

It's urgent to improve Social houses' overall quality in Toronto

A statistical description analysis based on the 'Apartment Building Evaluation' data

Yinuo Zhang

6 Feb 2022

Abstract

This study aims to investigate the Social houses' overall quality in Toronto based on the 'Apartment Building Evaluation' from Open Data Toronto Portal. The data includes about 10000 observations of buildings with about 40 attributes while a cleaned version used in this study. It was found that the number units of social houses is much smaller compared with private ones. It was also found although the average quality score for social houses is little higher now, but in several years, it could become lower and lower than that for private houses. The findings suggest the government should focus more on the social houses to improve quality for people to have a better living condition.

Keywords: social housing; Open Data Toronto Portal; Apartment Building Evaluation;

Introduction

Social houses refer the houses that prepared by government for people without a home. This could be an important indicator to measure the work of government in improving people's living condition and quality. In short, social housing gives people a home to satisfy the basic requirements for living. Social houses are always cheaper than private ones to rent and social houses are always providing long-term tenancy.

However, in recent years, due to lots of things, the house price becomes higher and higher over the world, lots of people can not afford such high pressure leading to a result that lots of people lost their home, this bring many social problems. Also, due to COVID-19, the issues become more serious, those people without their own houses faced with COVID-19 directly, even some of them are dead just in streets. Government should focus on the social housing assistance to make sure people has a basic living condition, the social housing is just a way to do the job. The government provides social houses to help those people to have a home which is an important part of housing benefit and social health development.

Due to the importance of the Social houses, this study focus on investigating the quality of social houses compared with private ones based on the "Apartment Building Evaluation" dataset provided by Toronto Open Data Portal which contains lots of free public available datasets. This study is important as the goal is to find out whether the government did enough for improving quality of social houses compared with those privates, if not, we could suggest the government focus on improving the quality of social houses by making better plans in future.

For this study, all of tables and figures were generated by R programming language (R Core Team 2020) [1], Rstudio and Rmarkdown [2][3], opendatatoronto package is used for importing original "Apartment Building Evaluation" dataset [4], psych package is used for data numerical description [5], tidyverse package is used for data cleaning [6], knitr package is used for generating tables [7], ggplot2 package is used for generating figures [8]¹.

¹Code and data are available at: <https://github.com/Amy527/socialhousing>

Data

Data source

The source of the data comes from the Open Data Toronto Portal using the opendatatoronto package. The Open Data Toronto portal has lots of public free available datasets stores in various formats such as CSV files. This study used the dataset named the “Apartment Building Evaluation” [9] which was collected through a bylaw enforcement program established in 2017. And the target of the program is to make sure that the owners and operators of the building houses could follow the Standards of Apartment Buildings. The dataset was last updated in the year 2021.

Methodology and Data Collection

The original “Apartment Building Evaluation” dataset includes about 10000 instances of buildings in Toronto with 40 attributes containing variables like property type of buildings, building IDs, years built, evaluated and registered of buildings, location addresses, evaluation items such as security score, entrance of doors and windows score, elevators score which are measured by a 5 point scale that the higher value the higher evaluation quality of corresponding item. Also the “Apartment Building Evaluation” dataset contains an overall item named “overall quality score” and confirmed units with the buildings.

For this study, not all of observations and variables were used. To investigate my goal of the study, I did some data cleaning processes. First, after I downloaded the original “Apartment Building Evaluation” data using opendatatoronto package, I selected 5 interested variables which are build year, property type, number of units, overall quality score and security score of buildings. Then I converted the build year, number of units, overall quality score and security score into numeric type and I converted property type into factor type. At last, I selected the renamed variables and removed missing values by rows.

There were only 40 observations dropped, the final cleaned data has 9718 observations with 5 variables, so the cleaned data is still a big one, however, this way of cleaning might introduce biasness in the analysis due to omitted observations and variables. Also, the scales of measures used in the original data set might be subjective that could introduce biasness too.

Data description

The cleaned data has 5 important variables: build year of buildings, property type of buildings, number of units in building, overall quality scores and security scores of buildings.

Table 1 illustrates the first several rows of the cleaned apartment building data, it can be found that the data variables are now in correct types after data cleaning, all of the following analysis are based on this cleaned data.

Table 1: First several rows of cleaned apartment building data

year	type	units	score	security
1964	PRIVATE	83	84	5
1964	PRIVATE	90	87	5
1968	PRIVATE	53	73	5
1962	TCHC	115	71	5
1969	PRIVATE	166	67	5
1960	PRIVATE	82	72	4

Table 2: Basic numerical description of apartment building data

	mean	sd	median	min	max
units	88.042	95.073	49	10	793
score	72.358	10.034	73	20	100
security	4.016	0.869	4	1	5

Table 2 illustrates the mean, median, standard deviation, minimum and maximum statistics for the cleaned data. The range of the built buildings started in the year 1805 and until the most recent the year 2021, the range is wide enough. For the number of units, the range is from 10 to 793, the average level is about 88 and the median level is about 49, the standard deviation is about 95, so the spread of number of units is very wide. For the overall quality score, the range is from 20 to 100, the average level is about 72 and the median level is about 73, the standard deviation is about 10, so the spread is much smaller and as mean and median values are close, the distribution is appear to be symmetric. For the security, the range is from 1 to 5, the average level is about 4 and the median level is about 4, the standard deviation is about 0.87, so the spread is small.

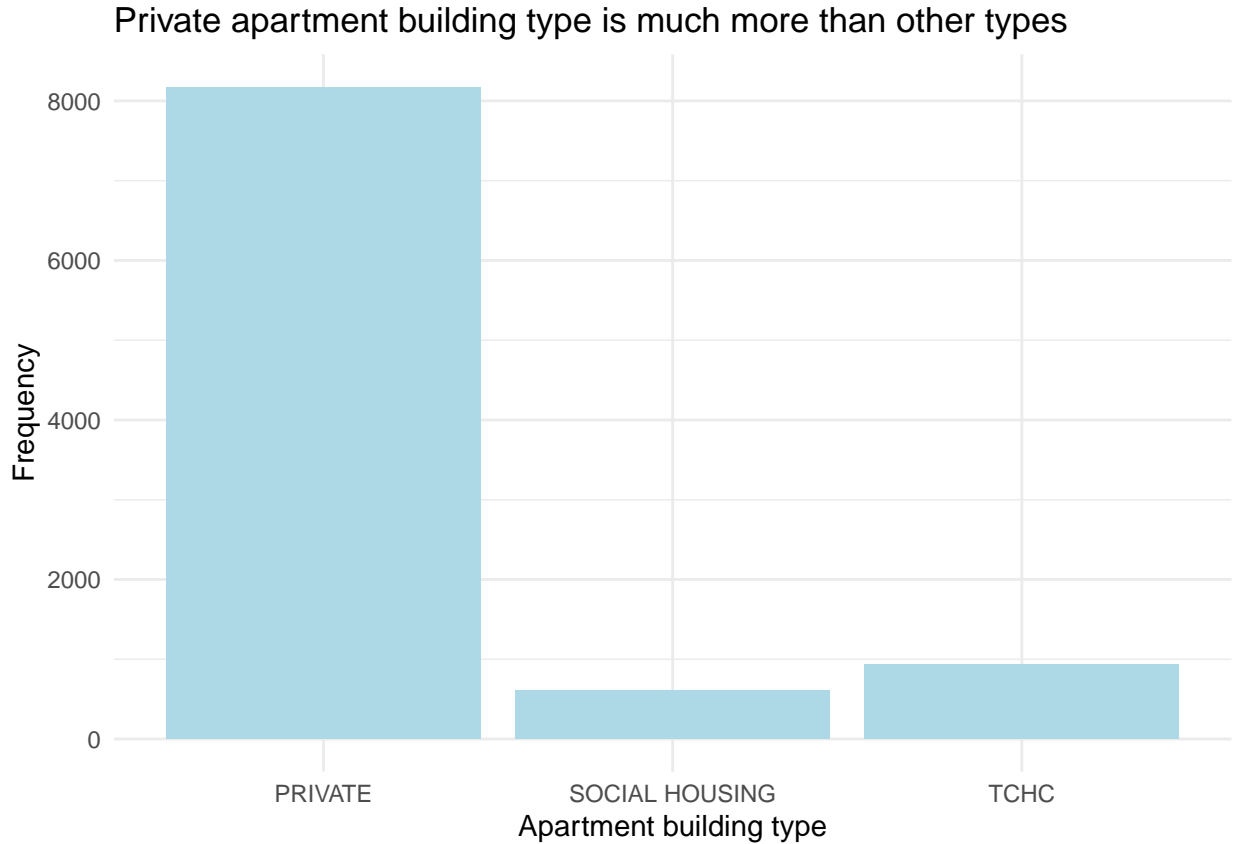


Figure 1: Number distributions of apartment buildings grouped by building type

Figure 1 shows the number distributions of apartment buildings grouped by building type. There are about 8200 private buildings, 600 social house building and 900 TCHC buildings. So that the private apartment building type is much more than other types, the social house buildings are much fewer. This finding indicates the government should focus their work more on the social house buildings, I think the government should better make some plans such as keep the ratio of the number of private houses versus the number of social houses in some ranges to make sure there are enough available social houses.

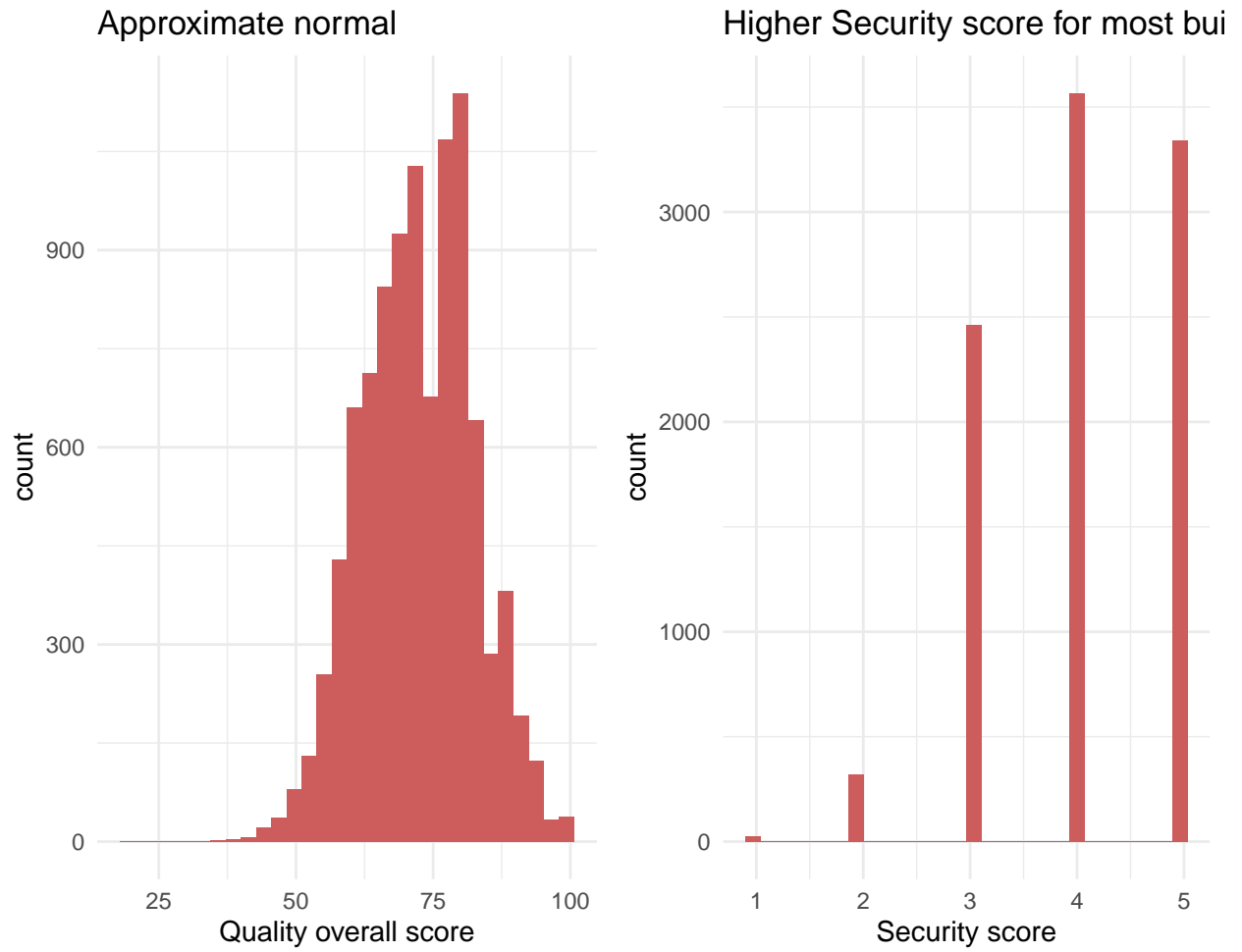


Figure 2: Distributions of Quality overall score and Security score of apartment buildings

Figure 2 shows the distributions of Quality overall score and Security score of apartment buildings. The distribution plots show that the Quality overall score is approximate normal distribution but the Security score tends to be high in average with a point scale around 4, this indicating the Security of houses is well enough now for most buildings but the overall quality of some houses still should be improved.

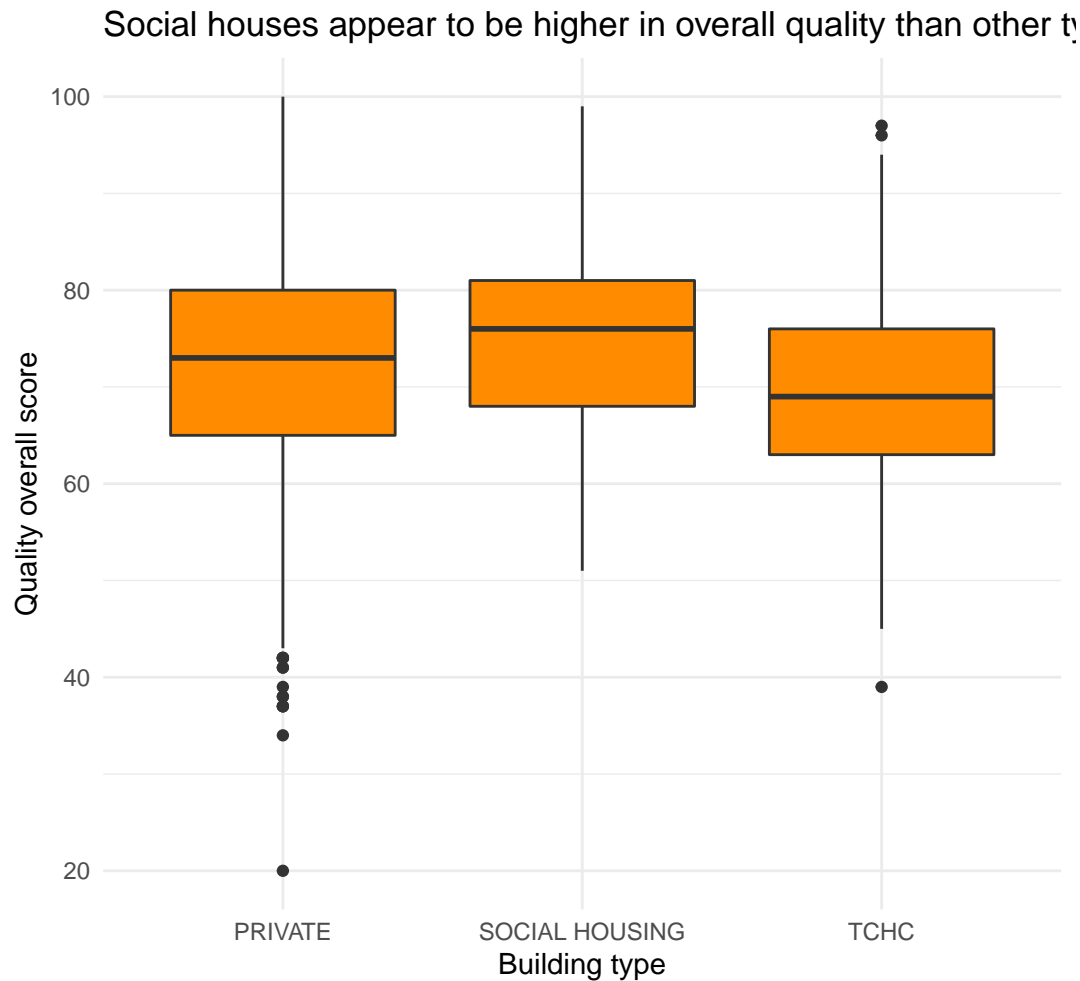


Figure 3: Distribution of Quality overall score by different types of apartment buildings

Figure 3 illustrates the distribution of Quality overall score by different types of apartment buildings. It seems that the social houses have a higher average level of quality overall score than other two types, the TCHC type is lowest while the private type is in the middle. This indicates the government did a not bad job that keep the social houses in a relative high quality.

Figure 4 illustrates the distribution of Quality overall score by different types of apartment buildings. Clearly, it can be found the average level of Quality overall score of buildings is increasing with the increasing of security score, it means that higher security score of buildings is strongly positive correlated with the overall quality score, it is not easy to understand as security is the first important thing of houses.

Figures 5 and 6 show the time trends of average overall quality score and average number of units grouped by building types. Clearly, figure 5 shows the Social housing units become much fewer and fewer compared with private houses, this means the government appear not focus on keep the number of units of social houses with the private ones in some ratio, but I think it is important due to we could prepare enough social houses to make sure people have basic conditions for living. Figure 6 shows the social housing overall quality score is not improving clearly, this could be a big issue, because although it was found that the social housing

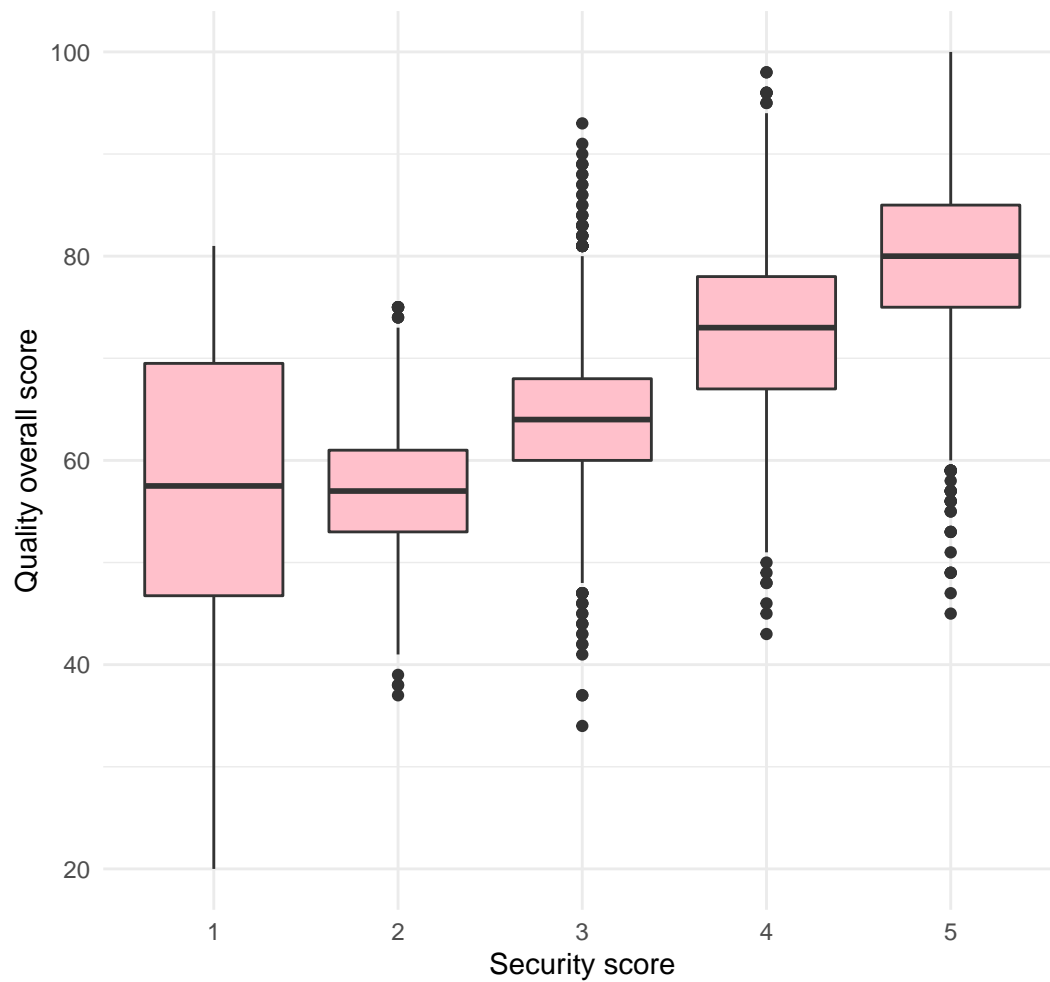


Figure 4: Distribution of Quality overall score by different types of apartment buildings

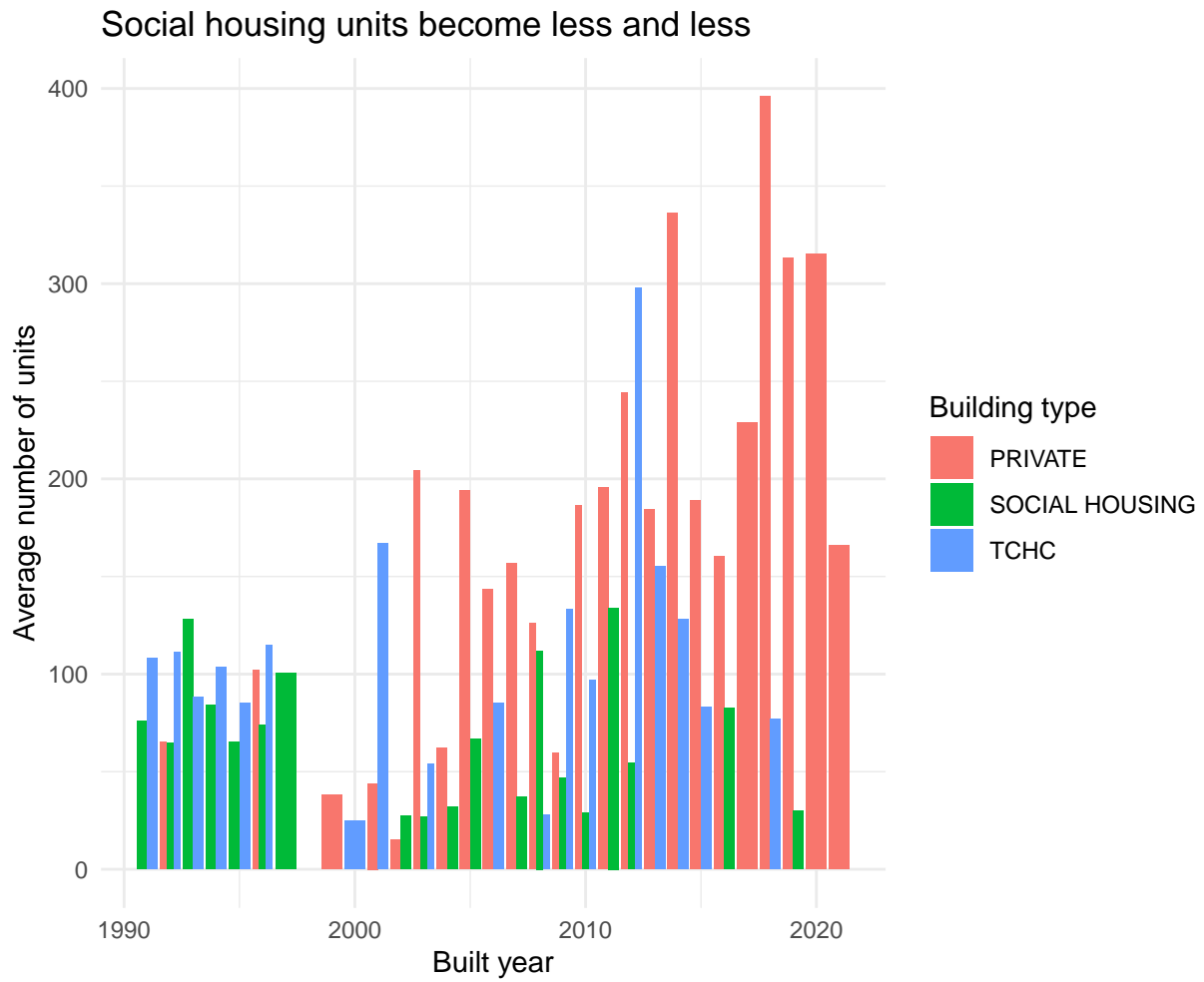


Figure 5: Average number of units grouped by building types trend in years

Social housing overall quality score is not improving clearly

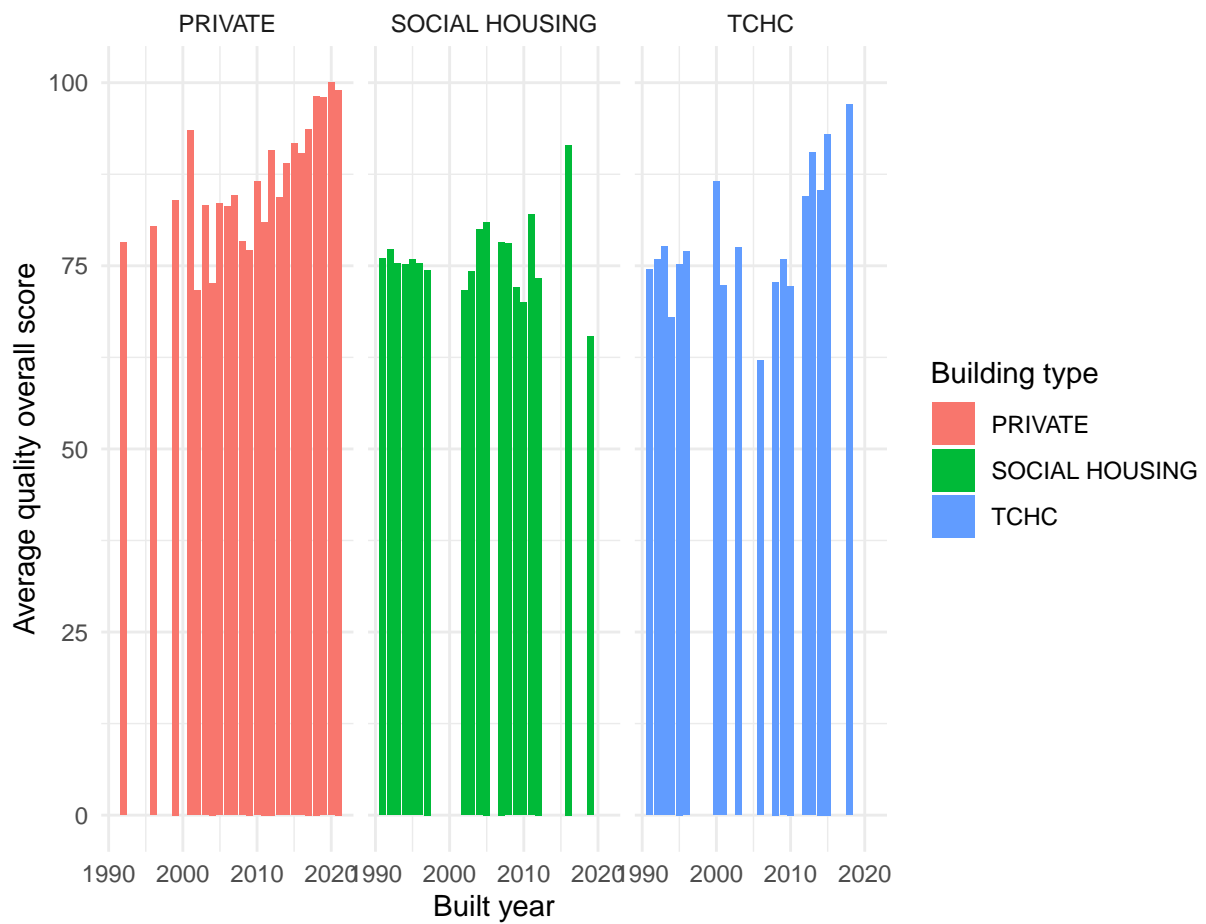


Figure 6: Average overall quality score grouped by building types trend in years

overall quality score for social houses is little higher than private ones, however, the time trend plot shows the overall quality score for private ones become higher and higher with increasing of time but overall quality score for social houses appear to be stay in same level across years, this means after several years, the overall quality score for social houses might be much lower than that for private ones, and if government is still not focus on this issue, the gap between the social houses and private ones in building quality might be larger and larger.

Discussion

Beside the findings, there are some weakness and limitations. First, the cleaned data used in this study might introduce biasness due to some observations are dropped and only interested variables are selected, there could be omitted variable biasness. Second, the data collection for the original data might introduce non-response biasness as not all of the buildings investigated could be recorded correctly. Third, the measures in quality and security scores could be biasness as these scores were subjective. At last, the results are only for Toronto, other locations might have different findings. In future work, the study could focus on more interested variables and try to apply more advanced methods such as using statistical models to investigate the factors which affect the quality of social houses.

Conclusion

Overall, this study finds that the government should focus more on the social house buildings in several aspects. First, the number units of social houses compared with private ones is much smaller, the number of social houses appear to be hard in making sure most of people could have a home. We suggest the government should make better plans in preparing enough social houses, for example, the government could keep a high enough ratio of social houses compared with private ones. Second, although in the current years, it was found that the average quality score of social houses is little higher than private ones, in a long-term trend or even in about 10 years, it was found by trend analysis that the average quality score of private ones could become higher and higher than that for social ones. The gap between private ones and social ones could be much larger in future if the government not focus on social houses. After several years, the overall quality score for social houses might be much lower than that for private ones, so we suggest the government make plans as quick as possible to deal with the issues of social housing.

Refereneces

1. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
2. Grolemund, G. (2014, July 16) *Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/articles_intro.html.
3. JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2021). rmarkdown: Dynamic Documents for R. R package version 2.11. URL <https://rmarkdown.rstudio.com>.
4. Sharla Gelfand (2020). opendatatoronto: Access the City of Toronto Open Data Portal. R package version 0.1.4. <https://CRAN.R-project.org/package=opendatatoronto>.
5. Revelle, W. (2020) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 2.1.3,.
6. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi: 10.21105/joss.01686.
7. Yihui Xie (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.36.
8. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
9. “Apartment Building Evaluation”. City of Toronto. Retrieved 2022-02-06.