



## Compte-rendu d'analyse

valant évaluation dans le cadre de :

**Diplôme** : Master M1 - Statistique pour l'Evaluation et la Prévision

**Année universitaire** : 2022-2023

**Module d'enseignement** : SEP 831

**Responsable** : Philippe Regnault

**Comptant pour** : 50 %

## Les déterminants des logements sociaux en France

Alexandre Brunet

Garance Gabaut

alexandre.brunet@etudiant.univ-reims.fr

garance.gabaut@etudiant.univ-reims.fr

### Métriques

**Finalisé le** : 05 mai 2023

**Page(s)** : 47

**Références(s)** : 0

**Figure(s)** : 14

**Table(s)** : 0

**Théorème(s)** : 0

**Résumé** : Aujourd'hui en France, on retrouve près de 5,3 millions de logements sociaux et 10 millions de locataires. À travers cette étude, on cherche à comprendre les différents potentiels déterminants de ces logements, que ce soit économiques, écologiques ou démographiques qui peuvent ou non expliquer la quantité et le prix des logements sociaux pour l'année 2019.

**Mots-clés** : Logements sociaux, départements.

### Matériel supplémentaire :

**git**: <https://github.com/pregnault/urcadown>

**data**: <https://www.data.gouv.fr/fr/datasets/logements-et-logements-sociaux-dans-les-departements-1/>

## Table des matières

1	Introduction	3
2	Impact de la démographie française sur le nombre, le prix et la nature des logements sociaux	3
2.1	Analyse de la répartition de la population selon l'âge . . . . .	3
2.2	Age et prix des logements sociaux . . . . .	6
2.3	L'influence de la localisation des logements sociaux au regard de leur nombre et de la fixation de leur prix . . . . .	6
2.4	Densité de population et taux de logements sociaux . . . . .	7
2.5	Les jeunes et l'environnement : une préoccupation grandissante en matière de logements ? . . . .	7

<b>3</b>	<b>L'équilibre démographique : un déterminant dans le taux de logements sociaux ?</b>	<b>8</b>
3.1	Solde naturel et solde migratoire selon les régions . . . . .	8
3.2	Est-ce qu'une différence significative entre les 2 soldes signifie un plus ou moins grand taux de logements sociaux ? . . . . .	8
<b>4</b>	<b>Situation économique française face au nombre et au prix au <math>m^2</math> des logements sociaux</b>	<b>9</b>
4.1	Les salaires confrontés au loyer mensuel d'un $m^2$ . . . . .	9
4.2	Taux de pauvreté . . . . .	9
4.3	Le taux de chômage et le prix du $m^2$ . . . . .	9
<b>5</b>	<b>Conclusion</b>	<b>11</b>
<b>A</b>	<b>Annexe - partie 2</b>	<b>12</b>
A.1	Corrélation entre la proportion de moins de 20 ans et la proportion des 60 ans et plus par département. . . . .	12
A.2	Corrélation entre la proportion d'une population d'un certain âge et le prix moyen au $m^2$ . . . . .	12
A.3	Etude du lien entre le prix moyen et le nombre de logements . . . . .	13
A.4	Densité de population et taux de logements sociaux . . . . .	18
A.5	La part des logements sociaux énergivores . . . . .	20
A.6	Age moyen des parcs sociaux et l'environnement . . . . .	21
<b>B</b>	<b>Annexe - partie 3</b>	<b>22</b>
B.1	Solde migratoire et solde naturel . . . . .	22
B.2	Impact de différence sur taux de logements . . . . .	39
<b>C</b>	<b>Annexe - partie 4</b>	<b>42</b>
C.1	Variable salaire confrontée à la variable loyer . . . . .	42
C.2	Taux de pauvreté et prix au carré du loyer . . . . .	43
C.3	Le chômage confronté au prix du $m^2$ des logements sociaux . . . . .	44
	<b>Bibliographie</b>	<b>47</b>

## 1. Introduction

La situation économique de la France depuis les années 1970 ne cesse d'être perturbée, en particulier à la suite de la crise liée à la Covid-19. Au cours de cette dernière, de nombreux emplois se sont retrouvés sur le fil du rasoir et ont été contraints de s'interrompre pendant une assez longue période. Les conséquences de cette crise sont nombreuses : pénurie de certains produits, engendrant ainsi une augmentation générale des prix des produits du quotidien, mais aussi de l'énergie (électricité, carburant), particulièrement au cours de l'hiver 2022-2023. Aujourd'hui, une inquiétude globale existe autour du pouvoir d'achat des Français..

Il existe différentes alternatives pour permettre de limiter ses dépenses au quotidien. Par exemple, la mise en place du télétravail de façon plus régulière permet de limiter sa consommation de carburant. Également, pour des foyers plus en difficulté, il est possible de se loger dans des logements plus abordables : les logements sociaux. Ces logements sont construits avec l'aide de l'État et proposent des loyers plus abordables. Des conditions doivent néanmoins être respectées pour pouvoir en bénéficier, notamment un revenu maximal à ne pas dépasser. Cependant, les conditions peuvent être modifiées étant donné la démographie du pays. En effet, la population française ne cesse d'augmenter et est marquée par un solde migratoire assez important (un des pays européens qui a le plus fort solde migratoire pour la période 2010-2015).

On en vient donc à s'interroger sur la situation démographique et l'impact qu'elle a eu sur les prix des logements sociaux en France. D'ailleurs, on note en 2019 une différence de près de 1000 € net par mois sur le salaire moyen entre différentes régions de France. Ainsi, nous avons fait le choix de comparer les différentes régions françaises en nous demandant dans quelle mesure la réalité économique et démographique des différentes régions de la France peut influencer le nombre et la nature des logements sociaux

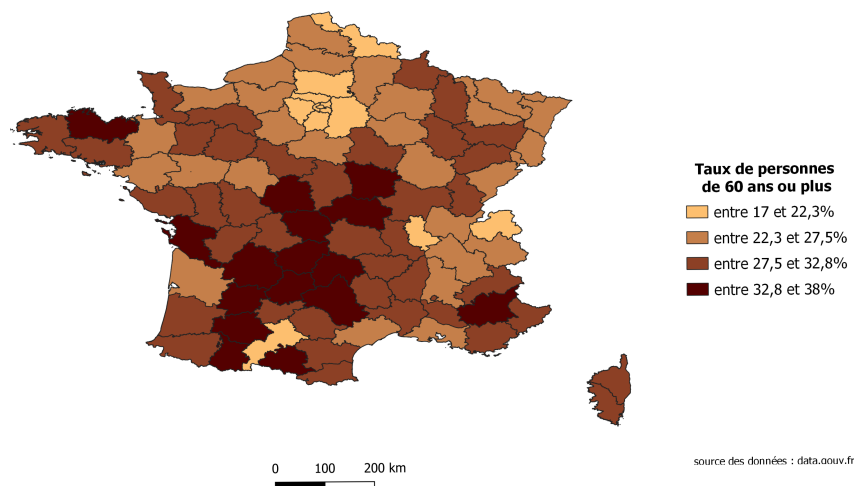
## 2. Impact de la démographie française sur le nombre, le prix et la nature des logements sociaux

Lorsque la question du logement est abordée, la démographie d'un pays est un facteur important à considérer. Il est donc pertinent de s'interroger sur l'impact de ces facteurs démographiques sur le nombre et sur les prix des logements sociaux. Dans un premier temps, nous examinerons s'il existe un lien entre l'âge des habitants et la répartition des logements sociaux. De même, nous étudierons si l'âge a une incidence sur les prix des logements. Ensuite, nous nous pencherons davantage sur la localisation géographique des logements sociaux ainsi que sur leur prix.

Nous poursuivrons notre analyse en examinant la densité de population et le nombre de logements disponibles. Enfin, nous aborderons un sujet d'actualité important, à savoir l'écologie. Nous nous demanderons si une population plus jeune est plus sensible à la prise en compte de l'environnement dans le choix de leur logement.

### 2.1. Analyse de la répartition de la population selon l'âge

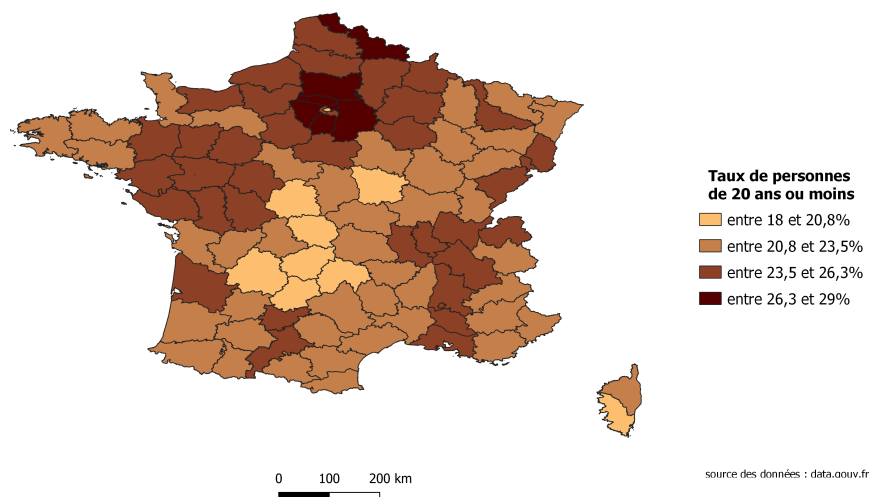
Comme évoqué dans l'article de RENAULT [Ren22], les seniors semblent se diriger vers le sud de la France ou les zones côtières. Nous avons souhaité vérifier si nos données concordaient avec les études menées par Le Figaro. Pour cela, nous avons utilisé le logiciel Qgis, qui nous a permis de produire les trois cartes visibles sur les figures (1), (2) et (3).



**FIGURE 1** – Proportion de personnes âgées de 60 ans ou plus en 2019, en France, en fonction des départements de France. *Lecture* : Dans le département de l'Aube, il y a entre 22,3 et 27,5% de personnes de 60 ans et plus.

La figure (1) nous permet de confirmer quelques propos de l'article : on constate en effet assez peu de personnes âgées dans les départements les plus dynamiques, comme ceux de la région Île-de-France. On observe également une forte concentration dans la diagonale du vide.

Lorsqu'on met en relation cette carte avec celle de la figure (2), on constate qu'elles sont assez complémentaires. Prenons l'exemple des départements situés en Île-de-France : on observe une forte concentration de jeunes et une très faible concentration de personnes de 60 ans et plus. À l'inverse, dans des départements tels que le Cantal, l'Aveyron ou même la Creuse, on note une importante concentration de personnes des 60 ans et plus contre une minorité de moins de 20 ans.



**FIGURE 2** – Proportion de personnes âgées de 20 ans ou moins en 2019, en France, en fonction des départements de France. *Lecture* : Dans le département de l'Aube, il y a entre 23,5 et 26,3% de personnes de moins de 20 ans.

Notre objectif est de mettre en relation la répartition des logements sociaux avec celle des personnes en fonction de l'âge. Pour cela, nous avons produit une dernière carte, visible sur la figure (3), qui indique le taux de logements sociaux en fonction des départements.

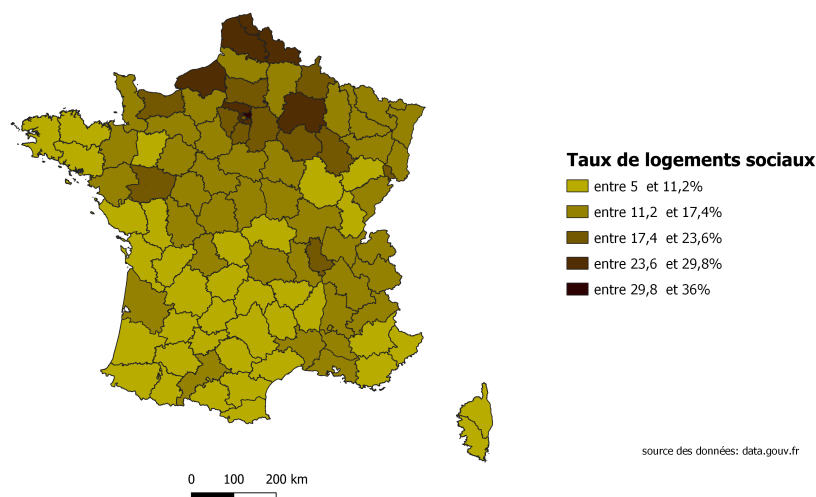


FIGURE 3 – Taux de logements sociaux en fonction des départements français en 2019. *Lecture* : Dans le département de l'Aisne, il y a entre 11,2 et 17,4% de logements sociaux.

Après avoir comparé les figures (1) et (2) avec la figure (3), nous ne sommes pas en mesure d'établir un lien évident entre la répartition de la population en fonction de l'âge et la répartition des logements sociaux. En effet, la carte (3) montre une répartition relativement homogène sur le territoire, à l'exception de 4 départements.

Cependant, une hypothèse intéressante reste à vérifier : il semblerait exister un lien entre la répartition des personnes des 60 ans et plus et celle des personnes de moins de 20 ans. Cette hypothèse est renforcée par la figure (4). En effet, ce graphique montre un lien linéaire entre les deux populations.

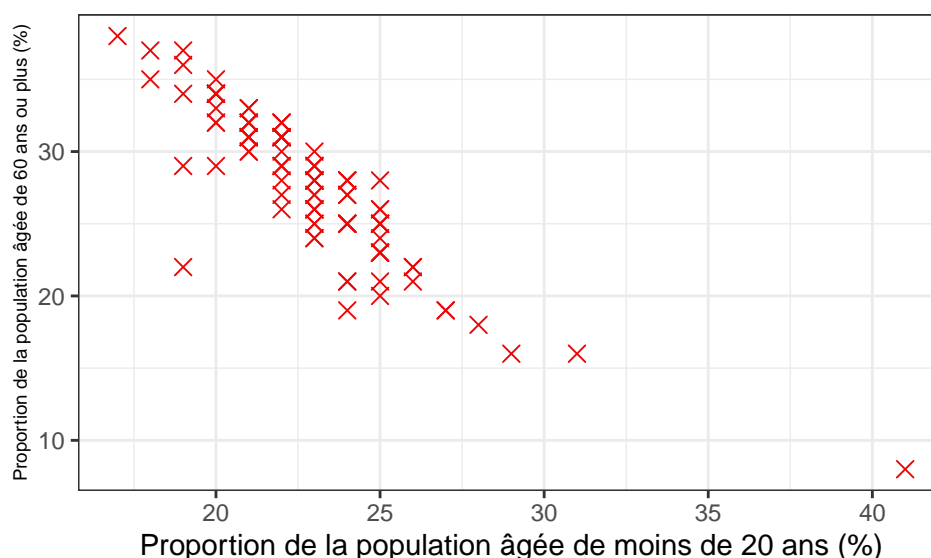


FIGURE 4 – Proportion de personnes âgées de 20 ans ou moins en 2019, en France, en fonction de la proportion de personnes de 60 ans ou plus. *Lecture* : Pour 41% de moins de 20 ans dans un département, on retrouve 9% de plus de 60 ans.

Nous avons confirmé cette hypothèse en effectuant un test de corrélation entre les données de la répartition des plus des 60 ans et plus et celles des moins de 20 ans, qui a révélé une forte corrélation négative. Ainsi, nous pouvons conclure que ces deux populations sont bien inversement corrélées. Autrement dit, si l'une des populations augmente, l'autre diminue.

En somme, bien que la répartition des logements sociaux ne semble pas être directement liée à la répartition de la population en fonction de l'âge, il existe des liens intéressants entre la répartition des différentes générations.

## 2.2. Age et prix des logements sociaux

On s'interroge désormais sur l'existence d'un lien entre le prix moyen au  $m^2$  et la proportion de la population âgée de 60 ans et plus et de la proportion de la population âgée de moins de 20 ans. D'après la figure (5), les données semblent effectivement liées. En effet, une proportion élevée des moins de 20 ans semble entraîner un prix moyen au  $m^2$  élevé, tandis qu'on constate le contraire en regardant le prix en fonction du pourcentage de personnes de plus de 60 ans.

Nous avons effectué un test de corrélation pour vérifier ces observations. Nous avons constaté qu'il existe bien une corrélation négative entre le prix moyen et la proportion de la population âgée de 60 ans et plus, ce qui signifie que les départements dans lesquelles la population est plus âgée ont tendance à fixer un prix moyen au  $m^2$  moins élevé que les autres départements. Par ailleurs, nous avons également constaté qu'il existe une corrélation entre le taux de personnes âgées de moins de 20 ans et le prix moyen, mais cette corrélation est très faible, d'environ 0,38.

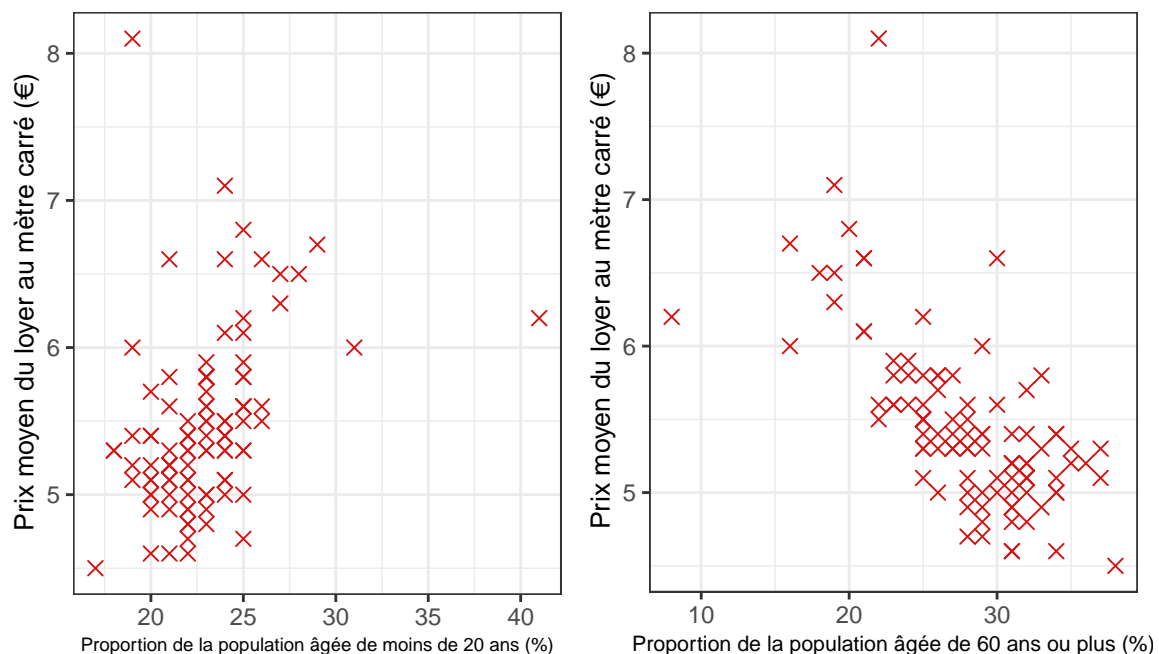


FIGURE 5 – Prix moyen du loyer au  $m^2$  en fonction de l'âge de la population en 2019, en France. *Lecture*: Un département dans lequel environ 18% de la population est âgée de moins de 20 ans possède des logements sociaux avec un prix moyen fixé à 4,5 €/m<sup>2</sup>.

## 2.3. L'influence de la localisation des logements sociaux au regard de leur nombre et de la fixation de leur prix

En théorie, plus un bien, ou service, est présent sur un marché, moins il vaut cher et inversement: "tout ce qui est rare est cher". En l'occurrence, nous nous attendons à ce que les régions qui proposent le moins de logements aient un prix au  $m^2$  plus élevé. Mais dans notre cas, il s'avère que le nombre de logements sociaux varie selon les régions elles-mêmes. En effet, l'Île-de-France est la région qui propose le plus de logements sociaux, suivie par les Hauts-de-France, alors que c'est l'Occitanie qui en propose

le moins en moyenne. Ce qui pose déjà un problème, car la théorie que nous avons exposée au début nécessite de poser une hypothèse selon laquelle le nombre de logements est également réparti sur toute la zone étudiée (la France).

Au niveau des prix moyens au  $m^2$ , on observe également une forte dépendance à la région. Cela contredit également la théorie de base dans la mesure où c'est dans l'Île-de-France que l'on observe le prix moyen par  $m^2$  le plus élevé.

## 2.4. Densité de population et taux de logements sociaux

On peut supposer que la densité de population est intimement liée au taux de logements sociaux. En effet, il semble cohérent de penser que plus la densité de population d'une région est élevée, plus le besoin de logements abordables se fera sentir. Il est donc raisonnable de penser que le taux de logements sociaux augmentera également.

En examinant les données, on constate qu'il existe effectivement une corrélation entre ces deux variables. La corrélation est positive, ce qui signifie que plus la densité de population est élevée, plus le taux de logements sociaux l'est également. Cette corrélation s'explique par le fait que les zones urbaines denses sont souvent caractérisées par des prix de l'immobilier plus élevés, ce qui peut rendre l'accès au logement difficile pour les personnes à faibles revenus. Les logements sociaux peuvent donc constituer une solution pour répondre à ce besoin.

## 2.5. Les jeunes et l'environnement : une préoccupation grandissante en matière de logements ?

L'écologie est devenue une préoccupation majeure pour une part croissante de la population, notamment chez les jeunes. On a notamment été témoin de l'émergence du terme "éco-anxiété". Ce terme, désignant « l'ensemble des émotions ressenties en réaction aux bouleversements environnementaux actuels et futurs. »<sup>1</sup>, est témoin d'une angoisse globale. Il est donc légitime de s'attendre à une augmentation de la prise de conscience environnementale, en particulier chez les jeunes, qui sont souvent les plus sensibles à ces questions.

Cependant, les résultats d'une étude sur la corrélation entre l'âge de la population et le nombre de logements énergivores dans les parcs sociaux ne confirment pas cette hypothèse. En effet, la corrélation entre une population jeune et la part de logements sociaux énergivores n'est que de 8%, ce qui signifie que les départements avec une population plus jeune ont tendance à avoir plus de logements sociaux énergivores. De même, la corrélation entre une population plus âgée et le nombre de logements énergivores dans les parcs sociaux est insignifiante. Ainsi, la proportion de personnes âgées dans un département n'affecte pas le nombre de logements sociaux ayant un impact énergétique important.

Néanmoins, l'âge du parc social dans lequel se trouvent les logements sociaux est un facteur bien plus déterminant. En effet, la corrélation entre l'âge moyen des parcs sociaux et leur impact énergétique est de 37%. Autrement dit, une augmentation d'un an de l'âge moyen des parcs sociaux entraîne une augmentation de 37% de l'impact énergétique de ces logements.

En conclusion, si l'âge de la population a une certaine influence sur l'impact énergétique des logements sociaux, celle-ci est limitée. L'âge du parc social est en revanche un facteur clé, qui doit être pris en compte dans toute stratégie visant à améliorer l'efficacité énergétique des logements sociaux.

---

1. Définition d'après Décathlon, <https://engagements.decathlon.fr/quest-ce-que-leco-anxiete>

### **3. L'équilibre démographique : un déterminant dans le taux de logements sociaux ?**

Le nombre de logements sociaux au sein d'une population est un enjeu majeur pour les politiques publiques car étroitement liées à la question de l'accès au logement pour les populations les plus défavorisées. C'est une des raisons pour laquelle il faut s'interroger sur les facteurs influençant le taux de logements sociaux en France. C'est un fait, il est nécessaire que le nombre de logements (sociaux ou non) disponibles suive la réalité démographique, au risque de ne plus pouvoir héberger correctement la population. Cette réalité peut être analysée grâce au solde naturel et au solde migratoire. La question principale ici est de savoir si l'équilibre entre les deux soldes impacte ou non le taux de logements sociaux en France.

#### **3.1. Solde naturel et solde migratoire selon les régions**

On constate qu'il peut exister une différence significative entre le solde naturel et le solde migratoire selon les régions, mais pas toutes. Parmi celles qui ne présentent pas de différence, nous pouvons citer : l'Auvergne-Rhône-Alpes, la Provence-Alpes-Côte d'Azur, la Bretagne, la Normandie, le Centre-Val de Loire, la Corse, les Pays de la Loire et la Bourgogne-Franche-Comté. Les régions qui présentent une différence significative sont : les Hauts-de-France, le Grand Est, l'Île-de-France ; les DOM-TOM ; l'Occitanie et la Nouvelle-Aquitaine.

#### **3.2. Est-ce qu'une différence significative entre les 2 soldes signifie un plus ou moins grand taux de logements sociaux ?**

Pour montrer qu'une différence entre les 2 soldes impacte ou non le taux de logements sociaux, nous avons décidé d'utiliser à la fois une région qui ne présente pas de différence entre ses soldes et une autre qui en présente une. En cela, nous allons utiliser la région de l'Auvergne-Rhône-Alpes et de la Nouvelle Aquitaine car ce sont les régions des deux groupes qui sont composés d'un plus grand nombre de départements.

Finalement, si une région affiche une différence significative entre son solde naturel et son solde migratoire, il y a des chances pour que le taux de logements sociaux soit significativement différent de celui d'une région qui n'affiche pas une différence significative des soldes. Par exemple, la région Auvergne-Rhône-Alpes qui ne présente pas de différence de ses soldes, a en moyenne 14% de ses logements qui sont des logements sociaux, alors que la région Nouvelle-Aquitaine présente une différence significative entre ses soldes et a en moyenne 10% de ses logements sont des logements sociaux.



## 4. Situation économique française face au nombre et au prix au $m^2$ des logements sociaux

Intéressons-nous maintenant à la situation économique du pays. Nous allons démontrer dans cette partie si la situation macroéconomique de la France en 2019 a impacté le prix des logements sociaux, ou plus exactement le prix du  $m^2$  des logements sociaux. Ainsi, nous avons étudié si le salaire net mensuel, le taux de pauvreté et le taux de chômage au 4e trimestre étaient indépendants dans la fixation du prix du  $m^2$  des logements sociaux.

### 4.1. Les salaires confrontés au loyer mensuel d'un $m^2$

Collectivement, nous pouvons penser que le prix d'un logement est plus élevé en fonction de sa localisation géographique. En général, les logements sont plus chers en Île-de-France que dans le Grand Est. Une des explications possibles serait que les salaires diffèrent selon les régions. C'est pour cela que nous avons regardé s'il existait réellement un lien entre le salaire net mensuel moyen des régions et le prix moyen du  $m^2$  des logements sociaux.

Nous avons alors décelé une corrélation monotone significative à hauteur de 57,4 % entre le salaire net mensuel moyen des régions et le prix moyen du  $m^2$  des logements sociaux. Autrement dit, lorsque le salaire net mensuel moyen augmente, le prix moyen du  $m^2$  des logements sociaux connaît une augmentation de 57,4 % de l'augmentation qu'a connue le salaire net mensuel moyen.

### 4.2. Taux de pauvreté

Le taux de pauvreté est un indicateur statistique qui indique la proportion de personnes dont le revenu est inférieur à un certain seuil considéré comme insuffisant pour subvenir à leurs besoins de base. En France, le seuil de pauvreté est défini comme étant égal à 60 % du niveau de vie<sup>2</sup> médian de la population française.

Ainsi, le taux de pauvreté est défini en fonction du niveau de vie de la population. De plus, l'accès au logement est pris en compte dans le calcul du niveau de vie. C'est pourquoi nous allons voir s'il existe une corrélation entre le taux de pauvreté et le prix moyen du  $m^2$  des logements sociaux. En effet, nous pouvons penser que plus le taux de pauvreté est élevé, moins le prix au  $m^2$  est faible pour diverses raisons, dont une selon laquelle l'État doit s'assurer une facilité d'accès au logement. Cependant, aucune corrélation significative n'a été observée entre le taux de pauvreté et le prix moyen du  $m^2$  des logements sociaux.

### 4.3. Le taux de chômage et le prix du $m^2$

Le taux de chômage est un indicateur économique qui mesure la proportion de la population active en âge de travailler qui n'a pas d'emploi. Les personnes sans emploi ne perçoivent donc pas de salaire, ce qui impacte à la baisse leur revenu disponible. Nous pourrions alors penser que le prix du  $m^2$  des logements sociaux pourrait s'adapter à la situation de chômage en diminuant. Cependant, Nous n'avons pas trouvé de corrélation significative entre le taux de chômage et le prix moyen du  $m^2$  des logements sociaux.

---

2. Le niveau de vie correspond au revenu disponible par unité de consommation, qui tient compte de la taille et de la composition du ménage, après déduction des impôts et des transferts sociaux (prestations sociales, allocations familiales, etc.).

Pour conclure cette partie, il semble que le prix des logements ne soit pas en phase avec la situation économique du pays. Le salaire net mensuel semble être le seul facteur économique présentant une corrélation (partielle) avec le prix des logements, le taux de pauvreté et le taux de chômage n'en montrent aucun.

## 5. Conclusion

Nous pouvons donc conclure que le nombre de logements et leur prix moyen du mètre carré sont dépendants de la démographie française (en incluant le solde migratoire et naturel). De plus, nous avons vu que le taux de logements sociaux énergivores est d'autant plus élevé que l'âge des parcs sociaux est élevé, alors que nous pensions que c'était l'âge des occupants qui expliquait la consommation énergétique des logements sociaux. La démographie explique mieux le nombre et la part des logements sociaux que l'économie du pays (hormis le salaire net moyen) n'explique la fixation du prix moyen du mètre carré des logements sociaux.

## A. Annexe - partie 2

### A.1. Corrélation entre la proportion de moins de 20 ans et la proportion des 60 ans et plus par département.

La figure (4) suggère une potentielle corrélation négative entre ses deux variables. Nous allons l'étudier afin de confirmer ou non cette hypothèse.

```
##  
## Shapiro-Wilk normality test  
##  
## data: Z  
## W = 0.76, p-value = 1.735e-11
```

La p-value du test de Shapiro-Wilk est inférieure au seuil de 5%. Par conséquent, on rejette l'hypothèse nulle selon laquelle les données suivent une distribution normale. On doit alors utiliser un test de corrélation de Spearman, plutôt qu'un test de corrélation de Pearson.

```
##  
## Spearman's rank correlation rho  
##  
## data: don$population_de_moins_de_20_ans and don$population_de_60_ans_et_plus  
## S = 312089, p-value < 2.2e-16  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## -0.872724
```

Le test de Spearman a une p-value inférieure à nouveau au seuil de significativité de 5%. Nous pouvons donc conclure à l'existence d'une assez forte corrélation négative entre ces deux variables, comme en témoignent le coefficient de corrélation rho de Spearman qui est négatif.

### A.2. Corrélation entre la proportion d'une population d'un certain âge et le prix moyen au $m^2$

La figure (5) suggère qu'il existe un lien entre la proportion d'avoir moins de 20 ans ou plus de 60 ans et le prix moyen au  $m^2$ .

```
##  
## Shapiro-Wilk normality test  
##  
## data: Z  
## W = 0.84014, p-value = 5.211e-09  
  
##  
## Shapiro-Wilk normality test  
##  
## data: Z  
## W = 0.88856, p-value = 4.258e-07
```

Les p-values des deux testes de Shapiro-Wilk sont inférieures au seuil de 5%. Par conséquent, on rejette l'hypothèse nulle selon laquelle les données suivent une distribution normale. Nous ne pouvons donc pas effectuer le teste de Pearson, nous allons exécuter le test de Spearman.

Regardons d'abord les résultats avec la proportion de personnes de moins de 20 ans par département.

```
##
## Spearman's rank correlation rho
##
## data: don$population_de_moins_de_20_ans and don$parc_social_loyer_moyen_en_eur_m2_mois
## S = 89371, p-value = 1.18e-06
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4637229
```

Les p-value est inférieure au seuil de significativité de 5%. On accepte donc l'hypothèse alternative selon laquelle il y a une corrélation significative entre les deux. De plus, le Rho de Spearman est positif, on conclut donc qu'il existe une corrélation positive entre la proportion de la population ayant moins de 20 ans et le prix moyen du loyer au  $m^2$  par département.

```
##
## Spearman's rank correlation rho
##
## data: don$parc_social_loyer_moyen_en_eur_m2_mois and don$population_de_60_ans_et_plus
## S = 281284, p-value = 2.66e-15
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.6878734
```

Les conclusions sont les mêmes que celles annoncées pour la variable “taux de personnes de moins de 20 ans”. Il convient de préciser que le Rho de Spearman est supérieure pour la variable “taux de personnes de 60 ans et plus” comparé à celle obtenue avec la variable “taux de personnes de moins de 20 ans”. Cela signifie que la corrélation entre le prix moyen du  $m^2$  et la proportion de la population âgée de 60 ans et plus est plus forte que celle avec la proportion de la population âgée de moins de 20 ans.

### A.3. Etude du lien entre le prix moyen et le nombre de logements

Dans un premier, nous avons regardé le nombre de logements sociaux, en groupant les données par régions.

```
## -----
## parc_social_nombre_de_logements ~ nom_region (don2)
##
## Summary:
## n pairs: 96, valid: 96 (100.0%), missings: 0 (0.0%), groups: 13
##
##
```

##	Auvergne-Rhone-Alpes	Bourgogne-France-Comte
## mean	46'192.083	26'520.250
## median	30'954.000	17'603.500
## sd	44'673.382	17'240.795
## IQR	29'816.500	24'741.750
## n	12	8
## np	12.500%	8.333%
## NAs	0	0
## Os	0	0
##		
##	Bretagne	Centre-Val de Loire
## mean	39'833.750	29'523.000
## median	41'753.500	26'627.500
## sd	26'381.848	17'775.622
## IQR	23'503.250	26'499.000
## n	4	6
## np	4.167%	6.250%
## NAs	0	0
## Os	0	0
##		
##	Corse	Grand Est
## mean	54'357.000	56'873.700
## median	54'357.000	42'599.500
## sd	45'821.934	56'383.545
## IQR	32'401.000	47'093.250
## n	2	10
## np	2.083%	10.417%
## NAs	0	0
## Os	0	0
##		
##	Hauts-de-France	Ile-de-France
## mean	115'203.200	142'993.750
## median	66'189.000	120'936.500
## sd	98'617.305	59'969.492
## IQR	119'396.000	73'111.250
## n	5	8
## np	5.208%	8.333%
## NAs	0	0
## Os	0	0
##		
##	Normandie	Nouvelle-Aquitaine
## mean	51'804.600	27'985.000
## median	37'708.000	18'376.500
## sd	51'667.096	26'267.438
## IQR	21'119.000	22'943.000
## n	5	12
## np	5.208%	12.500%
## NAs	0	0
## Os	0	0
##		
##	Occitanie	Pays de la Loire

```

## mean          13'217.385          45'782.000
## median        8'361.000          38'660.000
## sd            10'845.786          29'608.950
## IQR           11'769.000          42'375.000
## n              13                  5
## np            13.542%             5.208%
## NAs           0                   0
## Os            0                   0
##
##      Provence-Aples-Cote d'Azur
## mean          57'737.500
## median        47'304.000
## sd            51'216.517
## IQR           15'588.000
## n              6
## np            6.250%
## NAs           0
## Os            0
##
## Kruskal-Wallis rank sum test:
##   Kruskal-Wallis chi-squared = 40.048, df = 12, p-value = 7.061e-05

```

social\_nombre\_de\_logements ~ nom\_region (d

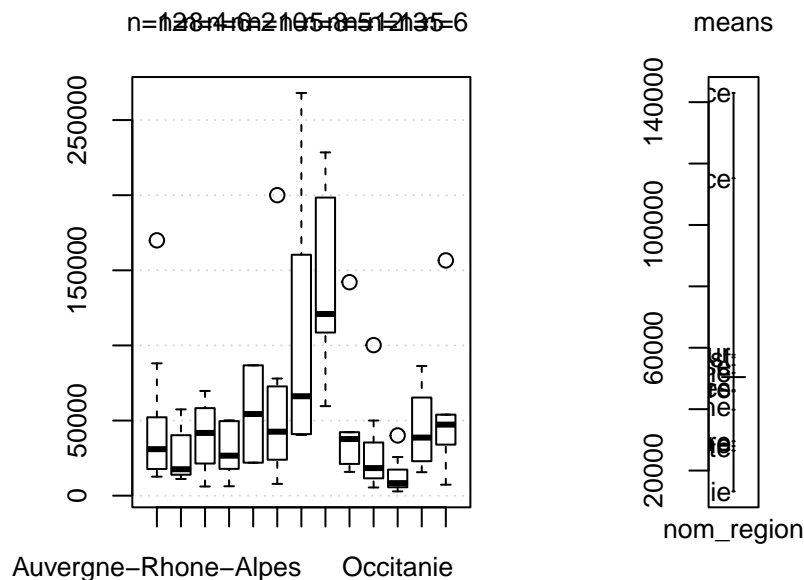


FIGURE 6 – Distribution du nombre de logements sociaux pour chaque région, et représentation de la moyenne du nombre de logements sociaux pour chaque région.

On constate que la moyenne du nombre de logements sociaux varie énormément d'une région à une autre. Le maximum est atteint, sans surprise, en Île-de-France, avec une moyenne de 142993.8 loge-

ments sociaux par département dans cette région. Le minimum de nombre de logements sociaux par département se trouve dans la région de l'Occitanie, avec 13217.4 logements sociaux en moyenne par département. On constate un écart assez important entre le minimum et le maximum, mais on remarque aussi sur le graphique (6) que deux valeurs se distinguent de la tendance générale : la plupart des moyennes se trouvent en dessous de 6 000 logements sociaux en moyenne par région, à l'exception de deux, qui dépassent les 110 000. Regardons désormais la distribution de ses données.

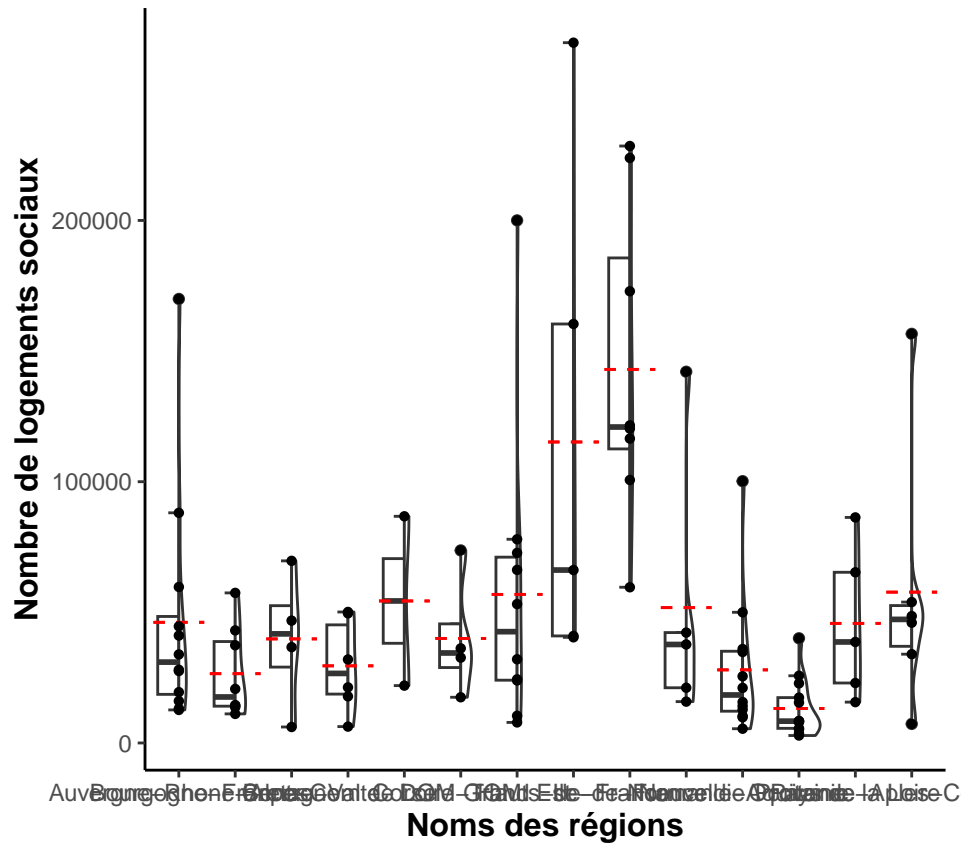


FIGURE 7 – Distribution du nombre de logements sociaux pour chaque région.

Aucune des distributions ne semble suivre une loi normale. Nous vérifions cette hypothèse à partir du test de Shapiro.

```
## $'Auvergne-Rhone-Alpes'
##
## Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.7, p-value = 0.002

## $'Auvergne-Rhone-Alpes'
##
## Anderson-Darling normality test
##
## data:  X[[i]]
## A = 1, p-value = 0.002

## $'Auvergne-Rhone-Alpes'
```



```
##
##  Jarque-Bera Normality Test
##
## data:  X[[i]]
## JB = 12, p-value = 0.002
## alternative hypothesis: greater
```

On rejette l'hypothèse de normalité pour au moins une sous population. Nous poursuivons en testant l'homoscédasticité et l'égalité des medianes.

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  parc_social_nombre_de_logements by nom_region
## Fligner-Killeen:med chi-squared = 18, df = 13, p-value = 0.2

##
##  Kruskal-Wallis rank sum test
##
## data:  parc_social_nombre_de_logements by nom_region
## Kruskal-Wallis chi-squared = 41, df = 13, p-value = 0.0001
```

La p-value du test de Fligner-Killeen est inférieure au seuil, on conclut que les variances entre les groupes ne sont pas significativement différentes. On conserve donc l'hypothèse d'homoscédasticité. De plus, le test de Kruskal-Wallis nous permet de conclure que les régions ont des distributions différentes en ce qui concerne le nombre de logements dans les parcs sociaux

```
##
##  Pairwise comparisons using Wilcoxon rank sum exact test
##
## data:  don$parc_social_nombre_de_logements and don$nom_region
##
##                               Auvergne-Rhone-Alpes Bourgogne-France-Comte Bretagne
## Bourgogne-France-Comte      1.000                -                -
## Bretagne                    1.000                1.000            -
## Centre-Val de Loire         1.000                1.000            1.000
## Corse                      1.000                1.000            1.000
## DOM-TOM                    1.000                1.000            1.000
## Grand Est                  1.000                1.000            1.000
## Hauts-de-France            1.000                1.000            1.000
## Ile-de-France              0.061                0.014            0.646
## Normandie                  1.000                1.000            1.000
## Nouvelle-Aquitaine         1.000                1.000            1.000
## Occitanie                   0.082                1.000            1.000
## Pays de la Loire           1.000                1.000            1.000
## Provence-Aples-Cote d'Azur 1.000                1.000            1.000
##
##                               Centre-Val de Loire Corse DOM-TOM Grand Est
## Bourgogne-France-Comte      -                -      -      -
## Bretagne                    -                -      -      -
## Centre-Val de Loire         -                -      -      -
```

```

## Corse 1.000 - - -
## DOM-TOM 1.000 1.000 - -
## Grand Est 1.000 1.000 1.000 -
## Hauts-de-France 1.000 1.000 1.000 1.000
## Ile-de-France 0.058 1.000 0.646 0.364
## Normandie 1.000 1.000 1.000 1.000
## Nouvelle-Aquitaine 1.000 1.000 1.000 1.000
## Occitanie 1.000 1.000 1.000 0.272
## Pays de la Loire 1.000 1.000 1.000 1.000
## Provence-Aples-Cote d’Azur 1.000 1.000 1.000 1.000
## Hauts-de-France Ile-de-France Normandie
## Bourgogne-France-Comte - - -
## Bretagne - - -
## Centre-Val de Loire - - -
## Corse - - -
## DOM-TOM - - -
## Grand Est - - -
## Hauts-de-France - - -
## Ile-de-France 1.000 - -
## Normandie 1.000 1.000 -
## Nouvelle-Aquitaine 0.504 0.003 1.000
## Occitanie 0.021 0.0009 1.000
## Pays de la Loire 1.000 0.504 1.000
## Provence-Aples-Cote d’Azur 1.000 0.962 1.000
## Nouvelle-Aquitaine Occitanie Pays de la Loire
## Bourgogne-France-Comte - - -
## Bretagne - - -
## Centre-Val de Loire - - -
## Corse - - -
## DOM-TOM - - -
## Grand Est - - -
## Hauts-de-France - - -
## Ile-de-France - - -
## Normandie - - -
## Nouvelle-Aquitaine - - -
## Occitanie 1.000 - -
## Pays de la Loire 1.000 0.755 -
## Provence-Aples-Cote d’Azur 1.000 0.719 1.000
##
## P value adjustment method: holm

```

*Lecture du tableau* : la comparaison entre la région Île-de-France et la région Bretagne a une p-value corrigée de 0,01383, ce qui est inférieure au seuil de 5%, donc il y a une différence significative dans le nombre de logements dans les parcs sociaux entre ces deux régions.

#### A.4. Densité de population et taux de logements sociaux

Commençons par visualiser nos données.

D’après la figure (8), il ne semble pas exister de corrélation entre les deux variables.

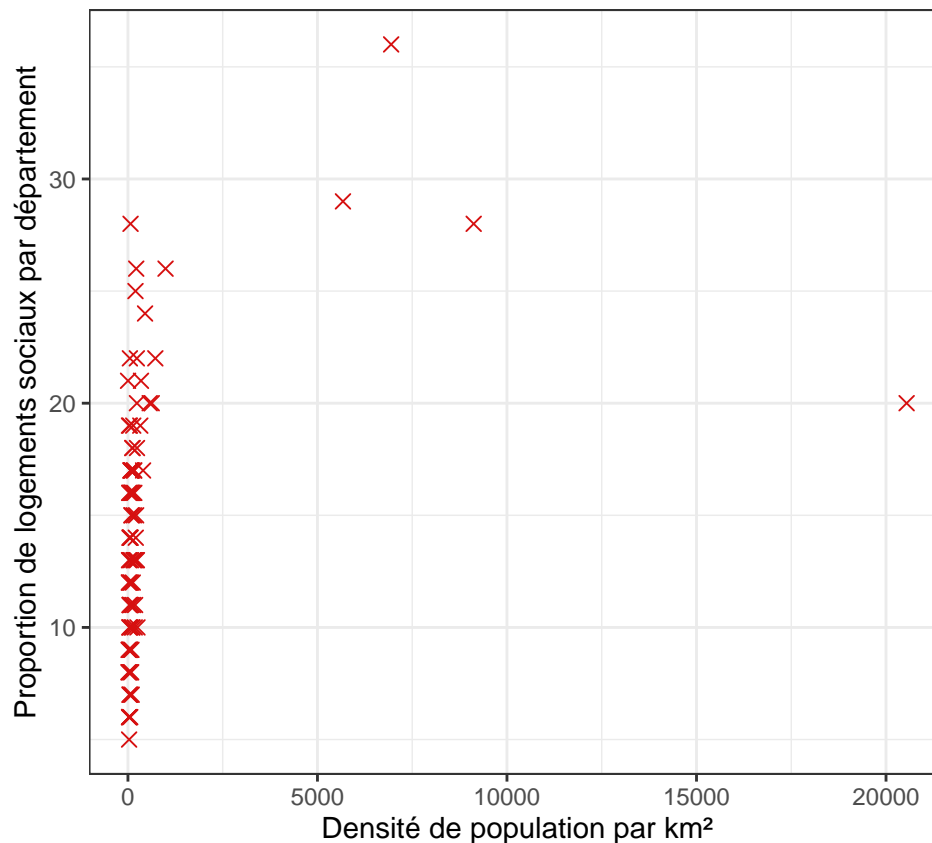


FIGURE 8 – Proportion de logements sociaux en fonction de la densité de population par région.

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.3, p-value <0.0000000000000002
```

Les p-values des deux tests de Shapiro-Wilk sont inférieures au seuil de 5%. Par conséquent, on rejette l'hypothèse nulle selon laquelle les données suivent une distribution normale. Nous ne pouvons donc pas effectuer le teste de Pearson, nous allons donc faire le test de Kendall.

```
##
## Spearman's rank correlation rho
##
## data: don$taux_de_logements_sociaux_en and don$densite_de_population_au_km2
## S = 70253, p-value = 0.0000000003
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.6
```

La p-value du test est inferieure au seuil de 5%, on accepte donc H1, hypothèse selon laquelle il y a une corrélation significative entre les deux variables. on conclut donc qu'il existe une corrélation positive entre le taux de logements sociaux par departements avec la densité de population au  $m^2$ , puisque le Rho de Spearman est positif.

## A.5. La part des logements sociaux énergivores

Commençons par une analyse descriptive de nos données.

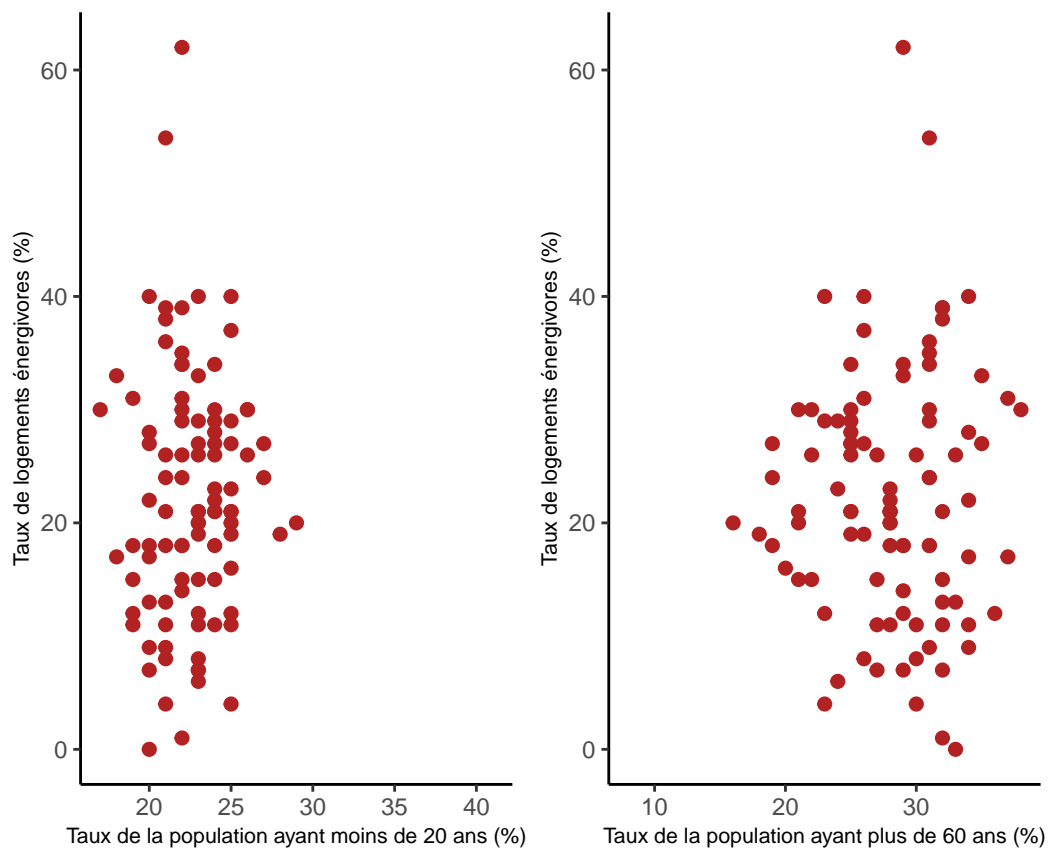


FIGURE 9 – Relation entre le taux de logements sociaux énergivores et la proportion de la population ayant moins de 20 ans (à gauche) et la proportion de la population ayant 60 ans et plus (à droite). *Lecture* : Un département dans lequel 30% de la population est âgée de moins de 20 ans possède environ 20% des logements sociaux qui sont énergivores, en 2019, en France.

À partir de ses graphiques, on peut émettre une hypothèse de corrélation entre le taux de logements énergivores et le taux de personnes de moins de 20 ans, mais il est plus difficile d'envisager une corrélation entre le taux de logements énergivores et le taux de personnes de plus de 60 ans. Regardons la normalité des couples, en commençant avec le taux de la population de moins de 20ans.

```
##
## Shapiro-Wilk normality test
##
## data:  Z
## W = 1, p-value = 0.03
```

Le second test, se concentre sur la proportion de la population âgée de 60 ans et plus.

```
##
## Shapiro-Wilk normality test
##
## data:  Z
## W = 1, p-value = 0.05
```

Dans les deux cas, les p-values sont inférieures au seuil de 5%, on rejette donc l'hypothèse de normalité. On fait donc un test de Spearman.

```
##
## Spearman's rank correlation rho
##
## data: don$parc_social_taux_de_logements_energivores_e_f_g_en and don$population_de_moins_d
## S = 135113, p-value = 0.4
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.08

##
## Spearman's rank correlation rho
##
## data: don$parc_social_taux_de_logements_energivores_e_f_g_en and don$population_de_60_ans_
## S = 151087, p-value = 0.8
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.02
```

Pour les deux variables, on accepte l'hypothèse de non-corrélation monotone.

## A.6. Age moyen des parcs sociaux et l'environnement

Nous avons tout d'abord souhaité regarder le nuage de points de nos données (cf figure (10)).

Nous avons ensuite testé la normalité.

```
##
## Shapiro-Wilk normality test
##
## data: t(parc_energ)
## W = 1, p-value = 0.00006
```

La p-value est inférieure au seuil de significativité, on rejette donc l'hypothèse de normalité. Nous allons donc effectuer le test de Spearman.

```
##
## Spearman's rank correlation rho
##
## data: don$parc_social_taux_de_logements_energivores_e_f_g_en and don$parc_social_age_moyen
## S = 92681, p-value = 0.0002
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.4
```

La p-value du test de Spearman est inférieure au seuil de 5%, on conclut alors à une corrélation significative entre les deux variables.

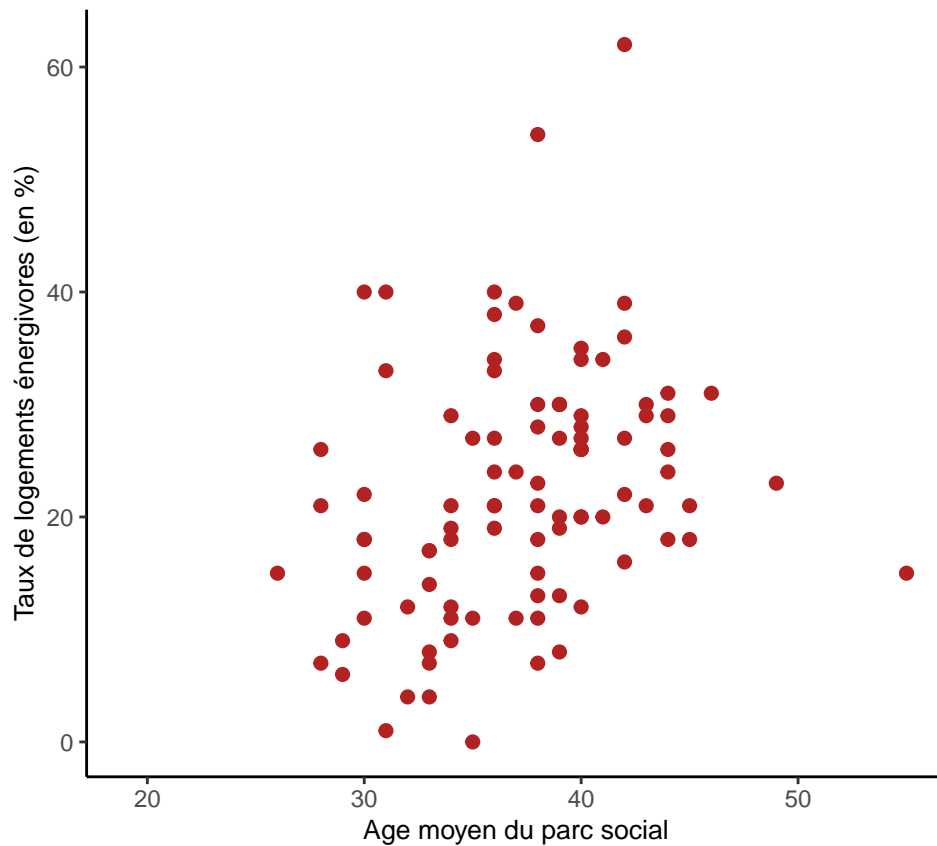


FIGURE 10 – Relation entre le taux de logements sociaux économes (en %) et l'âge moyen du parc social (en année) *Lecture* : En moyenne, les parcs sociaux d'un département sont âgés de 49 ans dans lesquels environ 22% des logements sociaux sont des logements sociaux économes.

## B. Annexe - partie 3

### B.1. Solde migratoire et solde naturel

Nous testons d'abord si nos données suivent une distribution normale.

```
## [1] "Résultats des tests de normalité pour le solde naturel de la région Auvergne-Rhone-Alp
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 1, p-value = 0.7
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.3, p-value = 0.6
##
```

```
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data: solde_nat
## JB = 0.6, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région Hauts-de-France est insu
## [1] "Résultats des tests de normalité pour le solde naturel de la région Hauts-de-France"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data: solde_nat
## W = 0.8, p-value = 0.06
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data: solde_nat
## A = 0.3, p-value = 0.6
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data: solde_nat
## JB = 0.7, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région Provence-Aples-Cote d'Az
## [1] "Résultats des tests de normalité pour le solde naturel de la région Provence-Aples-Cot
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data: solde_nat
## W = 0.9, p-value = 0.5
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data: solde_nat
## A = 0.3, p-value = 0.6
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
```

```
##
## data:  solde_nat
## JB = 0.6, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde naturel de la région Grand Est"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.7, p-value = 0.002
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_nat
## A = 1, p-value = 0.005
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 8, p-value = 0.02
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde naturel de la région Occitanie"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.9, p-value = 0.3
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.4, p-value = 0.3
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 0.9, p-value = 0.6
## alternative hypothesis: greater
##
```



```
## [1] "Résultats des tests de normalité pour le solde naturel de la région Bourgogne-France-C
## [1] "Test de Shapiro-Wilk"
##
##  Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.9, p-value = 0.6
##
## [1] "Test d'Anderson-Darling"
##
##  Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.3, p-value = 0.4
##
## [1] "Test de Jarque-Bera"
##
##  Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 0.09, p-value = 1
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région Bretagne est insuffisant
## [1] "Résultats des tests de normalité pour le solde naturel de la région Bretagne"
## [1] "Test de Shapiro-Wilk"
##
##  Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.9, p-value = 0.6
##
## [1] "Test d'Anderson-Darling"
##
##  Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.3, p-value = 0.4
##
## [1] "Test de Jarque-Bera"
##
##  Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 0.2, p-value = 0.9
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région Normandie est insuffisan
## [1] "Résultats des tests de normalité pour le solde naturel de la région Normandie"
## [1] "Test de Shapiro-Wilk"
##
```

```
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.8, p-value = 0.06
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.3, p-value = 0.4
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 0.7, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde naturel de la région Nouvelle-Aquitaine"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.9, p-value = 0.2
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.5, p-value = 0.2
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 2, p-value = 0.4
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région Centre-Val de Loire est"
## [1] "Résultats des tests de normalité pour le solde naturel de la région Centre-Val de Loir"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.9, p-value = 0.2
```

```
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data: solde_nat
## A = 0.5, p-value = 0.2
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data: solde_nat
## JB = 0.8, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région Corse est insuffisant po
## [1] "Le nombre d'observations pour le solde naturel dans la région Corse est insuffisant po
## [1] "Résultats des tests de normalité pour le solde naturel de la région Corse"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data: solde_nat
## W = 0.9, p-value = 0.2
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data: solde_nat
## A = 0.5, p-value = 0.2
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data: solde_nat
## JB = 0.3, p-value = 0.8
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde naturel de la région Ile-de-France"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data: solde_nat
## W = 0.9, p-value = 0.5
##
## [1] "Test d'Anderson-Darling"
##
```

```
## Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.4, p-value = 0.3
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 0.4, p-value = 0.8
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région DOM-TOM est insuffisant"
## [1] "Résultats des tests de normalité pour le solde naturel de la région DOM-TOM"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.9, p-value = 0.4
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_nat
## A = 0.4, p-value = 0.3
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_nat
## JB = 0.5, p-value = 0.8
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde naturel dans la région Pays de la Loire est insuffisant"
## [1] "Résultats des tests de normalité pour le solde naturel de la région Pays de la Loire"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_nat
## W = 0.9, p-value = 0.2
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_nat
```

```
## A = 0.4, p-value = 0.3
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data: solde_nat
## JB = 0.5, p-value = 0.8
## alternative hypothesis: greater
```

Pour au moins une sous population on rejette l'hypothèse de normalité. On regarde désormais pour la variable du solde migratoire.

```
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Auvergne-Rhone-
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data: solde_mig
## W = 0.9, p-value = 0.5
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data: solde_mig
## A = 0.3, p-value = 0.5
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data: solde_mig
## JB = 0.4, p-value = 0.8
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde migratoire dans la région Hauts-de-France est i
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Hauts-de-France
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data: solde_mig
## W = 0.9, p-value = 0.3
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data: solde_mig
## A = 0.3, p-value = 0.5
```

```
##
## [1] "Test de Jarque-Bera"
##
##   Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.4, p-value = 0.8
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde migratoire dans la région Provence-Aples-Cote d'Azur"
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Provence-Aples-Cote d'Azur"
## [1] "Test de Shapiro-Wilk"
##
##   Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.9, p-value = 0.3
##
## [1] "Test d'Anderson-Darling"
##
##   Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.3, p-value = 0.5
##
## [1] "Test de Jarque-Bera"
##
##   Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.7, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Grand Est"
## [1] "Test de Shapiro-Wilk"
##
##   Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.9, p-value = 0.3
##
## [1] "Test d'Anderson-Darling"
##
##   Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.4, p-value = 0.3
##
## [1] "Test de Jarque-Bera"
##
##   Jarque-Bera Normality Test
```

```
##
## data:  solde_mig
## JB = 1, p-value = 0.5
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Occitanie"
## [1] "Test de Shapiro-Wilk"
##
##  Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 1, p-value = 0.7
##
## [1] "Test d'Anderson-Darling"
##
##  Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.3, p-value = 0.6
##
## [1] "Test de Jarque-Bera"
##
##  Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.7, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Bourgogne-Franc
## [1] "Test de Shapiro-Wilk"
##
##  Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.7, p-value = 0.006
##
## [1] "Test d'Anderson-Darling"
##
##  Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.9, p-value = 0.009
##
## [1] "Test de Jarque-Bera"
##
##  Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 1, p-value = 0.6
## alternative hypothesis: greater
##
```

```
## [1] "Le nombre d'observations pour le solde migratoire dans la région Bretagne est insuffis
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Bretagne"
## [1] "Test de Shapiro-Wilk"
##
##  Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.7, p-value = 0.02
##
## [1] "Test d'Anderson-Darling"
##
##  Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.9, p-value = 0.009
##
## [1] "Test de Jarque-Bera"
##
##  Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.7, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde migratoire dans la région Normandie est insuffi
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Normandie"
## [1] "Test de Shapiro-Wilk"
##
##  Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.8, p-value = 0.09
##
## [1] "Test d'Anderson-Darling"
##
##  Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.9, p-value = 0.009
##
## [1] "Test de Jarque-Bera"
##
##  Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.7, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Nouvelle-Aquita
## [1] "Test de Shapiro-Wilk"
##
```



```
## Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.9, p-value = 0.2
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.5, p-value = 0.2
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 1, p-value = 0.5
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde migratoire dans la région Centre-Val de Loire e
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Centre-Val de L
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.7, p-value = 0.003
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.5, p-value = 0.2
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 3, p-value = 0.2
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde migratoire dans la région Corse est insuffisant
## [1] "Le nombre d'observations pour le solde migratoire dans la région Corse est insuffisant
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Corse"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
```

```
## data:  solde_mig
## W = 0.7, p-value = 0.003
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.5, p-value = 0.2
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.3, p-value = 0.8
## alternative hypothesis: greater
##
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Ile-de-France"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.9, p-value = 0.2
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.5, p-value = 0.1
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 2, p-value = 0.4
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde migratoire dans la région DOM-TOM est insuffisa
## [1] "Résultats des tests de normalité pour le solde migratoire de la région DOM-TOM"
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.9, p-value = 0.4
##
## [1] "Test d'Anderson-Darling"
```

```
##
## Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.5, p-value = 0.1
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.6, p-value = 0.7
## alternative hypothesis: greater
##
## [1] "Le nombre d'observations pour le solde migratoire dans la région Pays de la Loire est
## [1] "Résultats des tests de normalité pour le solde migratoire de la région Pays de la Loir
## [1] "Test de Shapiro-Wilk"
##
## Shapiro-Wilk normality test
##
## data:  solde_mig
## W = 0.8, p-value = 0.05
##
## [1] "Test d'Anderson-Darling"
##
## Anderson-Darling normality test
##
## data:  solde_mig
## A = 0.5, p-value = 0.1
##
## [1] "Test de Jarque-Bera"
##
## Jarque-Bera Normality Test
##
## data:  solde_mig
## JB = 0.8, p-value = 0.7
## alternative hypothesis: greater
```

Pour au moins une sous population on rejette l'hypothèse de normalité. Regardons désormais l'égalité des variances ainsi que l'égalité des moyennes.

```
##
## Bartlett test of homogeneity of variances
##
## data:  dont_contribution_du_solde_naturel_en by nom_region
## Bartlett's K-squared = 34, df = 13, p-value = 0.001
##
## Bartlett test of homogeneity of variances
##
## data:  dont_contribution_du_solde_migratoire_en by nom_region
## Bartlett's K-squared = 32, df = 13, p-value = 0.003
```

```
## [1] "Résultats pour la région Auvergne-Rhone-Alpes"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = -0.6, df = 11, p-value = 0.5
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -4 2
## sample estimates:
## mean difference
## -0.8
##
## [1] "Résultats pour la région Hauts-de-France"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = 5, df = 4, p-value = 0.01
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## 2 8
## sample estimates:
## mean difference
## 5
##
## [1] "Résultats pour la région Provence-Aples-Cote d'Azur"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = -0.4, df = 5, p-value = 0.7
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -5 4
## sample estimates:
## mean difference
## -0.7
##
## [1] "Résultats pour la région Grand Est"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = 3, df = 9, p-value = 0.02
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## 0.6 7.0
## sample estimates:
## mean difference
## 4
```

```
##
## [1] "Résultats pour la région Occitanie"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = -7, df = 12, p-value = 0.00001
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -8 -4
## sample estimates:
## mean difference
## -6
##
## [1] "Résultats pour la région Bourgogne-France-Comte"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = 1, df = 7, p-value = 0.2
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -0.9 3.4
## sample estimates:
## mean difference
## 1
##
## [1] "Résultats pour la région Bretagne"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = -3, df = 3, p-value = 0.07
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -8.1 0.6
## sample estimates:
## mean difference
## -4
##
## [1] "Résultats pour la région Normandie"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = 1, df = 4, p-value = 0.4
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -2 4
## sample estimates:
## mean difference
```

```
##          1
##
## [1] "Résultats pour la région Nouvelle-Aquitaine"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = -3, df = 11, p-value = 0.02
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -6.9 -0.8
## sample estimates:
## mean difference
##          -4
##
## [1] "Résultats pour la région Centre-Val de Loire"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = -0.5, df = 5, p-value = 0.6
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -7  5
## sample estimates:
## mean difference
##          -1
##
## [1] "Résultats pour la région Corse"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = -2, df = 1, p-value = 0.3
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -23  16
## sample estimates:
## mean difference
##          -4
##
## [1] "Résultats pour la région Ile-de-France"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = 4, df = 7, p-value = 0.009
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   3 17
## sample estimates:
```

```
## mean difference
##          10
##
## [1] "Résultats pour la région DOM-TOM"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = 8, df = 3, p-value = 0.005
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##    9 23
## sample estimates:
## mean difference
##          16
##
## [1] "Résultats pour la région Pays de la Loire"
##
## Paired t-test
##
## data: don$dont_contribution_du_solde_naturel_en[don$nom_region == region] and don$dont_con
## t = 0, df = 4, p-value = 1
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   -6  6
## sample estimates:
## mean difference
##           0
```

## B.2. Impact de différence sur taux de logements

Voici d'abord des boîtes à moustaches pour émettre une hypothèse sur la distribution des données.

On teste alors la normalité et, suite à ce test, dont les résultats sont ci-dessous, on rejette cette hypothèse. On teste ensuite l'égalité des variances. Suite aux résultats, on conserve cette hypothèse. On finit cette étude en comparant les moyennes. On conclut en rejetant l'hypothèse nulle selon laquelle la moyenne des deux groupes est significativement identique.

```
## $'Auvergne-Rhone-Alpes'
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
## W = 0.9, p-value = 0.4
##
##
## $'Nouvelle-Aquitaine'
##
## Shapiro-Wilk normality test
##
## data: X[[i]]
```

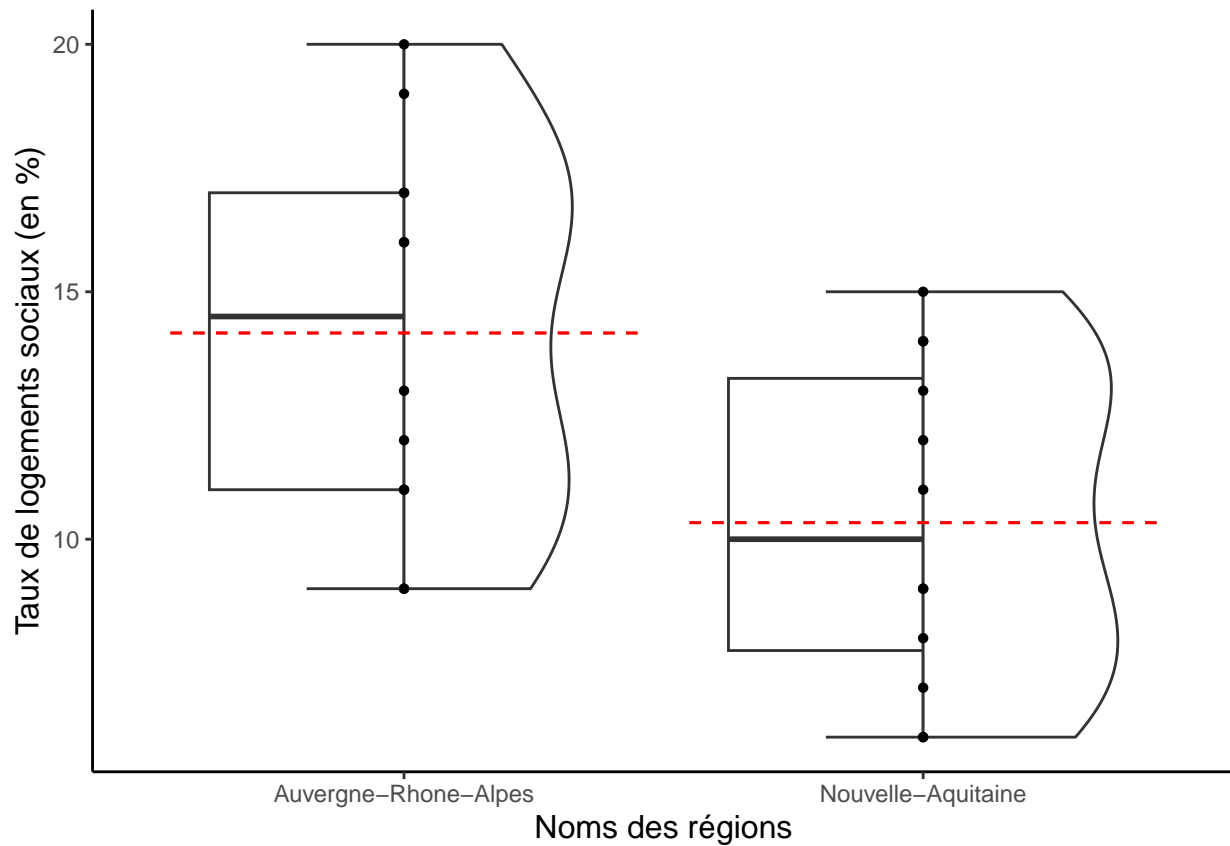


FIGURE 11 – Distribution du taux de logements sociaux en fonction des régions

```
## W = 0.9, p-value = 0.3
```

```
## $'Auvergne-Rhone-Alpes'
##
## Anderson-Darling normality test
##
## data: X[[i]]
## A = 0.4, p-value = 0.4
##
##
## $'Nouvelle-Aquitaine'
##
## Anderson-Darling normality test
##
## data: X[[i]]
## A = 0.4, p-value = 0.4
```

```
## $'Auvergne-Rhone-Alpes'
##
## Jarque-Bera Normality Test
##
## data: X[[i]]
## JB = 0.9, p-value = 0.6
## alternative hypothesis: greater
```



```
##
##
## $'Nouvelle-Aquitaine'
##
## Jarque-Bera Normality Test
##
## data: X[[i]]
## JB = 1, p-value = 0.6
## alternative hypothesis: greater

##
## F test to compare two variances
##
## data: taux_de_logements_sociaux_en by nom_region
## F = 1, num df = 11, denom df = 11, p-value = 0.6
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.4 4.8
## sample estimates:
## ratio of variances
## 1

##
## Two Sample t-test
##
## data: dif$taux_de_logements_sociaux_en by dif$nom_region
## t = 3, df = 22, p-value = 0.01
## alternative hypothesis: true difference in means between group Auvergne-Rhone-Alpes and gro
## 95 percent confidence interval:
## 0.8 6.8
## sample estimates:
## mean in group Auvergne-Rhone-Alpes mean in group Nouvelle-Aquitaine
## 14 10
```

## C. Annexe - partie 4

### C.1. Variable salaire confrontée à la variable loyer

Comme précédemment, on commence par visualiser les données.

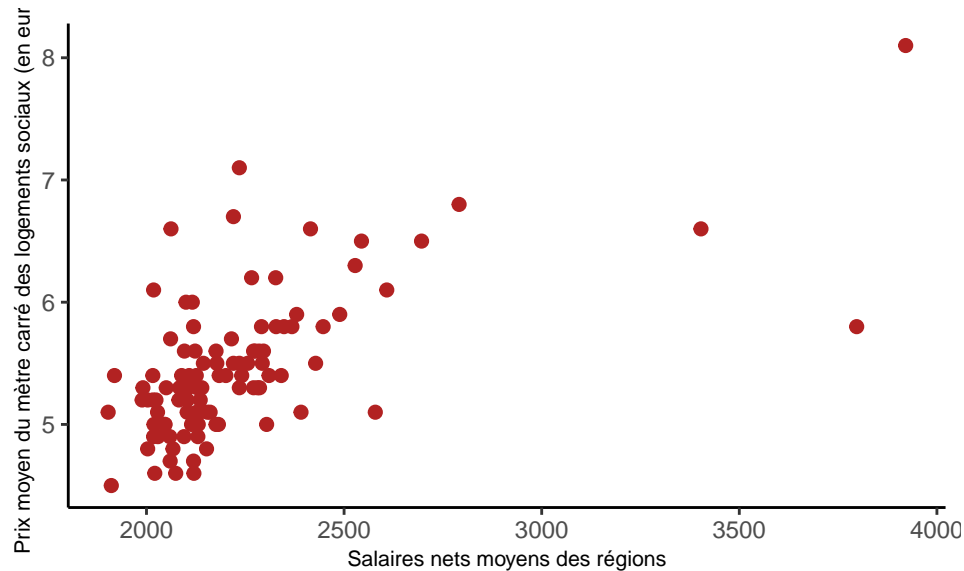


FIGURE 12 – Relation entre le prix moyen du mètre carré moyen des logements et les salaires nets des régions

On peut supposer que lorsque l'une des variables croît, l'autre croît également. Testons la normalité des variables pour adapter le test de corrélation en conséquence.

```
##
## Shapiro-Wilk normality test
##
## data:  Z
## W = 0.7, p-value = 0.000000000003
```

On rejette l'hypothèse de normalité, donc nous utilisons le test de corrélation de Spearman pour évaluer la relation entre les deux variables.

```
##
## Spearman's rank correlation rho
##
## data:  don$parc_social_loyer_moyen_en_eur_m2_mois and don$Salaire_n_mensuel
## S = 70859, p-value = 0.0000000004
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.6
```

Le rho de Spearman (rho) est de 0,57 et la p-value est inférieure au seuil de 5%, ce qui indique que la corrélation est significative. La direction de la corrélation est positive, ce qui signifie qu'il y a une relation croissante entre le prix du loyer mensuel au m<sup>2</sup> et le salaire mensuel net moyen des régions.

## C.2. Taux de pauvreté et prix au carré du loyer

On souhaite désormais confronter ces deux variables. Commençons par les visualiser.

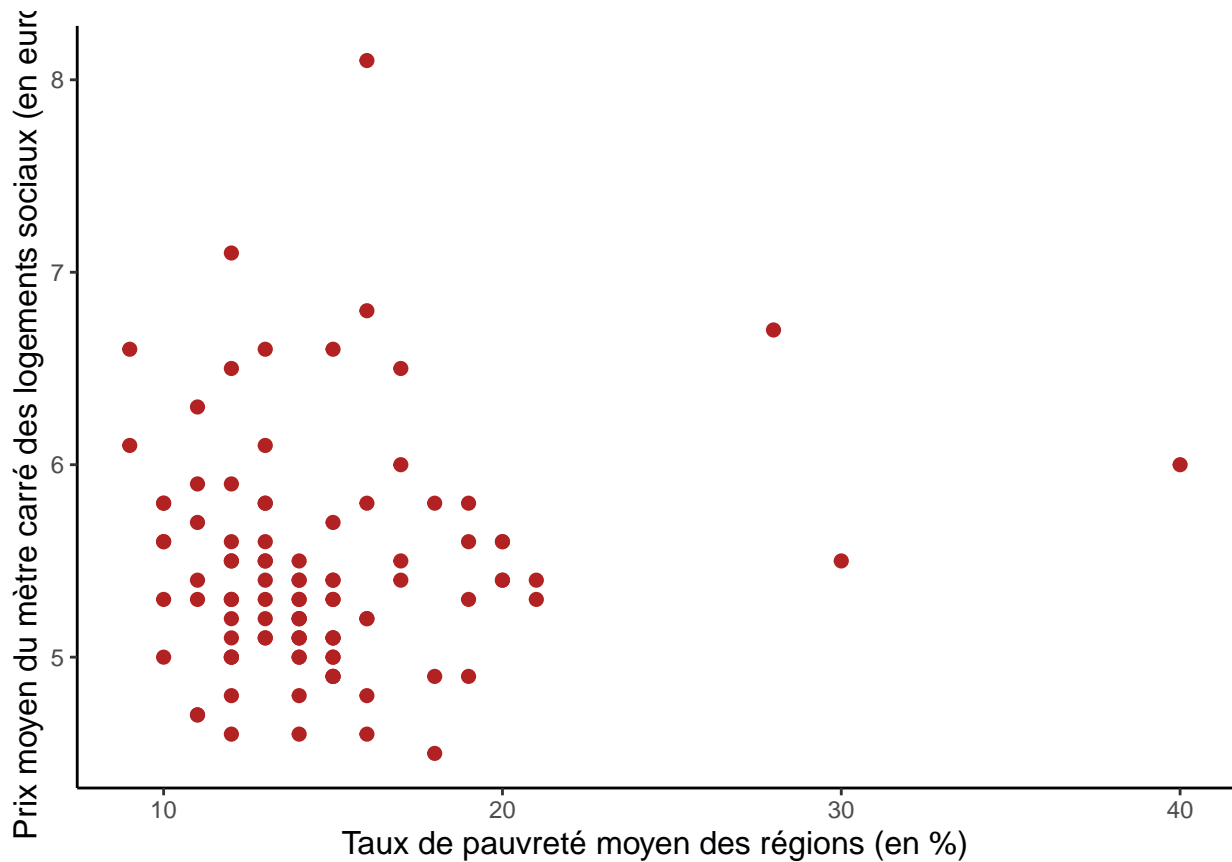


FIGURE 13 – Relation entre le prix moyen du mètre carré moyen des logements sociaux et le taux de pauvreté des régions

À première vue, il n'existe pas de relation linéaire qui lie ces deux variables. Regardons désormais la distribution du couple de ces variables suit une loi normale avec le test de Shapiro-Wilk.

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.8, p-value = 0.000000000003
```

D'après ce test, la p-value est inférieure au seuil, ce qui signifie que nous rejetons l'hypothèse nulle de la normalité de la distribution du couple. On teste donc la corrélation avec le test de Spearman.

```
##
## Spearman's rank correlation rho
##
## data: don$parc_social_loyer_moyen_en_eur_m2_mois and don$taux_de_pauvrete_en
## S = 160548, p-value = 0.8
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

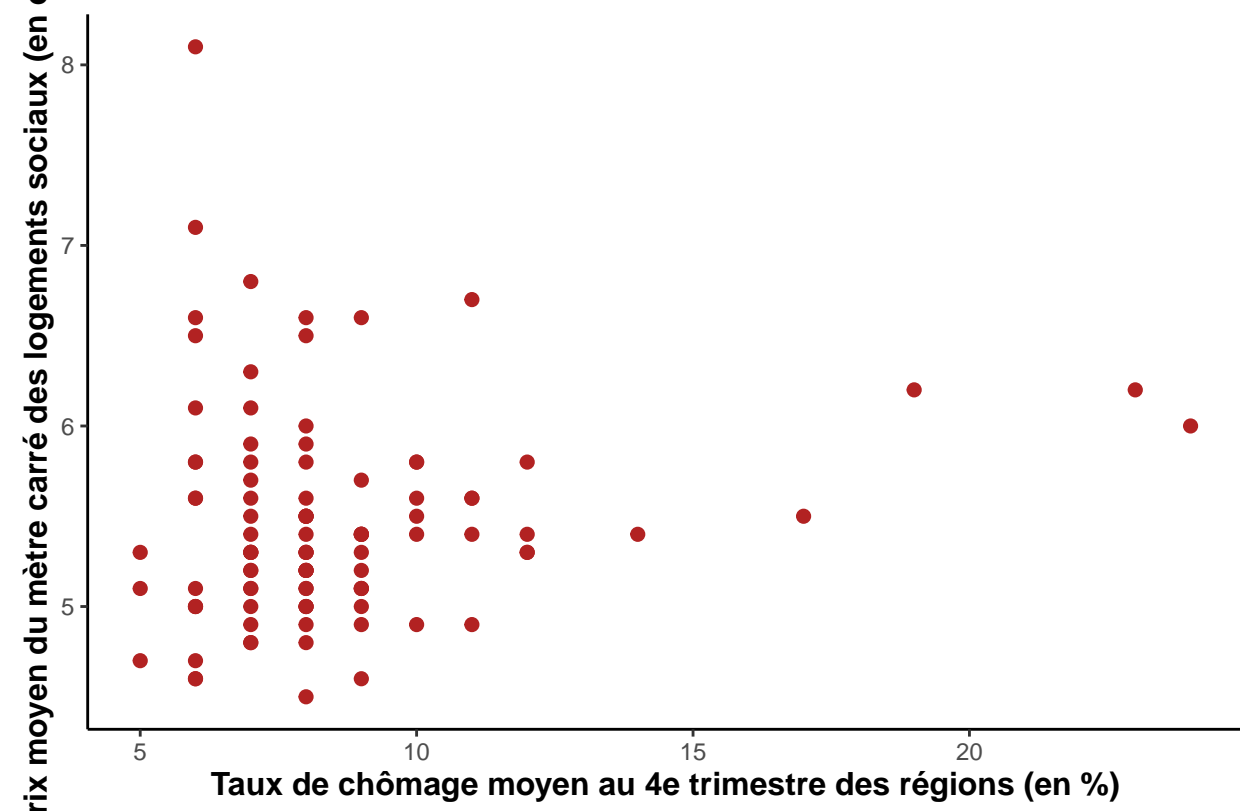
```
## rho
## -0.02
```

La p-value est de 0.8177, ce qui est supérieur au seuil de 5%. Par conséquent, nous concluons qu'il n'y a pas de corrélation statistiquement significative entre ces deux variables. On conclut donc que le prix du loyer d'un  $m^2$  ne dépend pas du salaire moyen de la région.

### C.3. Le chômage confronté au prix du $m^2$ des logements sociaux

Comme à chaque fois, commençons par un nuage de points.

Figure 1: Le prix moyen du mètre carré moyen des logements et le taux de chômage



```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.7, p-value = 0.00000000000008

##
## Spearman's rank correlation rho
##
## data: don$parc_social_loyer_moyen_en_eur_m2_mois and don$taux_de_chomage_au_t4_en
## S = 147700, p-value = 0.3
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.1
```

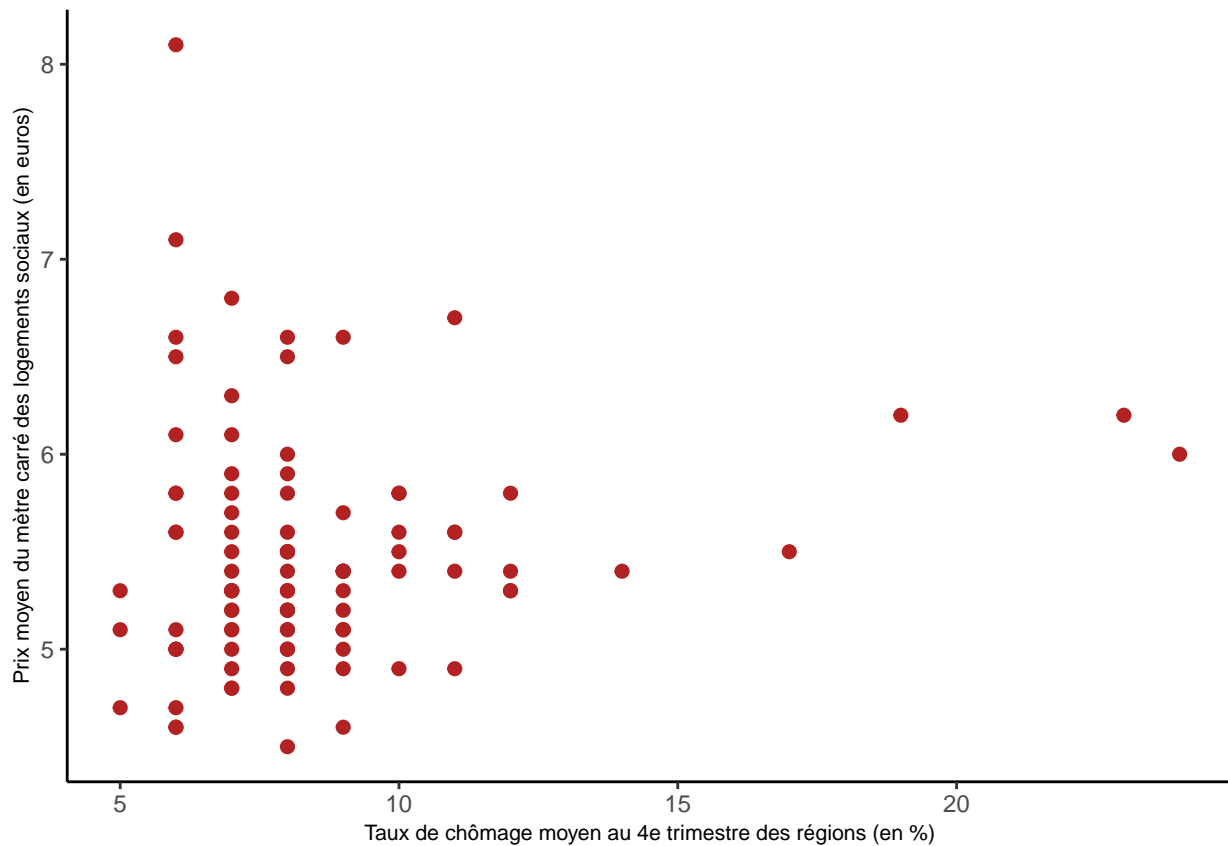


FIGURE 14 – Relation entre le prix moyen du mètre carré moyen des logements et le taux de chômage des régions

Le nuage de points ne nous permet pas de supposer une éventuelle corrélation linéaire entre ces deux variables. Nous avons donc testé la normalité de la paire de variables et nous avons dû rejeter cette hypothèse. Nous avons donc effectué un test de corrélation de Spearman.

```
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.7, p-value = 0.00000000000008

##
##  Spearman's rank correlation rho
##
## data:  don$parc_social_loyer_moyen_en_eur_m2_mois and don$taux_de_chomage_au_t4_en
## S = 147700, p-value = 0.3
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.1
```

La p-value du test de Spearman indique une corrélation non significative. On ne peut donc pas affirmer qu'il existe un lien significatif entre ses deux variables, alors aucune relation d'indépendance n'est à affirmer.

## Table des figures

1	Proportion de personnes âgées de 60 ans ou plus en 2019, en France, en fonction des départements de France. <i>Lecture</i> : Dans le département de l'Aube, il y a entre 22,3 et 27,5% de personnes de 60 ans et plus. . . . .	4
2	Proportion de personnes âgées de 20 ans ou moins en 2019, en France, en fonction des départements de France. <i>Lecture</i> : Dans le département de l'Aube, il y a entre 23,5 et 26,3% de personnes de moins de 20 ans. . . . .	4
3	Taux de logements sociaux en fonction des départements français en 2019. <i>Lecture</i> : Dans le département de l'Aisne, il y a entre 11,2 et 17,4% de logements sociaux. . . .	5
4	Proportion de personnes âgées de 20 ans ou moins en 2019, en France, en fonction de la proportion de personnes de 60 ans ou plus. <i>Lecture</i> : Pour 41% de moins de 20 ans dans un département, on retrouve 9% de plus de 60 ans. . . . .	5
5	Prix moyen du loyer au $m^2$ en fonction de l'âge de la population en 2019, en France. <i>Lecture</i> : Un département dans lequel environ 18% de la population est âgée de moins de 20 ans possède des logements sociaux avec un prix moyen fixé à 4,5 €/m <sup>2</sup> . . . . .	6
6	Distribution du nombre de logements sociaux pour chaque région, et représentation de la moyenne du nombre de logements sociaux pour chaque région. . . . .	15
7	Distribution du nombre de logements sociaux pour chaque région. . . . .	16
8	Proportion de logements sociaux en fonction de la densité de population par région. .	19
9	Relation entre le taux de logements sociaux énergivores et la proportion de la population ayant moins de 20 ans (à gauche) et la proportion de la population ayant 60 ans et plus (à droite). <i>Lecture</i> : Un département dans lequel 30% de la population est âgée de moins de 20 ans possède environ 20% des logements sociaux qui sont énergivores, en 2019, en France. . . . .	20
10	Relation entre le taux de logements sociaux énergivores (en %) et l'âge moyen du parc social (en année) <i>Lecture</i> : En moyenne, les parcs sociaux d'un département sont âgés de 49 ans dans lesquels environ 22% des logements sociaux sont des logements sociaux énergivores. . . . .	22
11	Distribution du taux de logements sociaux en fonction des régions . . . . .	40
12	Relation entre le prix moyen du mètre carré moyen des logements et les salaires nets des régions . . . . .	42
13	Relation entre le prix moyen du mètre carré moyen des logements sociaux et le taux de pauvreté des régions . . . . .	43
14	Relation entre le prix moyen du mètre carré moyen des logements et le taux de chômage des régions . . . . .	45

— Le premier ;

— Le deuxième.

## Webographie

- [Ren22] Marie-Cécile RENAULT. *À la retraite, les seniors fuient les grandes villes pour le Sud ou le littoral*. LEFIGARO. Section: Retraite. 10 nov. 2022. URL : <https://www.lefigaro.fr/retraite/a-la-retraite-les-seniors-fuient-les-grandes-villes-pour-le-sud-ou-le-littoral-20221110> (visité le 23/04/2023).