

Tp4

Sadou Ba

Introduction

Le Canada et le Québec vivent une crise du logement sans précédent. Ce phénomène s'est principalement vu révéler lors de la pandémie de covid-19. Depuis la médiatisation de la crise du logement plusieurs acteurs politiques se sont emparés du sujet en affirmant connaître la cause de celle-ci et ainsi, pouvoir attaquer le problème à la racine pour pouvoir l'endiguer. Dans les raisons évoquées, il y a la pression qu'exerce les immigrants sur le marché de l'habitation en augmentant la demande (Fillion, 2024). Il y a aussi l'hypothèse selon laquelle la pénurie est dû à une baisse des mises en chantiers. Ou encore, que cela est dû à l'augmentation des personnes vivant seules au sein de notre société (Nguyen, 2024). Si l'on s'attarde plutôt aux conséquences de la pénurie de logements, plusieurs indiquent que cela est la principale cause des loyers élevés. Ainsi, le but de ce travail sera d'observer s'il y a une corrélation entre le taux d'inoccupation des logements et le prix du loyer. Nous savons que dans la crise du logement, il y a aussi une crise de l'abordabilité de ceux-ci alors le but de notre recherche avec les données que nous utiliserons sera de mieux définir la raison de tels prix, ce qui pourrait aider les décideurs à mieux planifier leurs actions. Les éléments qui seront développés durant le reste de ce document seront la méthodologie, ainsi que l'analyse de nos résultats.

Méthodologie

Nos données proviennent du site web de l'observatoire québécois des inégalités, nous pouvons y trouver sous format texte les données souhaitées, soit les villes du Québec, le prix moyen du loyer par ville et le taux d'inoccupation. Ainsi, pour importer les données sur R pour y faire nos analyses, il a fallu installer une multitude de paquets dans R ainsi qu'une extension dans notre navigateur web. L'extension pour le navigateur web se nomme selector gadget et sert à sélectionner les éléments html de la page web. Ainsi une fois les éléments sélectionnés nous pouvons grâce à l'aide du paquetage rvest dans R dresser une liste de caractères pour chacune des variables. Nous avons ensuite pu rassembler nos listes en un tibble que nous avons ensuite transformé en data frame pour faire nos manipulations. Nous avons ensuite nettoyé nos données soit en changeant les virgules pour des points pour nos valeurs numériques ou encore d'enlever les espaces entre les valeurs numériques. Nous avons finalement enlevé les

valeurs manquantes. Pour montrer la corrélation entre le taux d'inoccupation et le prix moyen du loyer, nous avons effectué une régression linéaire.

Figure 1

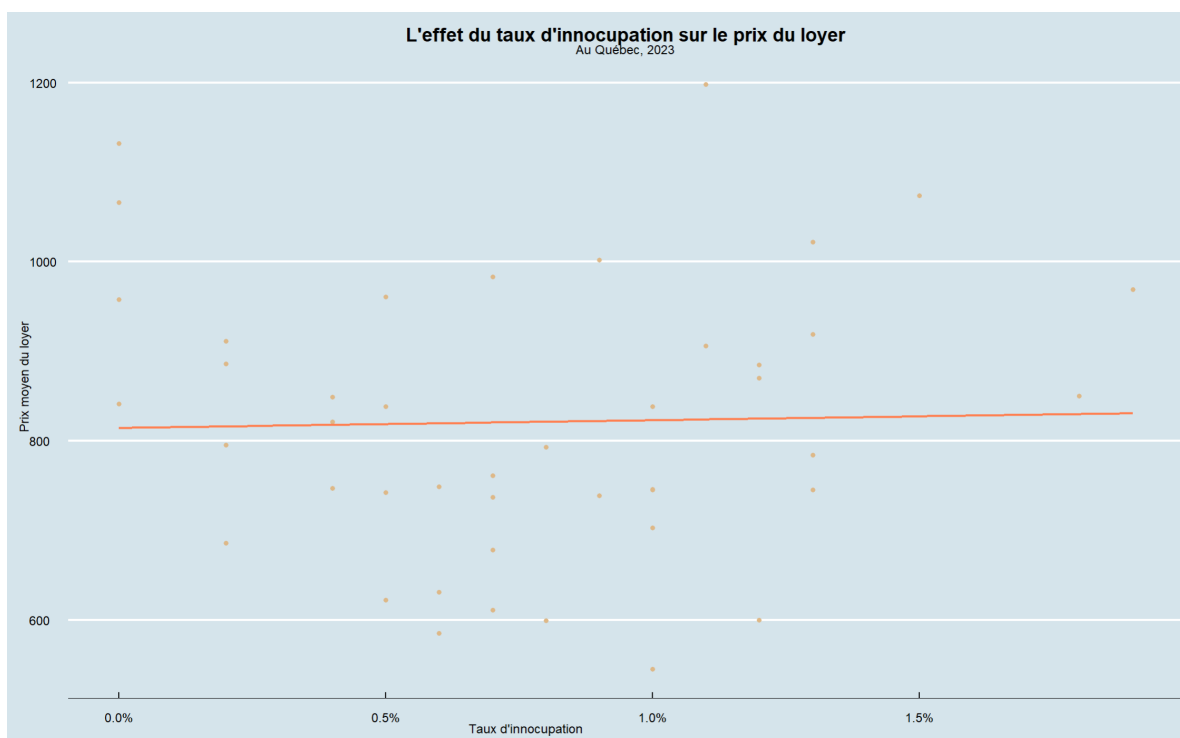
Description: df [6 x 3]

	Ville <chr>	Taux_innocation <dbl>	loyer_moyen <dbl>
1	Saguenay	1.3	745
2	Drummondville	0.5	838
3	Montréal	1.5	1074
4	Gatineau	1.1	1198
5	Québec	0.9	1002
6	Sherbrooke	1.3	919

6 rows

Résultats et discussion

Figure 2



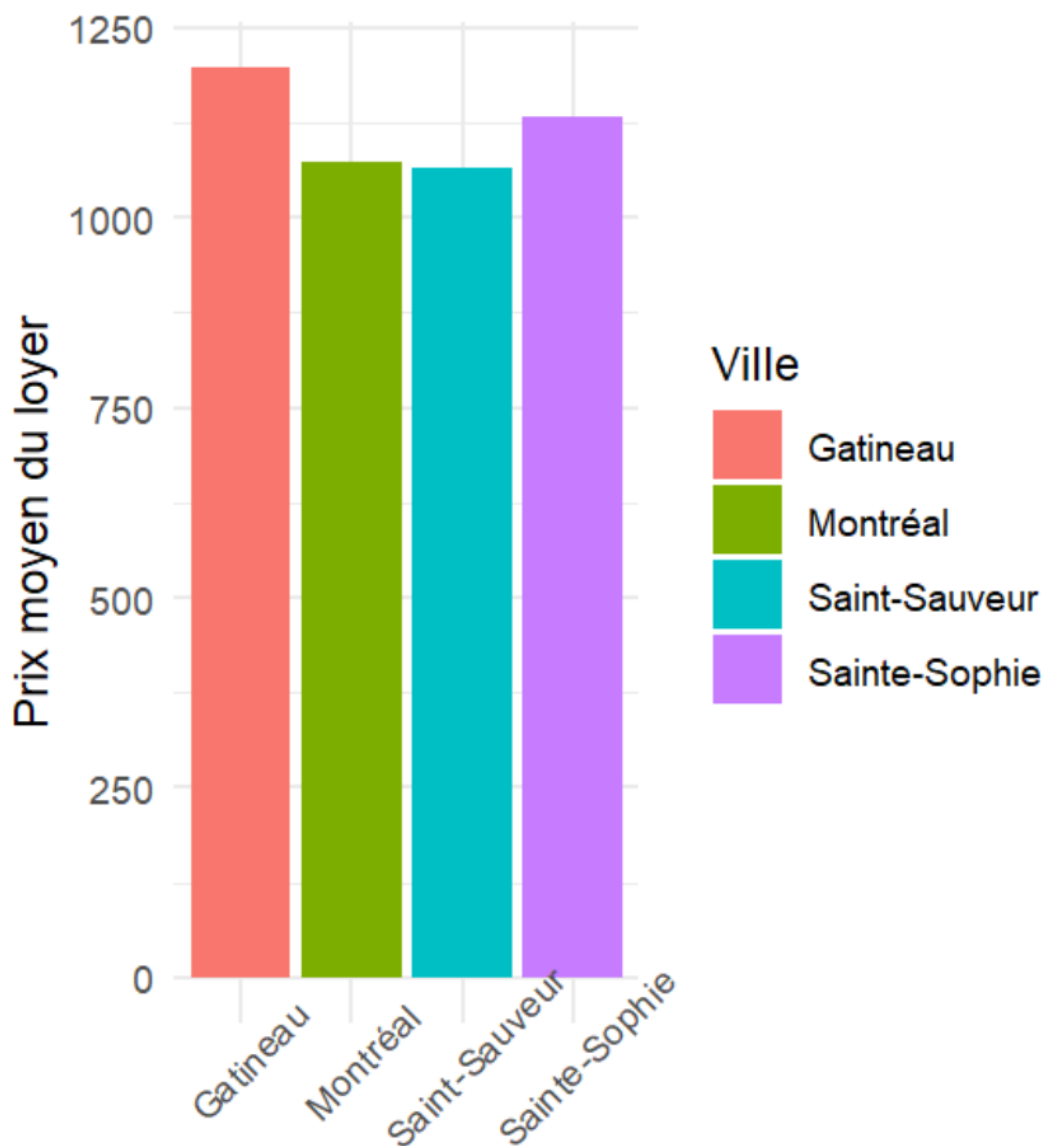
Comme nous pouvons le voir dans la figure 2 la corrélation entre le taux d'inoccupation est quasi nulle. Ce qui nous informe sur le fait que les loyers les plus chers ne sont pas nécessairement dû à la pénurie de logements, et il faudrait prendre en compte d'autres facteurs. Par exemple, le salaire moyen de la région. Où le type de logements disponibles, soit HLM

ou résidence étudiante. Toutefois, sur le même site de l'Observatoire québécois des inégalités, nous apprenons que le taux d'inoccupation influence grandement la variation du loyer. En effet, le marché local monte ainsi les prix lorsque l'offre diminue et nous pouvons voir que les villes avec les plus grandes hausses du prix moyen de loyer sont celles avec le plus bas taux d'inoccupation (Boucher, 2024).

Figure 3

Le prix des logements par ville

Au Québec, 2023



Dans la figure 3, il est possible de voir les 4 villes ayant un loyer moyen le plus élevé au Québec. Nous remarquons aussi que 2 des villes sont en zone urbaine, soit Gatineau et Montréal ce qui n'est pas surprenant mais aussi deux villes en zone rurales soit Saint-Sauveur et Sainte-Sophie,

ce qui est surprenant, mais qui aurait pu être la conséquence du tourisme.

Conclusion

Pour conclure, nous croyons que notre recherche est d'utilité publique et que nous avons eu l'éthique de chercheurs scientifiques. En effet, nous avons utilisés des données publiques et anonymiser, de plus le site web d'où nous avons extrait les données est la propriété d'un organisme de bienfaisance et nous utilisons ces données uniquement pour la diffusion scientifique et non en ayant pour but un intérêt financier. Ce qui selon nous, fais que notre travail est en concordance avec les normes éthiques.

Bibliographie

Geoffroy, Boucher, 31 janvier 2024, Rapport de la SCHL sur le marché locatif: la crise du logement risque d'exacerber les inégalités du Québec, <https://observatoiredesinegalites.com/marche-locatif-crise-logement/,G%C3%A9rald>

Fillion, 1 février 2024, Analyse Immigration et crise du logement : des nuances s'imposent, <https://ici.radio-canada.ca/nouvelle/2046130/immigration-crise-logement-demographie-croissance>,

Minh Nguyen, 26 octobre 2023, Crise du logement : portrait d'une dure réalité, Centrale des syndicats du Québec, <https://www.lacsq.org/actualite/crise-du-logement-portrait-dune-dure-realite/>,

Annexe

```
# Scrapping des données de la crise du logement
rm(list=ls())
library(lubridate)
```

Attachement du package : 'lubridate'

Les objets suivants sont masqués depuis 'package:base':

```
date, intersect, setdiff, union
```

```
# install.packages("rvest")
library(rvest)
library(tidyverse)
```

Warning: le package 'tidyverse' a été compilé avec la version R 4.3.2

Warning: le package 'ggplot2' a été compilé avec la version R 4.3.2

Warning: le package 'tidyr' a été compilé avec la version R 4.3.2

Warning: le package 'readr' a été compilé avec la version R 4.3.2

Warning: le package 'dplyr' a été compilé avec la version R 4.3.2

Warning: le package 'stringr' a été compilé avec la version R 4.3.2

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr   1.1.4      v readr   2.1.5
v forcats 1.0.0      v stringr 1.5.1
v ggplot2 3.4.4      v tibble  3.2.1
v purrr   1.0.2      v tidyr   1.3.1
```

```
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter()      masks stats::filter()
x readr::guess_encoding() masks rvest::guess_encoding()
x dplyr::lag()          masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
# install.packages("ggthemes")
library(ggthemes)
```

Warning: le package 'ggthemes' a été compilé avec la version R 4.3.3

```
# site web
pen_log <- read_html("https://observatoiredesinegalites.com/marche-locatif-crise-logement/
```

```

#pen_log_table <- pen_log %>% tibble(
# Ville = pen_log %>% html_element("tr+ tr b") %>% html_text(),
# loyer_moyen = pen_log %>% html_element("tr+ tr td:nth-child(3)") %>% html_text(),
# Taux_innocupation = pen_log %>% html_element("tr+ tr td:nth-child(2)") %>% html_text()
#)
#head(pen_log_table)

# colonne Villes
cities <- pen_log %>%
  html_elements("td:nth-child(1) b") %>%
  html_text()

# colonne loyer moyen
average_rents <- pen_log %>%
  html_elements("tr+ tr td:nth-child(3) b , td:nth-child(3) span") %>%
  html_text()

# colonne taux d'innocupation
occupancy_rates <- pen_log %>%
  html_elements("tr+ tr td:nth-child(2)") %>%
  html_text()

# Mise en forme d'un tableau
pen_log_table <- tibble(
  Ville = cities,
  Taux_innocupation = occupancy_rates,
  loyer_moyen = average_rents
)

print(pen_log_table)

```

A tibble: 48 x 3

	Ville	Taux_innocupation	loyer_moyen
	<chr>	<chr>	<chr>
1	Saguenay	1,3%	745 \$
2	Drummondville	0,5%	838 \$
3	Montréal	1,5%	1 074 \$
4	Gatineau	1,1%	1 198 \$
5	Québec	0,9%	1 002 \$
6	Sherbrooke	1,3%	919 \$
7	Trois-Rivières	0,4%	747 \$
8	Granby	0,5%	961 \$

```

 9 Joliette          1,8%          850 $
10 Rimouski         0,6%          749 $
# i 38 more rows

```

```

# conversion en dataframe + nettoyage de données
Qc_logements <- as.data.frame(pen_log_table)

Qc_logements <- Qc_logements %>%
  mutate(loyer_moyen = gsub(" ", "", loyer_moyen),
         Taux_innocation = gsub(",", ".", Taux_innocation)
        )

Qc_logements <- Qc_logements %>%
  mutate(
    loyer_moyen = parse_number(loyer_moyen),
    Taux_innocation = parse_number(Taux_innocation)
  )

```

```

Warning: There was 1 warning in `mutate()`.
i In argument: `Taux_innocation = parse_number(Taux_innocation)`.
Caused by warning:
! 4 parsing failures.
row col expected actual
 16 -- a number      -
 17 -- a number      -
 24 -- a number      -
 28 -- a number      -

```

```

Qc_logements <- na.omit(Qc_logements)

head(Qc_logements)

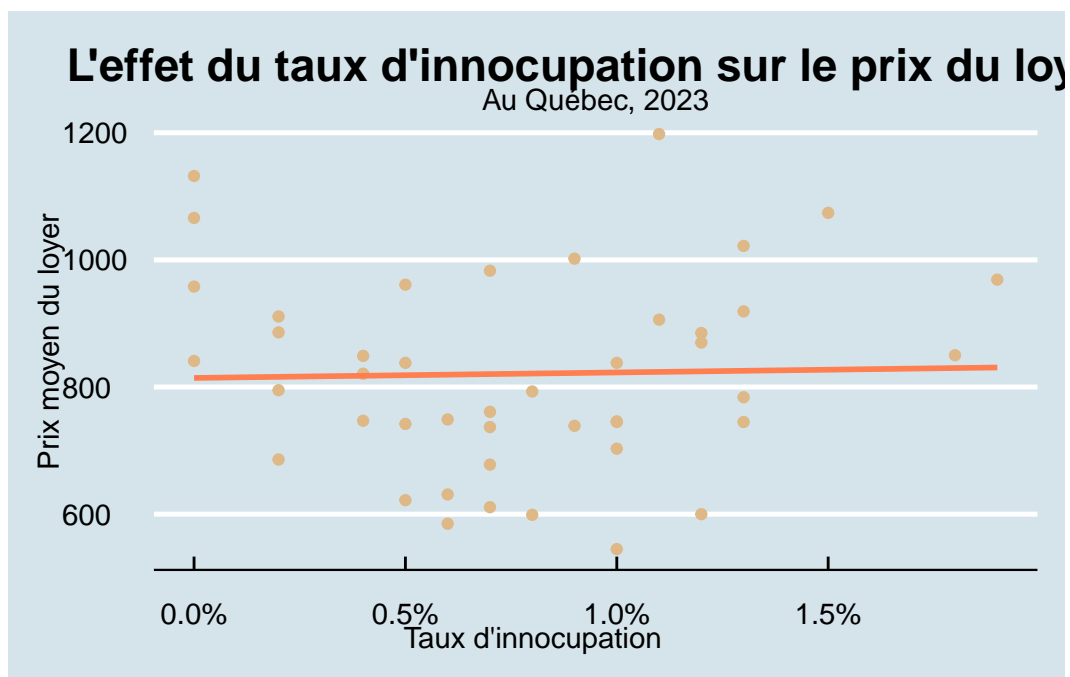
```

	Ville	Taux_innocation	loyer_moyen
1	Saguenay	1.3	745
2	Drummondville	0.5	838
3	Montréal	1.5	1074
4	Gatineau	1.1	1198
5	Québec	0.9	1002
6	Sherbrooke	1.3	919


```
# Régression linéaire
```

```
ggplot(data = Qc_logements, mapping = aes(x = Taux_innocation, y = loyer_moyen)) +  
  geom_point(color = "burlywood") +  
  geom_smooth(method = "lm", se = FALSE, color = "coral") +  
  labs(title = "L'effet du taux d'innocation sur le prix du loyer",  
        subtitle = "Au Québec, 2023",  
        x = "Taux d'innocation",  
        y = "Prix moyen du loyer") +  
  scale_x_continuous(labels = scales::percent_format(scale = 1)) +  
  theme_economist() +  
  theme(  
    plot.title = element_text(hjust = 0.5),  
    plot.subtitle = element_text(hjust = 0.5),  
    axis.text.x = element_text(hjust = 0.5),  
    axis.title.x = element_text(hjust = 0.4))
```

`geom_smooth()` using formula = 'y ~ x'



```
# Graphique top 4 villes les plus chères

logement_cher <- Qc_logements %>% top_n(4, loyer_moyen)

ggplot(data = logement_cher, mapping = aes(x = Ville, y = loyer_moyen, fill = Ville)) +
  geom_bar(stat = "identity") +
  labs(title = "Le prix des logements par ville",
        subtitle = "Au Québec, 2023",
        x = "Ville",
        y = "Prix moyen du loyer") +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 0.5),
    axis.title.x = element_text(hjust = 0.4)
  )
)
```

