1st Qu.:1650
Median :1720
Mean :1778
3rd Qu.:1908
Max. :2200

Quand on s'intéresse de près au salaire dans nos données, grâce au tableau 2, on trouve un minimum de 1200€, un maximum de 2200€, et un salaire médian de 1720€.

Croisement qualitatif vs qualitatif : Sexe et EtatCivil

Essayons de croiser deux variables, avec le sexe et l'état civil.

```
## Margins computed over dimensions
## in the following order:
## 1: Sexe
## 2: EtatCivil
```

Table 3: Tableau de contingence entre le sexe et l'état civil

	Célibataire	Divorcé(e)	Marié(e)	Veuf(ve)	sum
Femme	7	3	38	5	53
Homme	17	10	86	2	115
sum	24	13	124	7	168

On commence par croiser les effectifs de nos deux variables dans le tableau 3. On apprend par exemple que sur les 168 hommes de nos données, 89 sont mariés.

```
## Margins computed over dimensions
## in the following order:
## 1: Sexe
## 2: EtatCivil
```

Table 4: Tableau des fréquences croisée entre le sexe et l'état civil en %

	Célibataire	Divorcé(e)	Marié(e)	Veuf(ve)	sum
Femme	4.17	1.79	22.62	2.98	31.55
Homme	10.12	5.95	51.19	1.19	68.45
sum	14.29	7.74	73.81	4.17	100.00

Le tableau 4 nous donne les pourcentages du croisement entre nos deux variables. On remarque que 73.84% de nos individus sont mariés dont 22.62% sont des femmes.

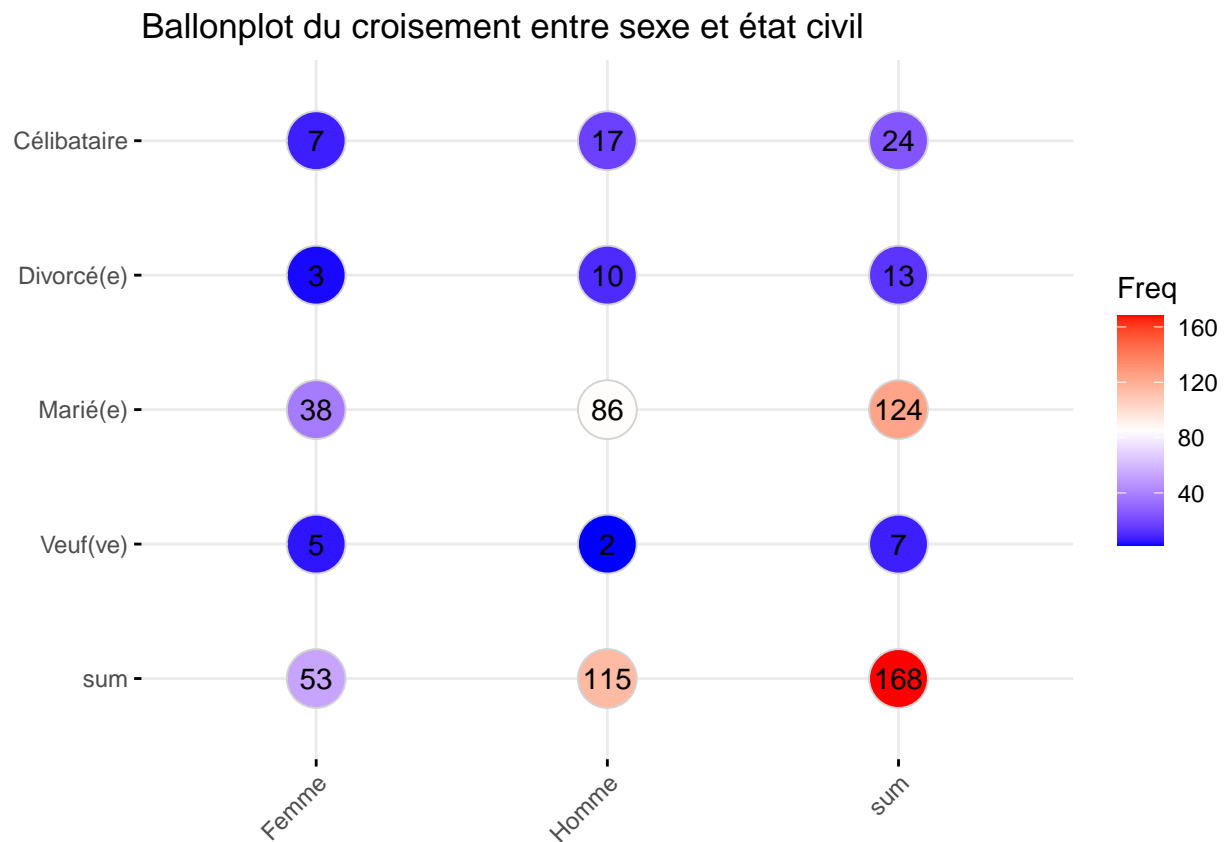


Figure 1: Ballonplot du croisement entre sexe et état civil

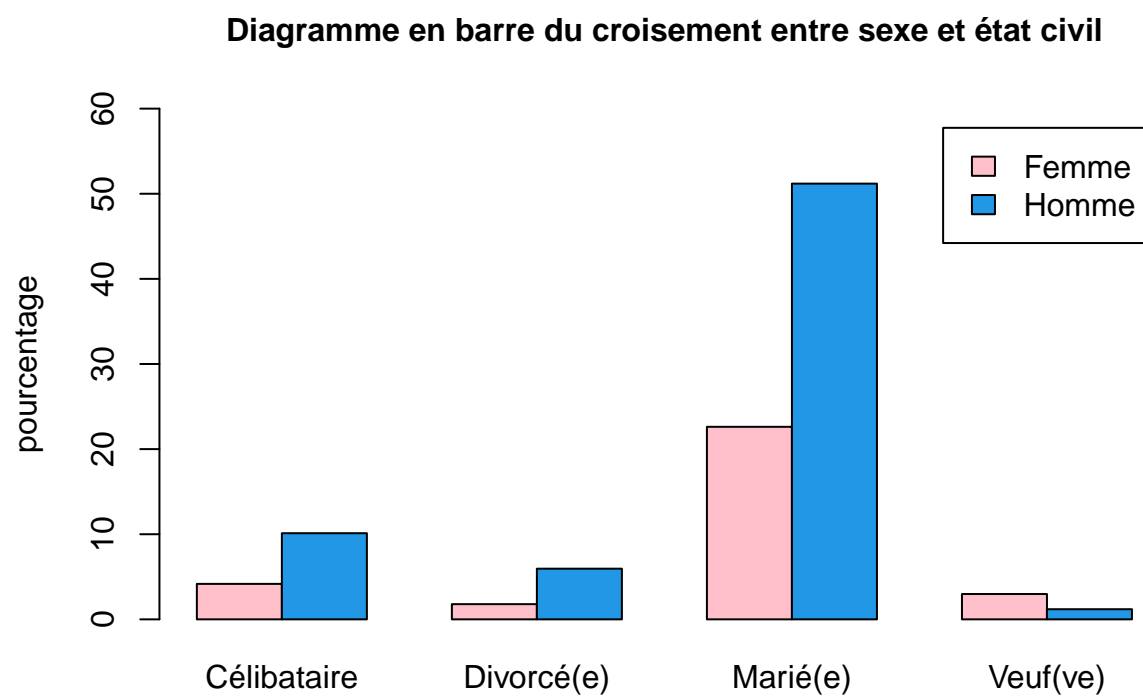


Figure 2: Diagramme en barre du croisement entre sexe et état civil

Nous pouvons rendre graphique les résultats de nos tableau de contingence, comme le font les figures 1 et 2.

Table 5: Distribution conditionnelle du sexe sachant l'état civil en %

	Célibataire	Divorcé(e)	Marié(e)	Veuf(ve)	sum
Femme	29.17	23.08	30.65	71.43	31.55
Homme	70.83	76.92	69.35	28.57	68.45
sum	100.00	100.00	100.00	100.00	100.00

Le tableau 5 nous apprend que parmi nos individus veufs, 71.3% sont des femmes.

Table 6: Distribution conditionnelle de l'état civil sachant le sexe en %

	Célibataire	Divorcé(e)	Marié(e)	Veuf(ve)	sum
Femme	13.21	5.66	71.70	9.43	100
Homme	14.78	8.70	74.78	1.74	100
sum	14.29	7.74	73.81	4.17	100

Le tableau 6 nous dit que parmi nos individus femmes, 13.21% sont célibataires.

```
##
## Pearson's Chi-squared test
##
## data:  TDC_E
## X-squared = 5.6972, df = 8, p-value = 0.6811
```

Quand on réalise le test pour savoir s'il ya indépendance entre nos deux variables, on trouve une p_valeur plutôt grande, ce qui nous ne permet pas de rejeter H_0 , il y a indépendance entre le sexe et l'état civil.

Table 7: Tableau des effectifs théoriques

	Célibataire	Divorcé(e)	Marié(e)	Veuf(ve)	sum
Femme	7.57	4.1	39.12	2.21	53
Homme	16.43	8.9	84.88	4.79	115
sum	24.00	13.0	124.00	7.00	168

Grâce au tableau 7 on peut obtenir la statistique de notre test. On trouve 5.69 ce résultat est visible dans les sorties de notre de test. Le quantile de la lois de khi2 est de 7.81, plus grand que notre stat de test. On ne rejette donc pas l'hypothèse d'indépendance.

Croisement quantitatif vs qualitatif : Stress vs EtatCivil

maintenant croisons l'état civil avec une autre variable qui est le stress.

Table 8: Statistique élémentaire de la variable stress

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.7	15.185	18.19	18.2044	21.115	31.84

Regardons les statistiques élémentaires de cette variable. Avec le tableau 8 on remarque par exemple que le stress moyen est de 18.20.

Boite à moustache de la variable stress

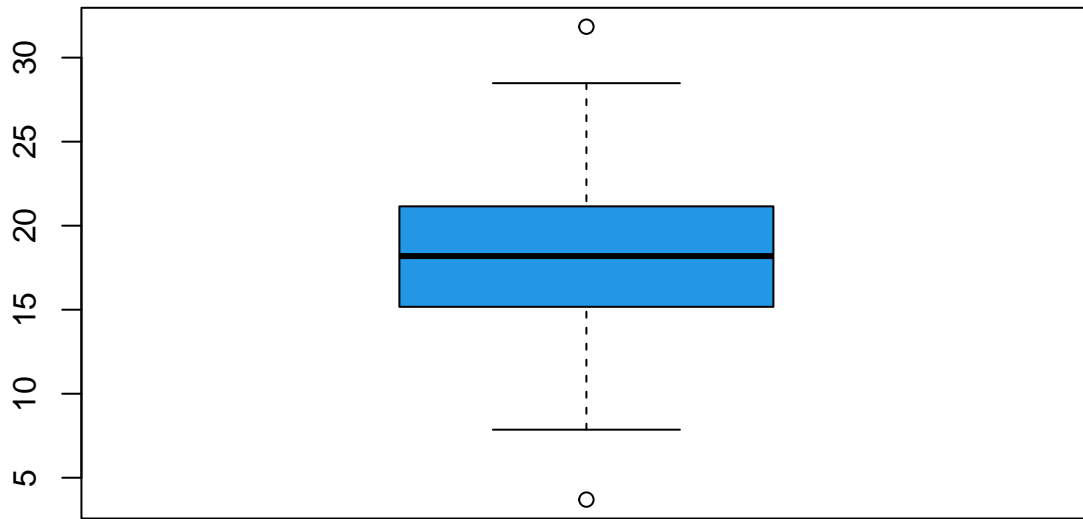


Figure 3: Boite à moustache de la variable stress

Illustrons ces statistiques. Sur la figure 3 on voit que cette variable est bien distribuer autour de la médiane. On voit également qu'il y a 2 valeurs aberrantes.

```
## Margins computed over dimensions
## in the following order:
## 1: EtatCivil
## 2: decoup_stress
```

Table 9: Croisement entre l'état civil et le stress

	(3.67,9.33]	(9.33,15]	(15,20.6]	(20.6,26.2]	(26.2,31.9]	sum
Célibataire	2	4	15	3	0	24
Divorcé(e)	1	3	7	1	1	13

	(3.67,9.33]	(9.33,15]	(15,20.6]	(20.6,26.2]	(26.2,31.9]	sum
Marié(e)	2	23	58	34	7	124
Veuf(ve)	1	4	2	0	0	7
sum	6	34	82	38	8	168

Pour analyser le croisement entre ces deux variables il faut au préalable découper en classe la variable stress. Nous la découpons en 5 classes de même amplitude. Une fois réalisé nous pouvons faire le tableau 9 du croisement des effectifs. On apprend par exemple que 58 individus mariés ont un stress entre 15 et 20.6.

```
## Margins computed over dimensions
## in the following order:
## 1: EtatCivil
## 2: decoup_stress
```

Table 10: Fréquence croisée en %

	(3.67,9.33]	(9.33,15]	(15,20.6]	(20.6,26.2]	(26.2,31.9]	sum
Célibataire	1.190	2.381	8.929	1.786	0.000	14.286
Divorcé(e)	0.595	1.786	4.167	0.595	0.595	7.738
Marié(e)	1.190	13.690	34.524	20.238	4.167	73.809
Veuf(ve)	0.595	2.381	1.190	0.000	0.000	4.166
sum	3.570	20.238	48.810	22.619	4.762	99.999

Le tableau 10 nous apprend que 2.38% des célibataires ont un stress entre 9.33 et 15.

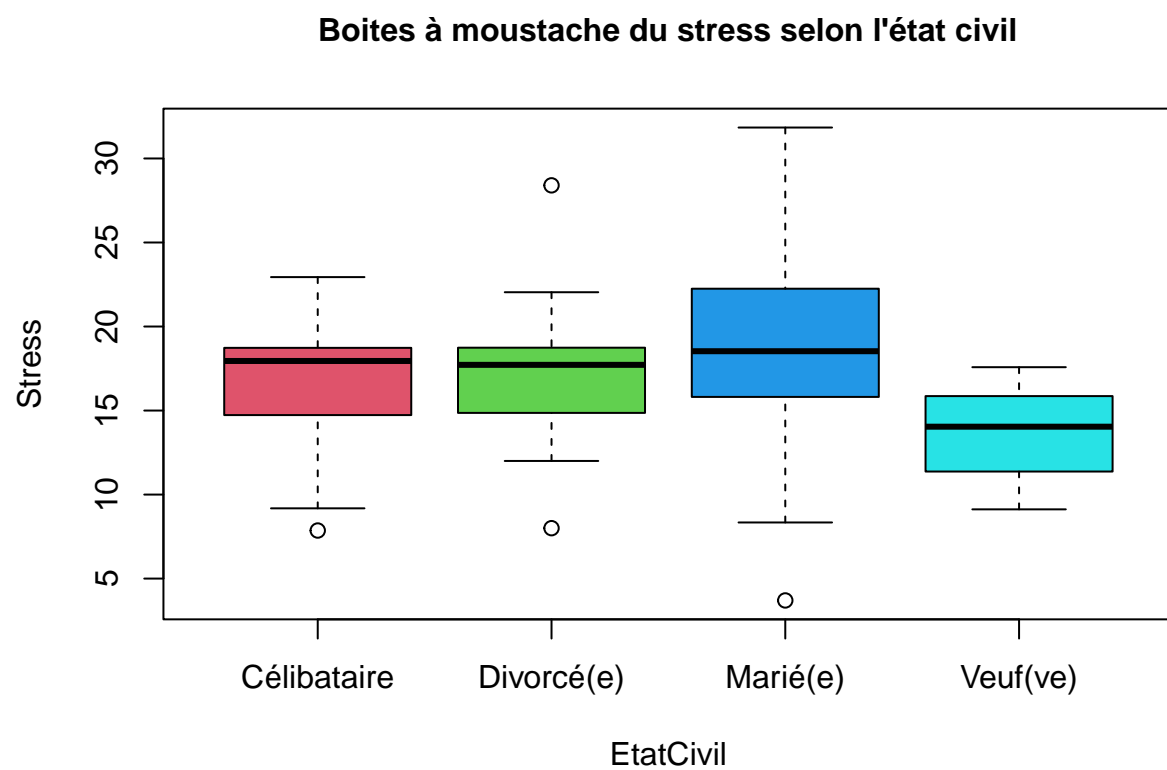


Figure 4: Boîtes à moustache du stress selon l'état civil

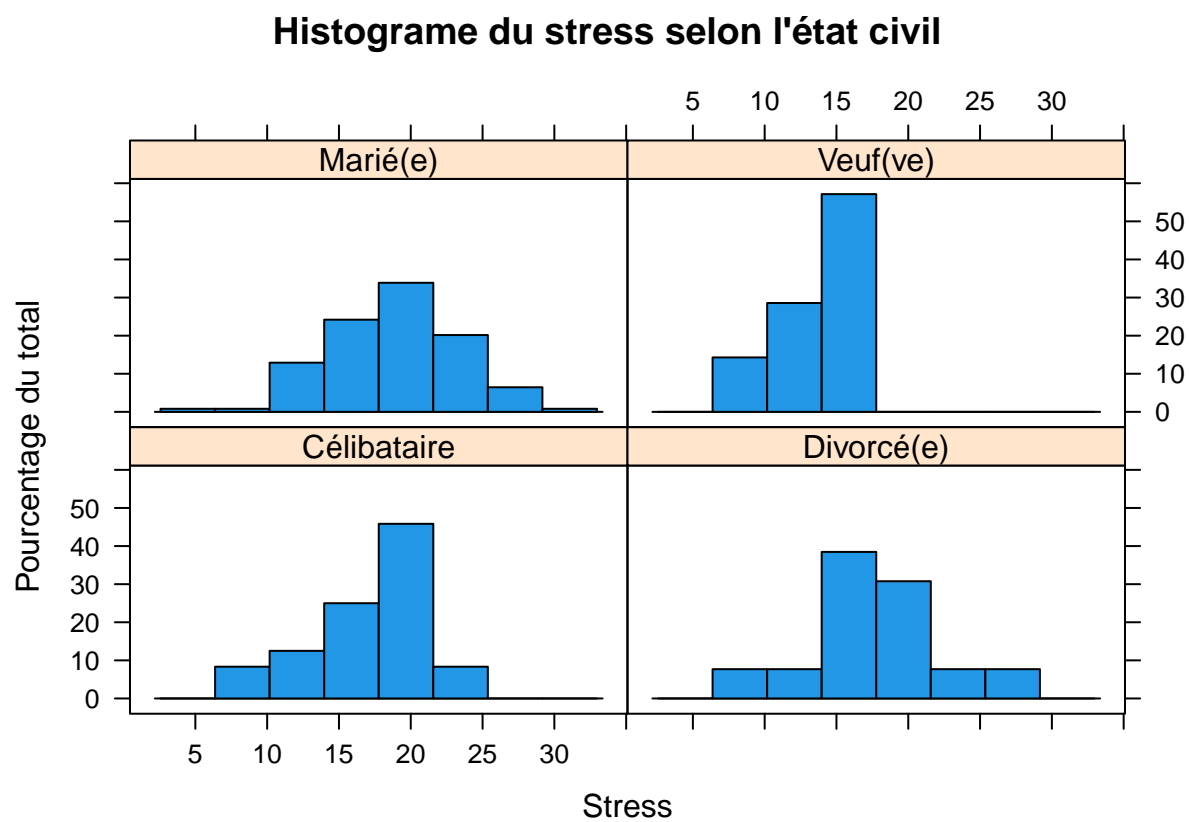


Figure 5: Histogramme du stress selon l'état civil

On remarque avec la figure 4 et 5 que la distribution du stress entre les célibataires et les divorcés est assez similaire. Pour les mariés on retrouve une distribution étendue qui va prendre des valeurs plus grandes. Alors que pour les veufs on retrouve un étendu plus faible, les valeurs du stress pour les veufs est plus faible.

```
## $Célibataire
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      7.86  14.91   17.95   16.76   18.66   22.94
##
## $'Divorcé(e) '
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      8.00  14.86   17.72   17.17   18.74   28.40
##
## $'Marié(e) '
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      3.70  15.82   18.53   18.85   22.25   31.84
##
## $'Veuf(ve) '
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      9.12  11.37   14.04   13.60   15.86   17.58
```

On peut regarder les statistiques élémentaires du stress selon l'état civil dans le tableau 11. On voit par la médiane pour les veufs est plus faible que pour les autres groupes. On retrouve les mêmes résultats que la figure 3.

On peut calculer le rapport de corrélation entre nos deux variables. On trouve 0.075. Cela nous dit que 7.5% de la variabilité du stress est expliquée par l'état civil.

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## EtatCivil      3      265    88.34   4.428 0.00506 **
## Residuals    164     3272    19.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On réalise un test pour si il y a une différence entre les états civils pour la valeur du stress. On trouve une p-valeur inférieure à 0.05, on peut rejeter l'hypothèse 0 et dire qu'il y a une différence entre la valeur du stress selon les différents états civils.

Croisement quantitatif vs quantitatif : Age vs Satisfaction

On va maintenant voir le croisement entre les variables age et satisfaction.

Table 11: Statistique élémentaire de la satisfaction

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.85	13.8375	19.17	20.43101	28.3125	38.45

Table 12: Statistique élémentaire de l'age

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
25	37	41	41.9881	49.25	57

Les tableaux 11 et 12 nous donnent les indicateurs statistiques sur nos variables. On remarque par exemple que la satisfaction moyenne est de 20.43 et celle de l'âge est de 41.99.

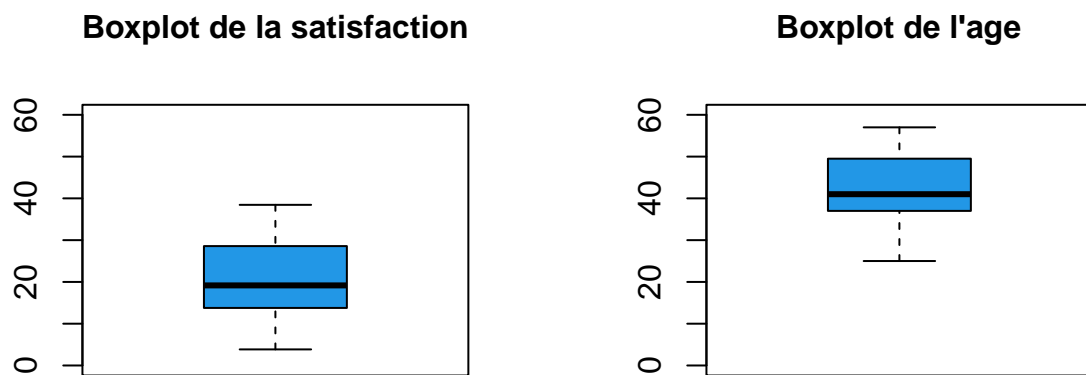


Figure 6: Boxplot des variables satisfaction et age

On peut rendre visuel nos résultats comme dans la figure 6. On remarque que, pour les deux variables, il n'y a pas de valeur aberrante.

```
## Margins computed over dimensions
## in the following order:
## 1: decoup_Satisfaction
## 2: decoup_age
```

Table 13: Tableau de contingence entre la satisfaction et l'âge

	(25,33]	(33,41]	(41,49]	(49,57]	sum
(3.82,10.8]	26	1	0	0	27
(10.8,17.7]	2	38	3	0	43
(17.7,24.6]	0	20	22	0	42
(24.6,31.5]	0	1	11	18	30
(31.5,38.5]	1	0	1	24	26
sum	29	60	37	42	168

Nous avons 2 variables quantitatives, il faut donc créer des classes chacune d'entre elles afin de pouvoir les croiser. Quand on croise les effectifs, on obtient la distribution résumée dans le tableau 13. On apprend par exemple que 38 de nos individus entre 33 et 41 ans ont une satisfaction comprise entre 10.8 et 17.7. Si on est bien attentif on remarque que quand l'âge augmente, la satisfaction augmente également. Regardons cela dans un graphique.

```
## Margins computed over dimensions
## in the following order:
## 1: decoup_Satisfaction
## 2: decoup_age
```

	(25,33]	(33,41]	(41,49]	(49,57]	sum
(3.82,10.8]	15.5	0.6	0.0	0.0	16.1
(10.8,17.7]	1.2	22.6	1.8	0.0	25.6
(17.7,24.6]	0.0	11.9	13.1	0.0	25.0
(24.6,31.5]	0.0	0.6	6.5	10.7	17.8
(31.5,38.5]	0.6	0.0	0.6	14.3	15.5
sum	17.3	35.7	22.0	25.0	100.0

Avant cela regardons d'autre façon d'illustrer le croisement entre nos deux variables. Avec le tableau 14 on retrouve le pourcentage de chaque croisement.

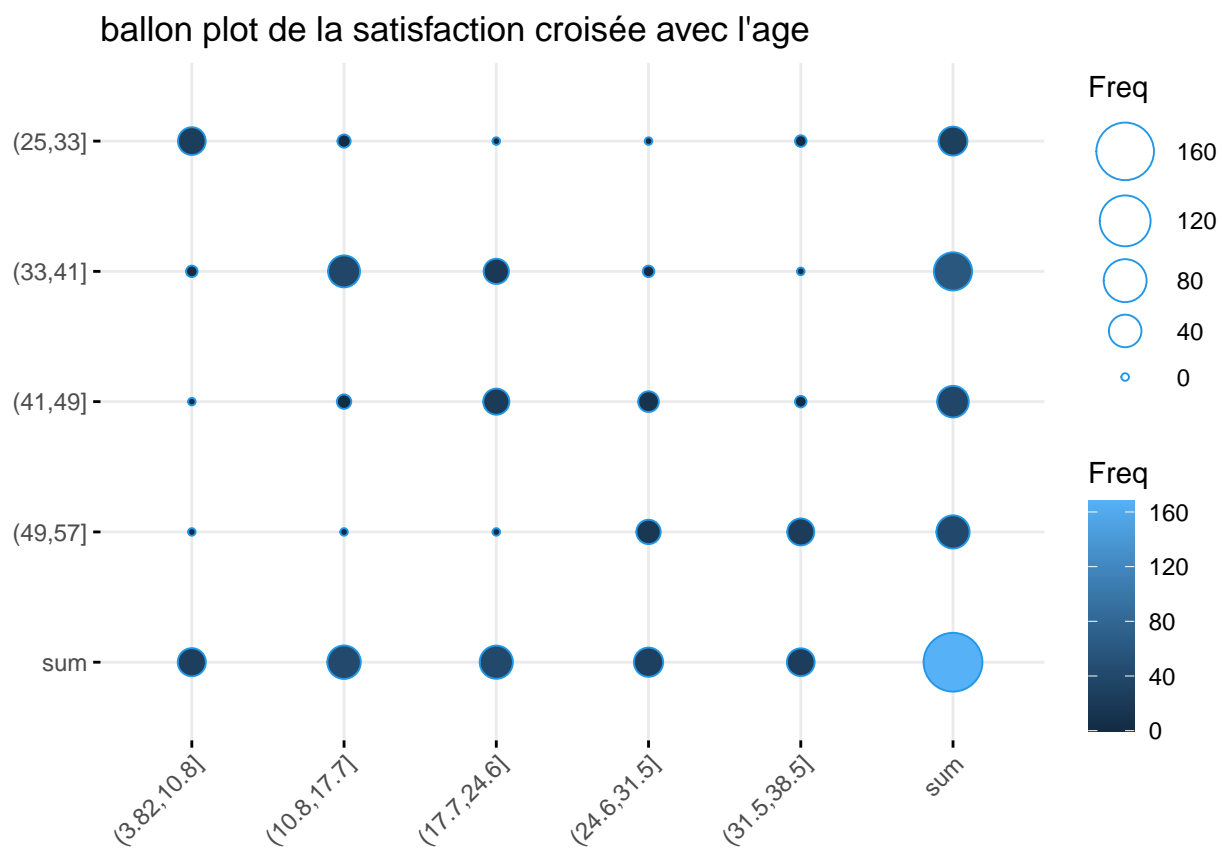


Figure 7: ballon plot de la satisfaction croisée avec l'age

Ou encore avec la figure 7 qui donne un cercle plus ou moins gros selon l'effectif du croisement.

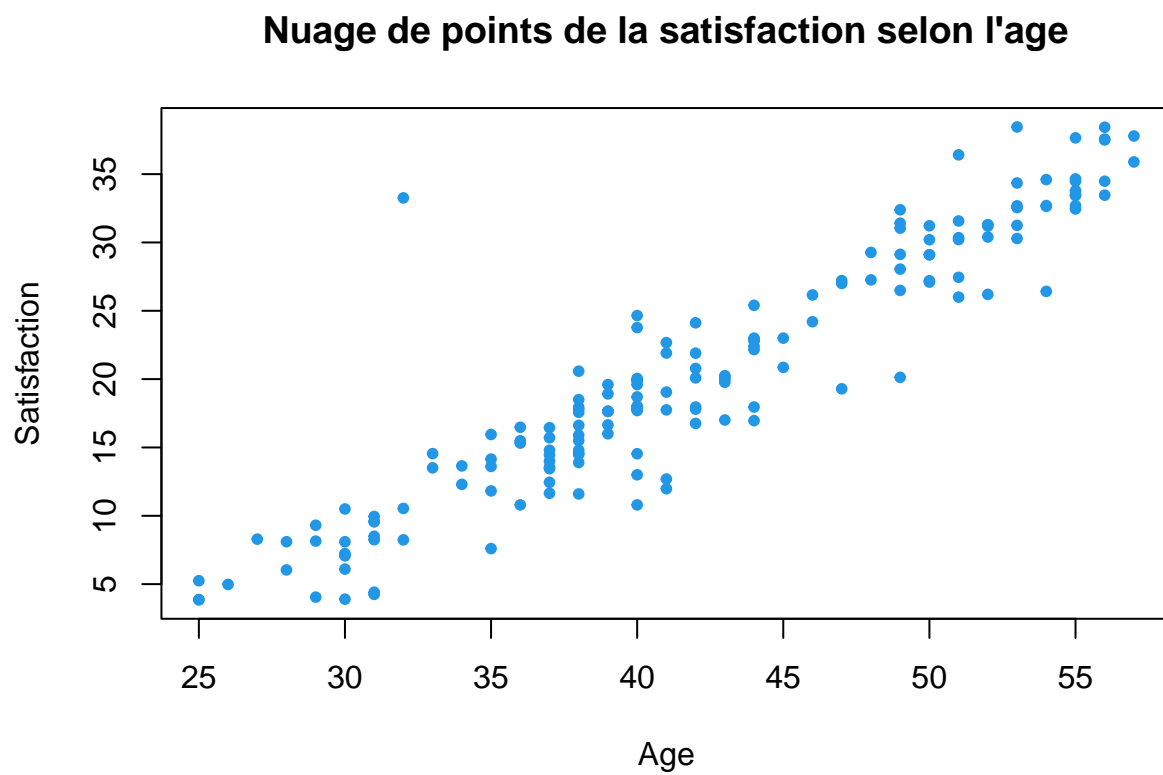


Figure 8: Nuage de points de la satisfaction selon l'age

Graphiquement on se rend compte directement de la liaison entre nos deux variables. La figure 8 nous montre un relation qui est croissante est positif. On voit un point qui ne se comporte pas comme les autres, un intrus.

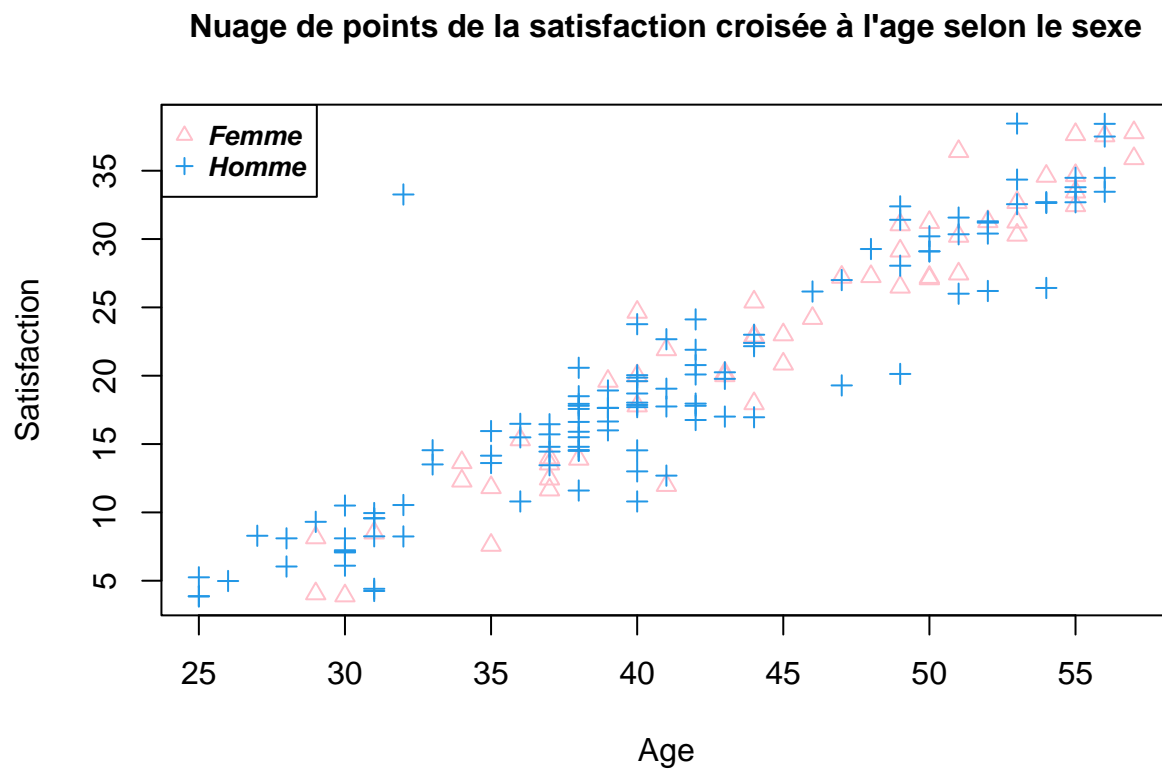


Figure 9: Nuage de points de la satisfaction croisée à l'age selon le sexe

On peut ajouter l'information sur le sexe comme sur la figure 9. Le sexe ne semble pas avoir d'influence sur notre croisement de la satisfaction et de l'âge.

Table 15: Matrice des corrélations

	Anciennete	Salaire	Satisfaction	Stress	EstimeSoi	Age
Anciennete	1.000	0.055	1.000	-0.620	0.995	0.932
Salaire	0.055	1.000	0.074	0.006	0.049	0.056
Satisfaction	1.000	0.074	1.000	-0.612	0.995	0.938
Stress	-0.620	0.006	-0.612	1.000	-0.666	-0.420
EstimeSoi	0.995	0.049	0.995	-0.666	1.000	0.935
Age	0.932	0.056	0.938	-0.420	0.935	1.000

Le tableau 15 nous donne les lien entre chaque variable. Plus le chiffre est proche de 1 plus le liens est fort positivement, plus il est proche de -1 plus le lien est fort mais négativement. Vers 0 la relation est faible. Pour la satisfaction et l'âge on voit un coefficient très proche de 1, ce qui confirme les résultats vue précédement.

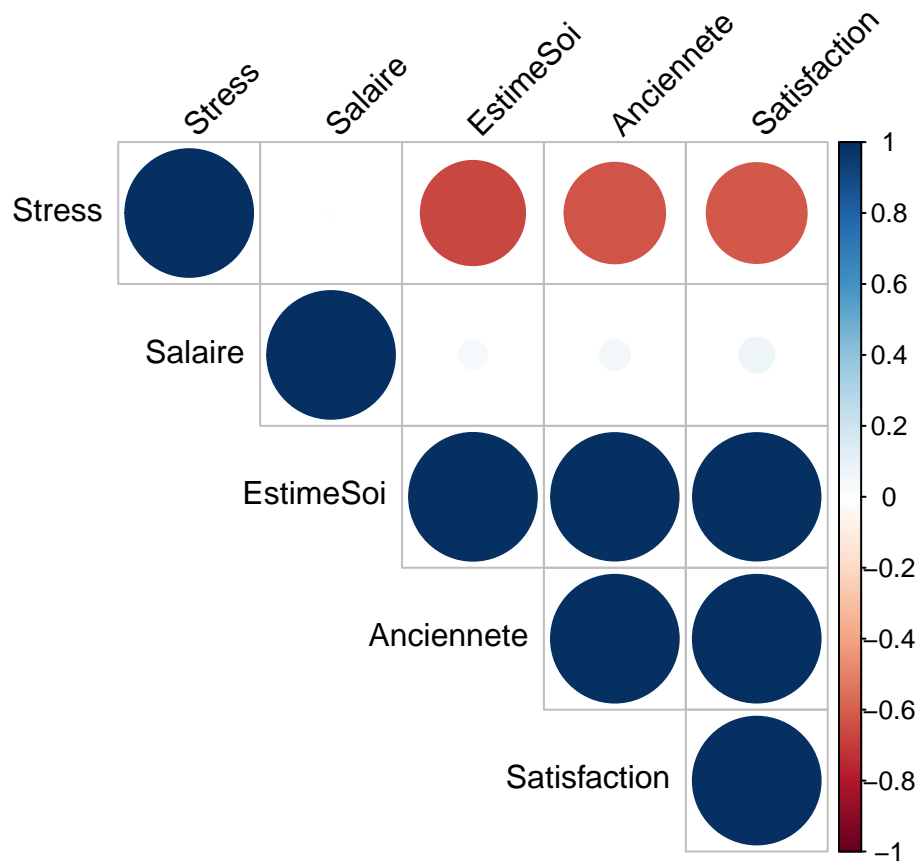


Figure 10: corrélogramme

On peut illustrée ce tableau comme le montre la figure 10. Plus le cercle est bleu plus la corrélation est forte, comme par exemple avec l'ancienneté et la satisfaction.