# TLe_FinalProject

Tien Le

2024-12-09

# Table of Contents

## Abstract

This study examines the impact of proximity to expressways, Mass Rapid Transit (MRT) stations, and storey levels on housing resale prices in Singapore. Using data from the Housing Development Board, National Map Line (SLA), and OneMap Reverse Geocode, the dataset includes 220,971 records of public housing resale transactions from 2015 to 2024.

The analysis reveals significant relationships between housing prices and storey levels. Flats located further from expressways tend to result in higher prices due to reduced noise, traffic pollution and different socio-economic neighborhood. Additionally, higher storey flats, particularly those categorized as "Very High," are valued more highly, likely due to enhanced privacy, social status, and views. The interplay between expressway proximity and storey levels emphasizes the multifaceted influences on property values where lands are scarce, in a densed country with high living costs, and a country with the 2nd highest GDP per capita (worldometers 2022).

This study's findings demonstrates the importance of infrastructure and housing design in shaping housing market dynamics in densely populated cities, and how unique regulations and culture may not be able to translate to other countries. Future research could integrate variables such as median salary and proximity to MRT stations to better capture the complexities of urban housing markets.

## Introduction

The United States relies heavily on car ownership and private transportation, with 91.7% of households owning at least one vehicle in 2022 (Caporal 2024). However, in dense metropolitan areas such as Chicago, Boston, and San Francisco, among the country's most populated and economically dynamic cities, public transportation plays an essential role. High living costs, limited real estate availability, and well-developed transit systems make proximity to public transit a critical factor in determining property values (Becker_etal. 2013).

This study's motivation aims from my study abroad experience in Singapore and question about whether the above's phenomenon applies to other dense cities and countries as well. Singapore is a country where the Mass Rapid Transit (MRT) system serves as the primary mode of transportation. I relied on the MRT daily to travel across the country, gaining firsthand experience of how Singapore's infrastructure revolves around its transit system. With limited land and a dense population, Singapore's urban planning mirrors the challenges of dense U.S. cities. Land, the country's most valuable natural resource, is highly sensitive to changes based on infrastructure development. Additionally, proximity to MRT stations introduces trade-offs, such as increased noise, traffic, and reduced privacy, despite the convenience it offers. To account for these factors, this study also examines the impact of building height, or "storey" level, on housing value.

This research aims to analyze how proximity to expressways and MRT stations, and storey levels affects housing prices in Singapore.

## Data

My data (Goh 2024) comes from the Singapore Housing Development Board, National Map Line (SLA), and OneMap Reverse Geocode. This dataset represents 220,971 records of resale prices for Singapore public housing (HDB) flats or apartments from 2015 to 2024. In this project, I examine the relationships between housing resale prices (in SGD), expressway proximity (meters), and storey range level (floor number).

This data is significant because it provides key information about Singapore's housing market, making it relevant for examining the effects of public amenities such as MRT systems and expressways on housing prices, as well as residents' attitudes toward Singapore's scarce and valuable asset: real estate (Department 2024).

A limitation of this data is that it does not account for the socio-economic status of real estate owners in Singapore. Wealthier families may afford private modes of transportation, allowing them to live further from working-class areas and MRT stations. Additionally, the resale prices were recorded from 2015 to 2024, a period during which housing prices may have changed dramatically, making comparisons challenging. However, while U.S. housing prices rose by 4.3% in Q2 2024 (AGENCY 2024), Singapore's housing market saw a modest 1.39% increase (Delmendo 2024). Therefore, we do not separate the resale flat prices by year, given the relatively lower price volatility in Singapore. Instead, the focus is on the effects of amenities like the MRT on flat resale prices.

In the future, I would separate the years into four-year intervals and include an additional variable to categorize salary ranges, enabling a better understanding of household lifestyles and their reliance on public transportation.

Throughout this analysis, I will maintain consistency with the original dataset's terminology: "flat" refers to condominiums or apartments in Singapore, "storey range" refers to the floor level, and "proximity" indicates the distance to an MRT or expressway.

```r
library(dplyr)
library(ggplot2)
library(bookdown)
```

```r
SG.flats <- read.csv("/Users/tienle/Downloads/hdbpresaleprices.csv")
SG.flats <- na.omit(SG.flats)

## Remove unused columns and information
SG.flats <- SG.flats |> select(year, resale_price, flat_type, storey_range_category, distance_from_expre


summary(SG.flats)
```

```
##       year       resale_price      flat_type         storey_range_category
##  Min.   :2015   Min.   : 140000   Length:220971      Length:220971
##  1st Qu.:2017   1st Qu.: 365000   Class :character   Class :character
##  Median :2020   Median : 458000   Mode  :character   Mode  :character
##  Mean   :2020   Mean   : 490867
##  3rd Qu.:2022   3rd Qu.: 585000
##  Max.   :2024   Max.   :1588000
##  distance_from_expressway
##  Length:220971
##  Class :character
##  Mode  :character
##
##
##
```

## Variable definitions

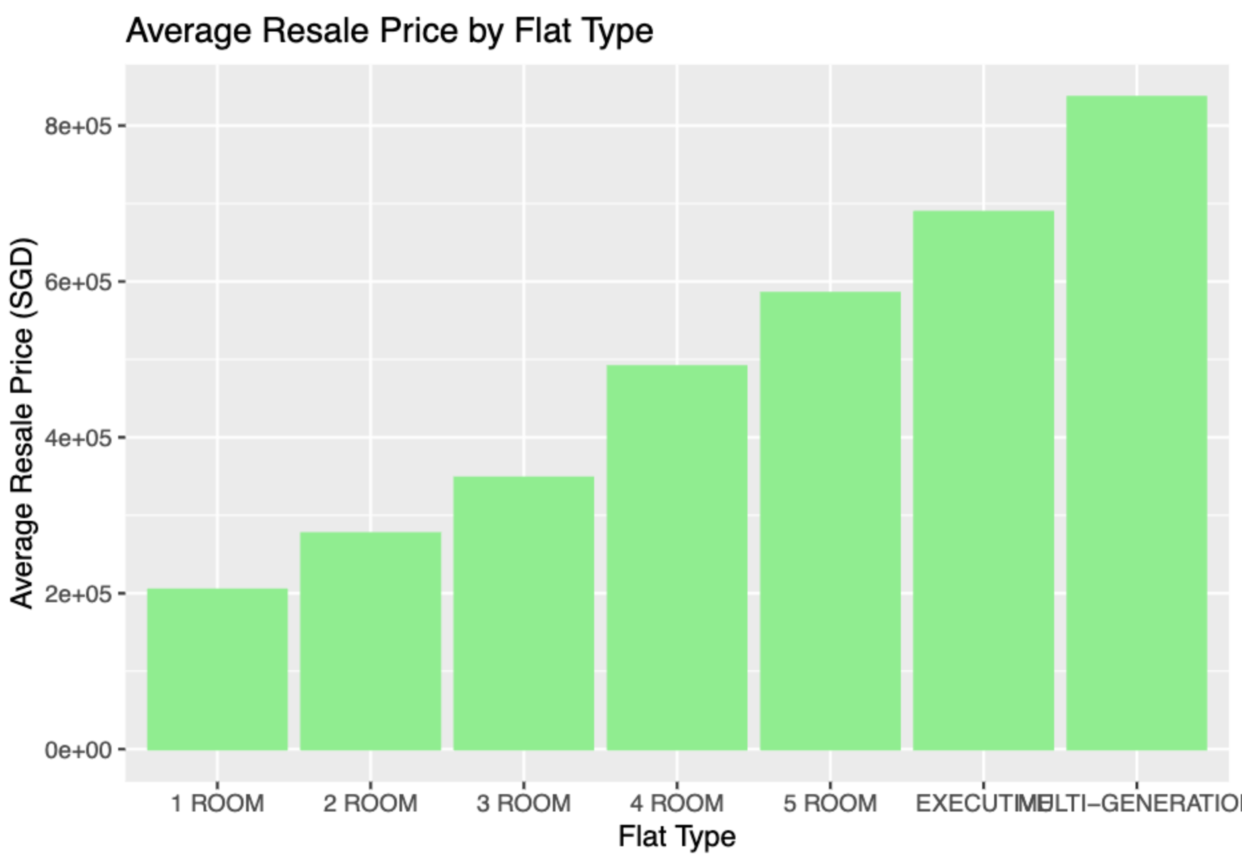
- **resale_price**: transaction price (in Singapore dollars)

- **flat_type**: type of flat (e.g. 1-room, 2-room, 5-room, executive)
- **storey_range_category**: categorical variable created based on "storey_range" (Low (01-06); Low-Mid (07-12); Mid (13-18); High (19-30); Very High (>30))

- **distance_from_expressway**: distance from the building to the nearest expressway (<=50m, 51-100m, 101-150m, 151-300m, 301-500m, >500m)

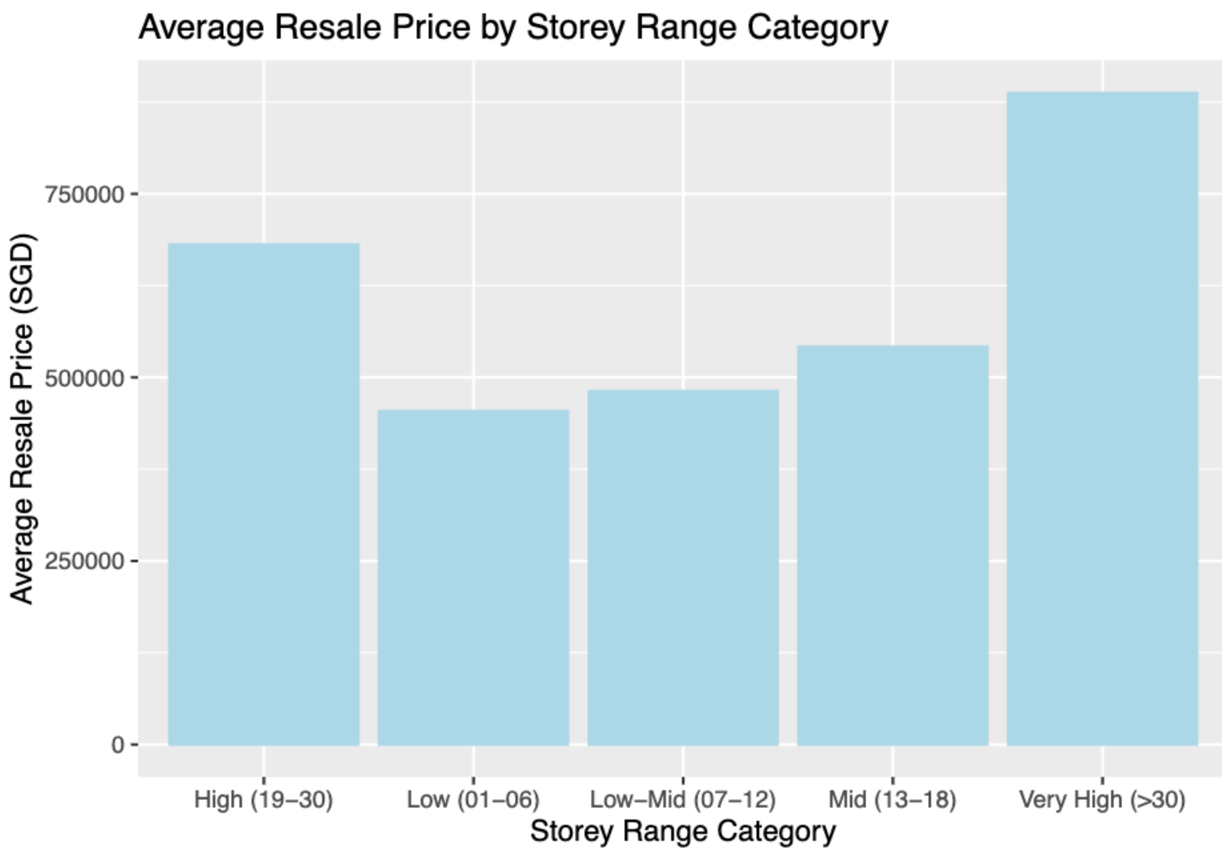
### Signifance of these variables

```r
average_prices <- SG.flats %>%
  group_by(flat_type) %>%
  summarise(avg_resale_price = mean(resale_price, na.rm = TRUE))

ggplot(average_prices, aes(x = flat_type, y = avg_resale_price)) +
  geom_bar(stat = "identity", fill = "lightgreen", color = "lightgreen") +
  labs(x = "Flat Type", y = "Average Resale Price (SGD)", title = "Average Resale Price by Flat Type") +
  theme(legend.position = "none")
```




```r
# Calculate the average resale price by storey range category
storey_range_avg <- SG.flats %>%
  group_by(storey_range_category) %>%
  summarise(avg_resale_price = mean(resale_price, na.rm = TRUE))

# Plot average resale price by storey range category
ggplot(storey_range_avg, aes(x = storey_range_category, y = avg_resale_price)) +
  geom_bar(stat = "identity", fill = "lightblue", color = "lightblue") +
  labs(x = "Storey Range Category", y = "Average Resale Price (SGD)",
```

```r
       title = "Average Resale Price by Storey Range Category") +
  theme(legend.position = "none")
```




```r
# Calculate the average resale price by distance from expressway
expressway_avg <- SG.flats %>%
  group_by(distance_from_expressway) %>%
  summarise(avg_resale_price = mean(resale_price, na.rm = TRUE))

# Plot average resale price by distance from expressway
ggplot(expressway_avg, aes(x = distance_from_expressway, y = avg_resale_price)) +
  geom_bar(stat = "identity", fill = "lightpink", color = "lightpink") +
  labs(x = "Distance from Expressway", y = "Average Resale Price (SGD)",
       title = "Average Resale Price by Distance from Expressway") +
  theme(legend.position = "none")
```


```
flattypeMedian <- median(SG.flats$flat_type)
storey_rangeMedian <- median(SG.flats$storey_range_category)
distance_fromexpressMedian <- median(SG.flats$distance_from_expressway)
```

Based on the barplots @ref(fig:Price Surface Level Relationship), the most common flats, or the median flats, are the 4 ROOM flat type, flats that are Low-Mid (07-12) floors and >500m away from the expressway.

Based on the observations, we can make a few naive predictions:
- As the size of flats increase, flat_type as reference, the resale price does not follow an ascending trend, but smaller and private flats are valued on a higher price.
- Flats on the middle levels, Low-Mid (07-12) and High (19-30), are more valuable than those on the lowest and highest. Possibly due to convenience from accessing elevators, stairs and amenities on the mid-level, and avoiding noise and providing a sense of superiority by avoiding the low levels.
- Flats that are far away from the expressway are valued higher than others due to noise, pollution and the owner's ability to commute without using public transportation.

Note: Some outliers are the executive and multi-generation flats, due to its limited availability and extension governmenet application to acquire those spaces. For example, executive and multi-generation flats are only available to Singapore citizens and permanent residents family, despite having almost 40% of occupants being foreigners.

# Analysis & Result

## Linear Model on Individual Variables

```
flats.fit <- lm(resale_price ~ distance_from_expressway + storey_range_category, data = SG.flats)
model_sum <- summary(flats.fit)
model_sum
```

```
##
## Call:
## lm(formula = resale_price ~ distance_from_expressway + storey_range_category,
##     data = SG.flats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -565094 -115125  -28770   90908 1156670
##
## Coefficients:
##                                      Estimate Std. Error  t value Pr(>|t|)
## (Intercept)                            638337       3085  206.940  < 2e-16 ***
## distance_from_expressway>500m           45795       2666   17.176  < 2e-16 ***
## distance_from_expressway101-150m        34440       3309   10.409  < 2e-16 ***
## distance_from_expressway151-300m        31663       3189    9.927  < 2e-16 ***
## distance_from_expressway301-500m        27777       3650    7.610 2.75e-14 ***
## distance_from_expressway51-100m         19854       3143    6.316 2.69e-10 ***
## storey_range_categoryLow (01-06)      -227007       1679 -135.205  < 2e-16 ***
## storey_range_categoryLow-Mid (07-12)  -200040       1682 -118.963  < 2e-16 ***
## storey_range_categoryMid (13-18)      -139980       1835  -76.290  < 2e-16 ***
## storey_range_categoryVery High (>30)   205962       3920   52.536  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 158100 on 220961 degrees of freedom
## Multiple R-squared:  0.1382, Adjusted R-squared:  0.1382
## F-statistic:  3938 on 9 and 220961 DF,  p-value: < 2.2e-16
```

```
p_value <- model_sum$coefficients["(Intercept)", "Pr(>|t|)"]
```

According to the linear model fit, using flat prices with High range, and with all p-values– $0 < \alpha = 0.01$–it suggests that these categories of expressway proximity and storey range can confidently estimate flat resale prices, the estimated figures are significant.

## Estimated Effects of Expressway Proximity

```
model_coef <- summary(flats.fit)$coefficients
distancecoef <- model_coef[grep("distance_from_expressway", rownames(model_coef)), ]

#make a new dataframe to plot
```

```r
estimatedchangeDF <- data.frame(
  category = rownames(distancecoef),
  estimate = distancecoef[, "Estimate"],
  std_error = distancecoef[, "Std. Error"]
)
estimatedchangeDF$category <- gsub("distance_from_expressway", "Dist", estimatedchangeDF$category)

# Flip for readability for each category and use reorder to sort ascending
ggplot(estimatedchangeDF, aes(x = reorder(category, estimate), y = estimate)) +
  geom_bar(stat = "identity", fill = "purple", color = "black") +
  coord_flip() +
  labs(
    title = "Estimated Effects on Resale Price Based on Expressway Proximity",
    y = "Estimated Change in Price (SGD)",
    x = "Proximity to Expressway"
  )
```

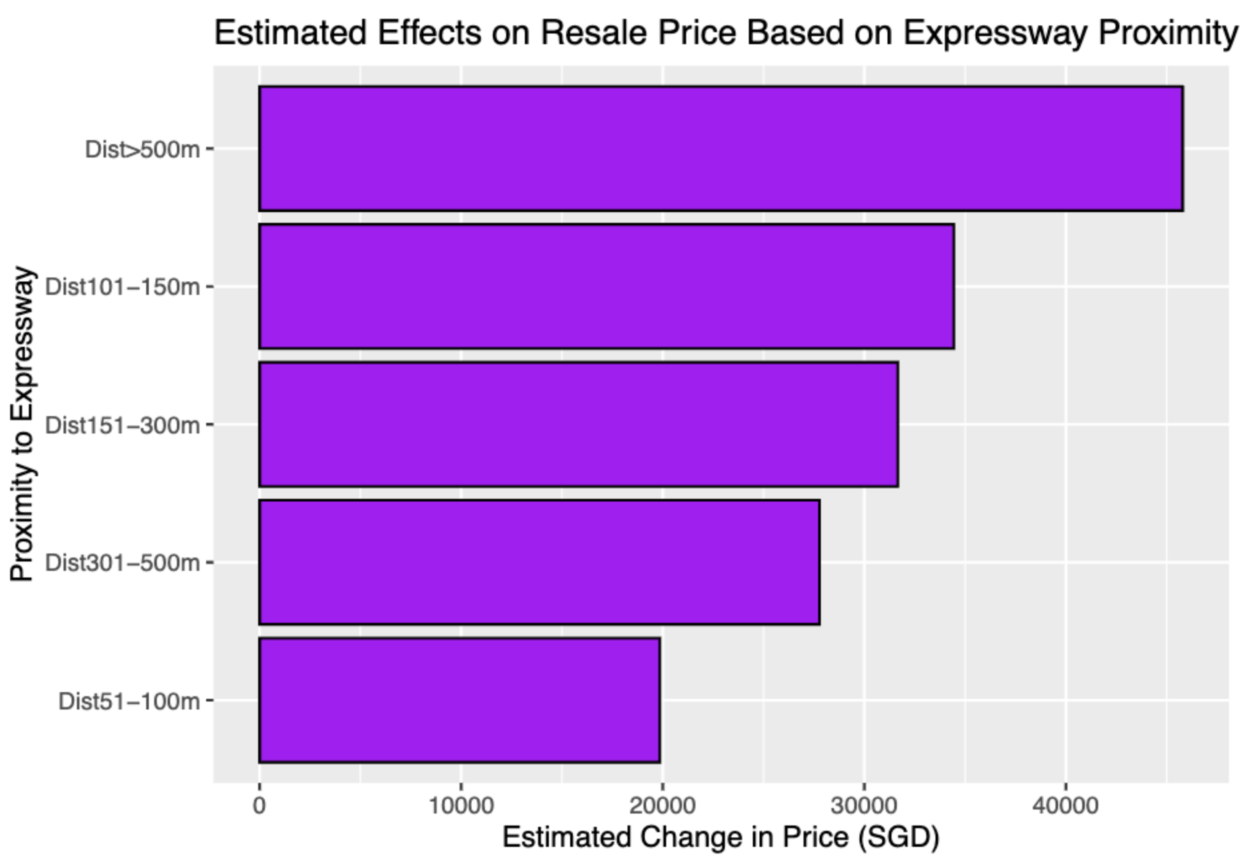
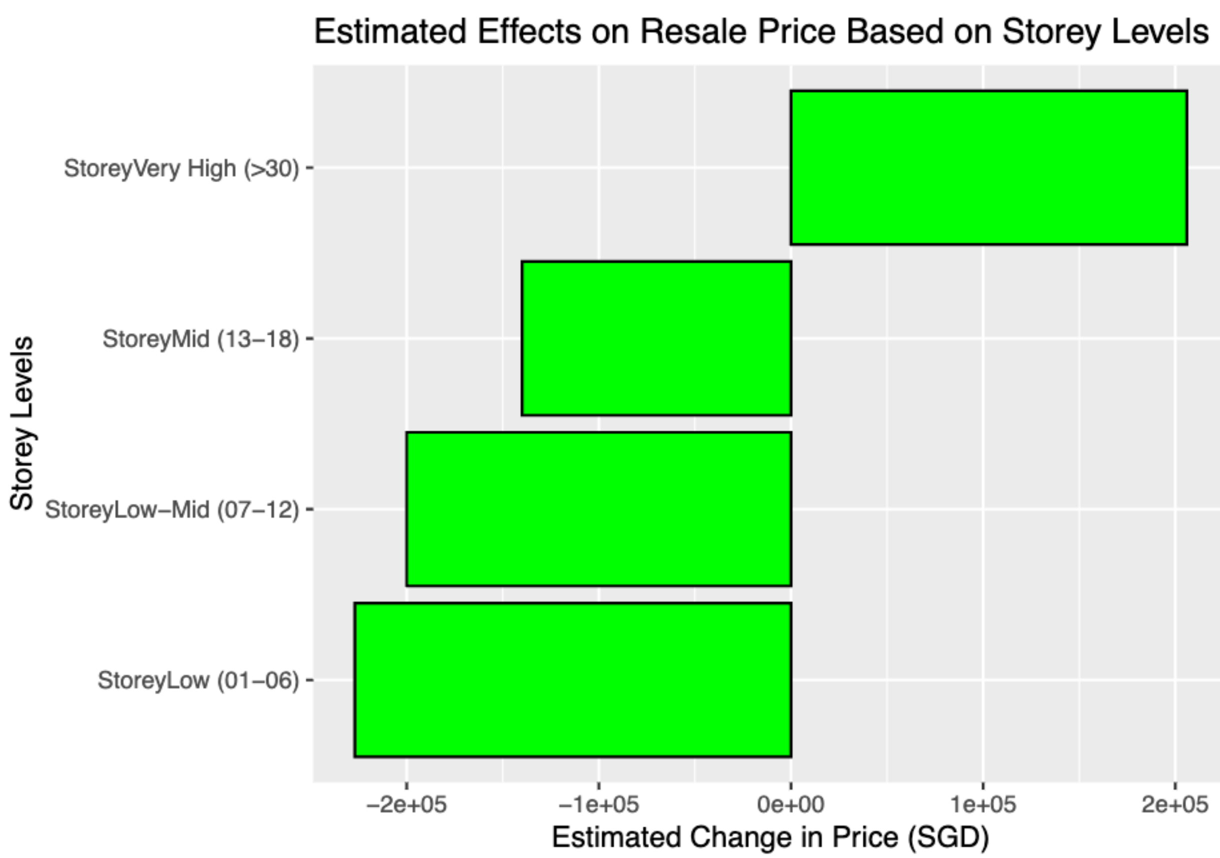



Figure 1: Estimated Resale Price Affected by Expressway Proximity

As shown in Figure @ref(fig:estchangeBarplot), flats with distances from expressway that are greater than the reference point, <50m, will have a positive estimated change in price and higher valued as the flats are higher level. These estimates are statistically significant as mentioned before due to the p-value being $< \alpha = 0.01$.

### Storey Level Estimated Effects on Resale Prices

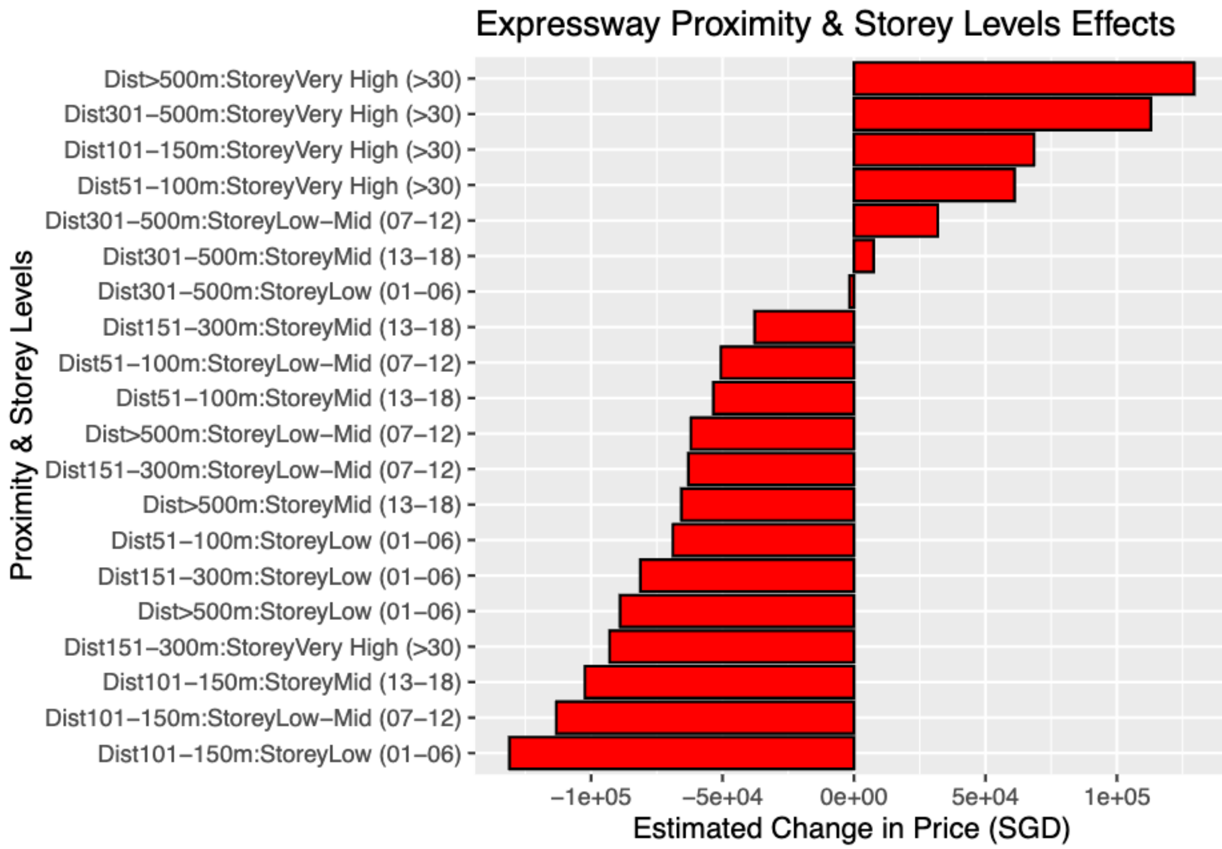
```r
distancecoef <- model_coef[grep("storey_range_category", rownames(model_coef)), ]

#make a new dataframe to plot
estimatedchangeDF <- data.frame(
  category = rownames(distancecoef),
  estimate = distancecoef[, "Estimate"],
  std_error = distancecoef[, "Std. Error"]
)
estimatedchangeDF$category <- gsub("storey_range_category", "Storey", estimatedchangeDF$category)

# Flip for readability for each category
ggplot(estimatedchangeDF, aes(x = reorder(category, estimate), y = estimate)) +
  geom_bar(stat = "identity", fill = "green", color = "black") +
  coord_flip() +
  labs(
    title = "Estimated Effects on Resale Price Based on Storey Levels",
    y = "Estimated Change in Price (SGD)",
    x = "Storey Levels"
  )
```




On figure @ref(fig:storeylevelBarplot), flats that are lower than the reference point, High category (19-30), will have less value than flats that are on the Very High category (>30). These estimates are statistically significant as mentioned before due to the p-value being $< \alpha = 0.01$.

### Relationship Between Storey Levels and Proximity

```r
storeyproximity <- lm(resale_price ~ distance_from_expressway * storey_range_category, data = SG.flats)
model_summary <- summary(storeyproximity)


storeyproximity_coef <- summary(storeyproximity)$coefficients
coefcollect <- storeyproximity_coef[grep(":", rownames(storeyproximity_coef)), ]

# Create a new dataframe to plot
estrelation <- data.frame(
  category = rownames(coefcollect),
  estimate = coefcollect[, "Estimate"],
  std_error = coefcollect[, "Std. Error"]
)

# Shorten category name using gsub()
estrelation$category <- gsub("distance_from_expressway", "Dist", estrelation$category)
estrelation$category <- gsub("storey_range_category", "Storey", estrelation$category)

# Plot the results with flipped coordinates for readability
ggplot(estrelation, aes(x = reorder(category, estimate), y = estimate)) +
  geom_bar(stat = "identity", fill = "red", color = "black") +
  coord_flip() +
  labs(
    title = "Expressway Proximity & Storey Levels Effects",
    y = "Estimated Change in Price (SGD)",
    x = "Proximity & Storey Levels"
  )
```


Expressway Proximity & Storey Levels Effects
Proximity & Storey Levels
Dist>500m:StoreyVery High (>30)
Dist301–500m:StoreyVery High (>30)
Dist101–150m:StoreyVery High (>30)
Dist51–100m:StoreyVery High (>30)
Dist301–500m:StoreyLow–Mid (07–12)
Dist301–500m:StoreyMid (13–18)
Dist301–500m:StoreyLow (01–06)
Dist151–300m:StoreyMid (13–18)
Dist51–100m:StoreyLow–Mid (07–12)
Dist51–100m:StoreyMid (13–18)
Dist>500m:StoreyLow–Mid (07–12)
Dist151–300m:StoreyLow–Mid (07–12)
Dist>500m:StoreyMid (13–18)
Dist51–100m:StoreyLow (01–06)
Dist151–300m:StoreyLow (01–06)
Dist>500m:StoreyLow (01–06)
Dist151–300m:StoreyVery High (>30)
Dist101–150m:StoreyMid (13–18)
Dist101–150m:StoreyLow–Mid (07–12)
Dist101–150m:StoreyLow (01–06)
−1e+05
−5e+04
0e+00
5e+04
1e+05
Estimated Change in Price (SGD)

The analysis, illustrated in Figure @ref(fig.storeyproximity), focuses on the relationship between expressway proximity and storey range. This model uses flats located >500m from expressways and categorized as "High" storey as reference points. Estimated price changes are compared to these reference flats.

Based on Figure (**ref?**)(fig.storeyproximity), flats closer to expressways and on lower storeys tend to be the least valued. These flats are often impacted by noise, pollution, and reduced social status. In contrast, flats on higher storeys and farther from expressways command higher values, likely due to quieter environments, better views, and enhanced privacy. The impact of these changes are more pronoucened as we traverse in highest and lowest storey range and distance from expressway–as we see with **Dist101-150m:StoreyLow(01-06)** and **Dist>500m:StoreyVery High(>30)**.

These pattern underscores the trade-offs between convenience and desirability in Singapore's dense housing market.

# Conclusion

The analysis of Singapore's public housing resale prices reveals significant relationships between proximity to expressways, storey levels, and housing prices. Flats located further from expressways tend to have higher resale prices, likely due to reduced noise and pollution, reflecting residents' preferences for quieter environments. Similarly, flats on higher storey levels, especially those categorized as "Very High," command premium prices, possibly driven by better views, privacy and social status.

The interaction between expressway proximity and storey levels further highlights that housing prices are influenced by a combination of factors, with preferences for quieter environments and more desirable locations playing a critical role. These findings underscore the importance of considering both proximity to infrastructure and housing design when evaluating property values. As a result, Singaporean housing markets and its public transportation relationships are not applicable to U.S. cities like New York City, Chicago and San Francisco, as the two culture, regulations and financial demographics are entirely different.

Future research could incorporate additional variables, such as median salary in respective areas, access to MRT stations and amenities, to provide a more comprehensive understanding of Singapore's unique housing market.

# Work Cited


AGENCY, FEDERAL HOUSING FINANCE. 2024. “U.s. House Prices Rise.” https://www.fhfa.gov/news/news-release/u.s.-house-prices-rise-4.3-percent-over-the-prior-year-up-0.7-percent-from-the-second-quarter-of-2024#:~:text=Washington%2C%20D.C.%20%E2%80%93%20U.S.%20house%20prices,the%20second%20quarter%20of%202024.

Becker_etal., Sofia. 2013. “The New Real Estate Mantra Location Near Public Transportation.” https://www.fool.com/money/research/car-ownership-statistics/?furi=%2Fbuying-stocks%2F&ltyp=txt&luri=%2Fnewsroom%2F.

Caporal, Jack. 2024. “Car Ownership Statistics.” https://www.fool.com/money/research/car-ownership-statistics/?furi=%2Fbuying-stocks%2F&ltyp=txt&luri=%2Fnewsroom%2F.

Delmendo, Lalaine C. 2024. “Singapore’s Residential Property Market Analysis 2024.” https://www.globalpropertyguide.com/asia/singapore/price-history.

Department, Statista Research. 2024. “Real Estate - Singapore.” https://www.statista.com/outlook/fmo/real-estate/singapore.

Goh, Adison. 2024. “Resale Flat Prices with Distance from Expressway.” https://www.kaggle.com/datasets/adisongoh/resale-flat-prices-with-distance-from-expressway/data.

worldometers. 2022. “GDP Per Capita.” https://www.worldometers.info/gdp/gdp-per-capita/.