**Modeling the Impact of Factors on Housing Prices in Qianzhen District, Kaohsiung**  
　111ZU1059 Chou Yi Lin

112266002 Eric

111405182 Tzu Jui Wang

113266005 TENG Yun

112266009 Sawa Tatsuki

1. **Introduction**

In recent years, housing prices in Taiwan have experienced a rapid and sustained increase, prompting widespread concerns regarding housing affordability and the sustainability of urban development. According to the Ministry of the Interior (MOI), the national housing price index has shown a consistent upward trend since 2017, driven by factors such as persistently low interest rates, speculative investment in real estate, and urban population concentration (MOI, 2024). These developments have disproportionately affected younger and middle-income households, who increasingly find homeownership to be out of reach.

Among the cities undergoing notable transformation, Kaohsiung stands out for its rapid urban redevelopment. In particular, the Qianzhen District has transitioned from a primarily industrial zone into a mixed-use area combining residential, commercial, and public infrastructure. The establishment of transportation facilities such as Light Rail Transit has altered the landscape of the district, potentially reshaping the determinants of housing prices.

This study focuses on identifying the key drivers of housing prices in Qianzhen District by analyzing actual transaction data from 2017 to 2024. Using a multiple linear regression model, we examine the influence of property attributes such as building area, land area, number of rooms, age of the house, and the presence of parking or property management services. The goal is to provide a data-driven assessment of which factors significantly affect the unit price per square meter. The findings are expected to offer practical implications for urban policy, real estate development, and investment decision-making.

1. **Research Problem**

The primary aim of this research is to investigate the structural and locational factors that drive housing price variation in Kaohsiung’s Qianzhen District. As housing prices continue to climb, understanding the composition of price determinants becomes increasingly important for ensuring equitable and efficient urban development. This research is motivated by the need to provide empirical evidence on how various features of a property contribute to market value in a rapidly evolving urban environment.

We focus on the following core question:  
Which property-specific factors contribute most significantly to rising housing prices in Qianzhen District?

To answer this question, we analyze a set of key variables, including building area, land area, number of rooms and bathrooms, house age, and whether the property includes amenities such as parking or a management organization. The study employs a multiple linear regression model to estimate the extent to which each factor affects the unit price of housing, while controlling for the effects of other variables.

The specific research objectives are as follows:

1. To assess the relationship between structural characteristics (e.g., house age, room layout) and housing prices.
2. To evaluate the impact of additional features such as parking spaces and the presence of management organizations on property values.
3. To investigate whether traditional price determinants remain significant in an urban district undergoing rapid redevelopment.

By addressing these objectives, this research contributes to a better understanding of price formation mechanisms in urban housing markets and offers guidance for stakeholders including homebuyers, urban planners, and policymakers.

1. **Data Description**

3.1 Data Source and Scope

Property transaction records for this analysis were obtained through the Real Estate Actual Price Registration System (MOI, Taiwan) website, which provides public information on real estate transactions nationwide and records the actual transaction prices, avoiding asymmetry of market information and ensuring market transparency. The data covers detailed information such as the area of land and building transfer, the number of rooms, the age of the house, and the parking space. This study selects the transaction data from January 2017 to November 2024, which covers the housing price information in Qianzhen District, to explore the influence of Kaohsiung's LRT system on residential property prices.

3.2 Key Variables

1. Total Area of Land Transferred

As a key element in real estate transactions, the size of land directly affects the development potential and market value of a property. Generally speaking, a larger land area provides more flexibility in utilization and may exert an upward effect on property prices, whether it is used as a building site or has the potential to increase in value in the future. Therefore, the coefficient is expected to be positive and is expressed in “m²”.

1. Total Area of Transferred Buildings

Building Area is a key indicator of the amount of usable space in a home. Larger building area usually means more living space, which can meet the needs of different family structures and lifestyles, thus increasing the value of the property. Therefore, the coefficient is expected to be positive and is expressed in “m²”.

1. Current Building Layout (Room)

The number of rooms plays a crucial role in shaping buyers' decision-making when purchasing a home. A greater number of rooms generally enhances living functionality and boosts market demand, which is expected to positively influence home prices. As a result, the coefficient is anticipated to be positive, with the unit measured in count.

1. Current Building Layout (Bathroom)

The number of bathrooms plays a crucial role in shaping buyers' decision-making when purchasing a home. A higher number of bathrooms typically improves living convenience and overall functionality, especially for larger households, which can increase a property's market appeal. Consequently, the coefficient is expected to be positive, with the unit measured in count.

1. Current Building Layout (Living Room)

The number of living rooms plays a significant role in influencing buyers' decision-making when purchasing a home. More living room space generally enhances comfort and facilitates social interaction, which can make a property more attractive to potential buyers. As a result, the coefficient is expected to be positive, with the unit measured in count.

1. Age of Estates (age)

This study also includes age as an important variable affecting housing prices. In general, newer homes are preferred because they have better building quality and amenities, lower maintenance costs, and are more in tune with modern living needs. In general, property age is anticipated to have a negative influence on housing prices, i.e., housing prices are likely to decline as homes age. Therefore, the expected coefficient is negative and is expressed in “years”.

1. Parking Space

The availability of a parking space is also a key factor in the price of a home. As urban development and vehicle ownership increase, the need for parking increases, especially in areas where parking space is limited, and the added value of a parking space is significant. As a result, properties with parking spaces often attract more buyers and increase prices. It is anticipated that this variable will positively impact home prices, resulting in a positive coefficient. This variable will be represented as a dummy variable, where owning a parking space will be represented as 1 and not owning a parking space will be represented as 0, in units of “1” and “0”.

1. Management Organization

The presence of a management organization is also a key factor in the pricing of a home. In residential communities, especially in urban settings, a well-structured management organization can enhance property maintenance, security, and overall living quality. This added value can make properties with such organizations more appealing to buyers and drive up housing prices. Therefore, the existence of a management organization is expected to have a positive impact on home prices, resulting in a positive coefficient. This variable will be represented as a dummy variable, where the presence of a management organization is coded as 1, and its absence as 0, in units of “1” and “0”.

1. Availability of Light Rail Transit stations

Light Rail Transit (LRT) system can enhance regional accessibility and increase the price of housing, especially due to the lower operating cost and flexibility of LRT. According to Al Mosaind (1993), Du et al. (2006), the price of housing within 500 meters of a LRT station exerts a positive influence on housing prices relative to properties located beyond 500 meters. Therefore, the study considers a house within 500 meters of a Light Rail Transit station as having Light Rail Transit, and the expected coefficient is positive. This variable is represented as a dummy variable, taking the value of 1 if a Light Rail Transit station is present, and 0 otherwise, with units of “1” and “0”.

1. Distance from Light Rail Transit stations

Generally speaking, homes closer to Light Rail Transit stations enjoy greater commuting convenience, which can help boost market demand and prices. However, if the proximity is too great, negative externalities such as noise, privacy disturbances or crowds of people may affect the quality of living, which in turn suppresses housing prices. Therefore, it is possible that a positive effect may be expected only if the Light Rail Transit station is located beyond a certain distance. It is therefore expected that the coefficient is not necessarily positive or negative and is expressed in meters.

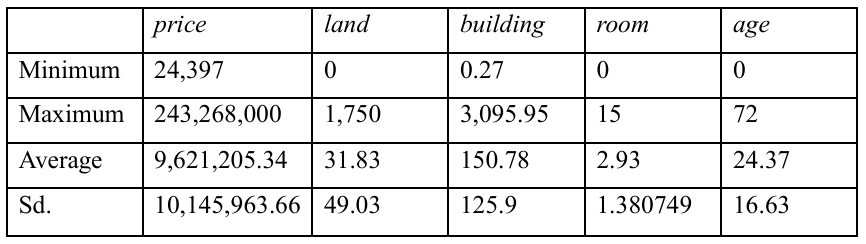
3.3 Data Processing

To ensure the robustness of the regression analysis, the data underwent preprocessing. Outliers, such as transactions exceeding 1.4 billion NTD or having over 400 rooms, were removed. Missing values (NA) were replaced with zeros. Binary variables were created to indicate the presence of parking space and management organization.

As for enhancing data accuracy and applicability, this study initially transformed the transaction addresses into XY coordinate values, and excluded unrecognizable addresses and missing data to ensure the correctness of the geographic information. After screening and cleaning, 9,642 valid transactions were obtained, which served as the basis for the subsequent empirical analysis.

**4. Descriptive Statistics**

Table 1. Narrative Statistics of Building Properties



The lowest total housing price was $24,397 and the highest was $243 million. The average transaction price was $9,621,200 with a standard deviation of $10,146,000, indicating a significant degree of variation in housing prices. In terms of building space, the transferable area of buildings ranged from 0.27 square meters to 3,095.95 square meters, with an average of 150.78 square meters and a standard deviation of 125.9 square meters, reflecting that the size of the objects of different transactions varied greatly. The minimum area of land transfer was 0 square meters and the maximum was 1,750 square meters, with an average of 31.83 square meters and a standard deviation of 49 square meters, indicating that some of the transactions did not include land ownership. In terms of spatial pattern, the maximum number of rooms was 15, the minimum was 0, and the average was about 3. In terms of age, the oldest building in the sample was 72 years old and the youngest was a new building, with an average age of about 24 years and a standard deviation of 16.63 years, indicating that the data covered buildings of different ages.

1. Methodology

To examine the relationship between LRT access and housing prices from multiple variables, four models are specified. These models differ in how they operationalize the concept of LRT proximity—either as a binary indicator, a linear distance measure, a nonlinear distance specification, or a segmented (threshold-based) distance function. Each model is described below along with its rationale.

#### **Model 1: Binary LRT Accessibility Indicator**

The first model includes a **binary variable lrt**, which equals 1 if a property is located within 500 meters of an LRT station and 0 otherwise. The model tests whether simply being near a station confers a price premium or discount. The specification is:

**model1 <- lm(`總價` ~ 房 + 廳 + 衛 + 車位 + 有無管理組織 + 屋齡 + 土地移轉面積 + 建物移轉面積 + lrt, data = data)**

**summary(model1)**

This approach is intuitive and policy-relevant, commonly used in urban planning studies to assess “walkable access” effects on real estate value.

#### **Model 2: Linear Distance to LRT (lrtdis)**

The second model replaces the binary variable with a **continuous distance variable lrtdis** (measured in meters), representing the direct distance from each property to the nearest LRT station. The model estimates the marginal effect of each additional meter of distance from a station:

**Price = β₀ + β₁⋅Rooms + β₂⋅Living Rooms + β₃⋅Bathrooms + β₄⋅Parking + β₅⋅Management + β₆⋅Age + β₇⋅Land Area + β₈⋅Building Area + β₉⋅LRT (binary) + ε**

This specification allows a more nuanced understanding of how distance influences value, assuming the effect is constant across the range.

#### **Model 3: Quadratic Distance Specification**

The third model incorporates both lrtdis and its squared term lrtdis² to account for potential nonlinear effects of distance:

**model3 <- lm(`總價` ~ 房 + 廳 + 衛 + 車位 + 有無管理組織 + 屋齡 + 土地移轉面積 + 建物移轉面積 + lrtdis + I(lrtdis^2), data = data)**

**summary(model3)**

This formulation allows for a curvature in the price-distance relationship. It is especially relevant when proximity to LRT stations initially increases housing value, but the effect may diminish—or even reverse—beyond a certain range due to externalities such as noise or congestion.

#### **Model 4: Segmented Regression with Estimated Threshold**

Finally, the fourth model applies a segmented linear regression approach, using the segmented package in R to estimate a structural break point in the relationship between distance and price:

**library(segmented)**

**base\_model <- lm(`總價` ~房 + 廳 + 衛 + 車位 + 有無管理組織 + 屋齡 + 土地移轉面積 + 建物移轉面積 + lrtdis, data = data)**

**model4 <- segmented(base\_model, seg.Z = ~lrtdis)**

**summary(model4)**

**plot(model4)**

Here, LRT Distance represents the slope change after an estimated breakpoint. This method enables the identification of a distance threshold beyond which the effect of LRT distance on housing price significantly changes.

Together, these four models allow for a comprehensive understanding of how LRT proximity relates to housing prices under varying assumptions.

1. Result and Visualization

**Model 1: Binary LRT Accessibility Indicator**

The first model uses a binary variable lrt to represent whether a property is located within 500 meters of a Light Rail Transit (LRT) station. This approach tests for the presence of a simple walkability premium by comparing properties near stations to those further away. The regression results indicate that proximity within this threshold has a significantly positive impact on total housing price, with the lrt variable showing a positive coefficient (Estimate ≈ 454,629) and p-value < 0.01. This suggests that properties within walking distance to LRT stations are valued higher, likely due to improved accessibility and convenience.

The model’s R-squared value is 0.6828, indicating that approximately 68% of the variation in housing prices can be explained by the included variables. While this binary approach is intuitive and commonly used in planning literature, it assumes a fixed benefit cutoff and may not capture the full spatial dynamics of accessibility.

**Model 2: Linear Distance to LRT**

The second model replaces the binary variable with a continuous variable lrtdis, representing the exact distance (in meters) from each property to the nearest LRT station. This allows for a more refined estimation of how housing prices respond to incremental changes in accessibility. The results show that lrtdis has a statistically significant negative effect on total price (Estimate ≈ –1,239; p < 0.001), meaning that each additional meter of distance reduces the property’s value.

This finding aligns with expectations from urban economic theory, where accessibility is capitalized into land prices. With an R-squared of 0.6872, the model fits slightly better than the first, suggesting that a continuous measure of distance provides more explanatory power than a binary threshold alone.

**Model 3: Quadratic Distance Specification**

In Model 3, a quadratic term lrtdis² is introduced alongside the linear distance variable to capture potential nonlinearities in the relationship between distance and housing price. The results show that the squared distance term is highly significant and negative (p < 0.001), while the linear term becomes statistically insignificant. This implies a curved relationship: housing prices decline slowly with distance initially, but the negative impact becomes stronger at greater distances.

This formulation reflects real-world conditions where very close proximity may bring benefits (e.g., convenience), but beyond a certain point, the marginal accessibility diminishes and other factors (e.g., noise, reduced connectivity) may dominate. The R-squared rises to 0.6882, supporting the usefulness of including a nonlinear term to better model spatial variation.

**Model 4: Segmented Regression with Estimated Threshold**

The fourth model employs segmented regression to detect and estimate a breakpoint in the relationship between distance and price. The estimated breakpoint occurs at approximately 1,518 meters from an LRT station. The slope of lrtdis before this breakpoint is –504, while the slope after is significantly steeper at –2,306, indicating that the negative impact of distance on housing price becomes much more pronounced beyond 1.5 kilometers.

This segmented approach offers a more flexible alternative to polynomial models, directly identifying structural changes in accessibility effects. With an R-squared of 0.6885, the model has the highest explanatory power among the four models, and provides important policy-relevant insights about critical access zones.

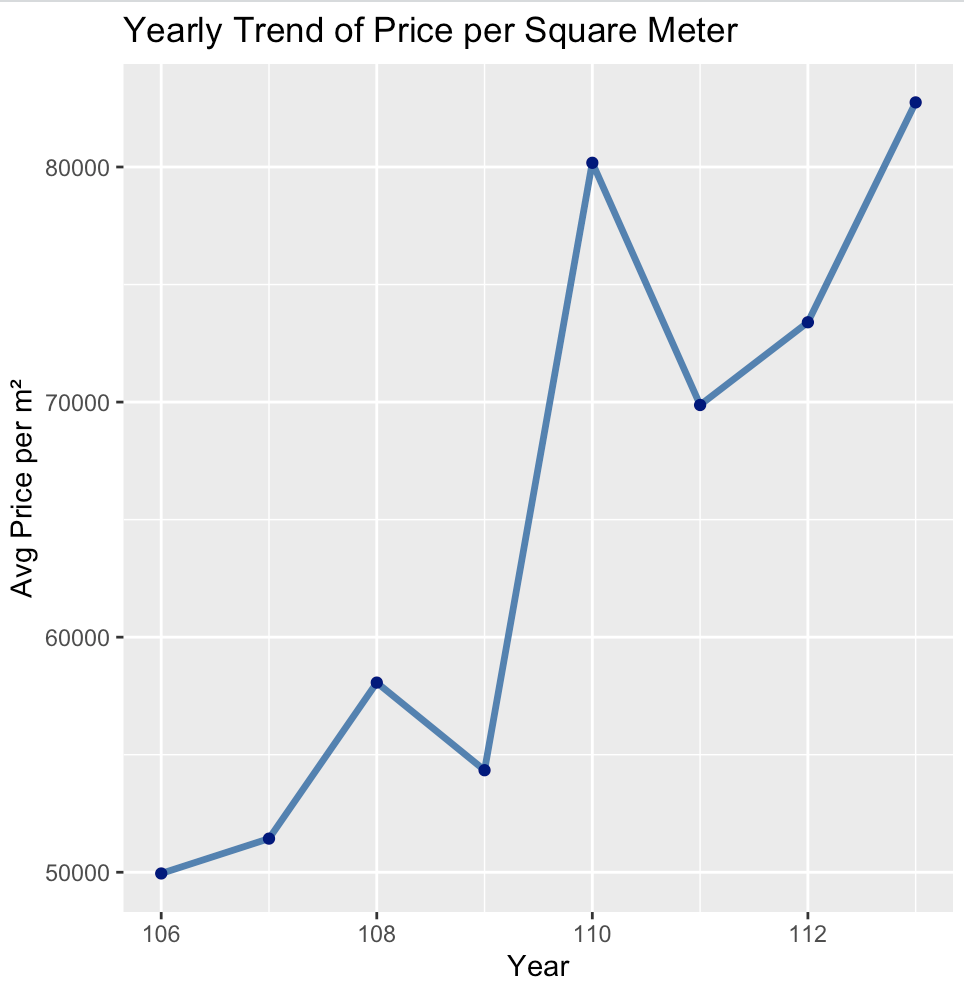
To better understand the factors influencing housing prices, we present a series of visualizations based on our dataset:

**Figure 1** shows the yearly trend of average unit prices from Year 106 to 113. A significant price surge is observed after Year 109, indicating possible market shocks or policy effects.

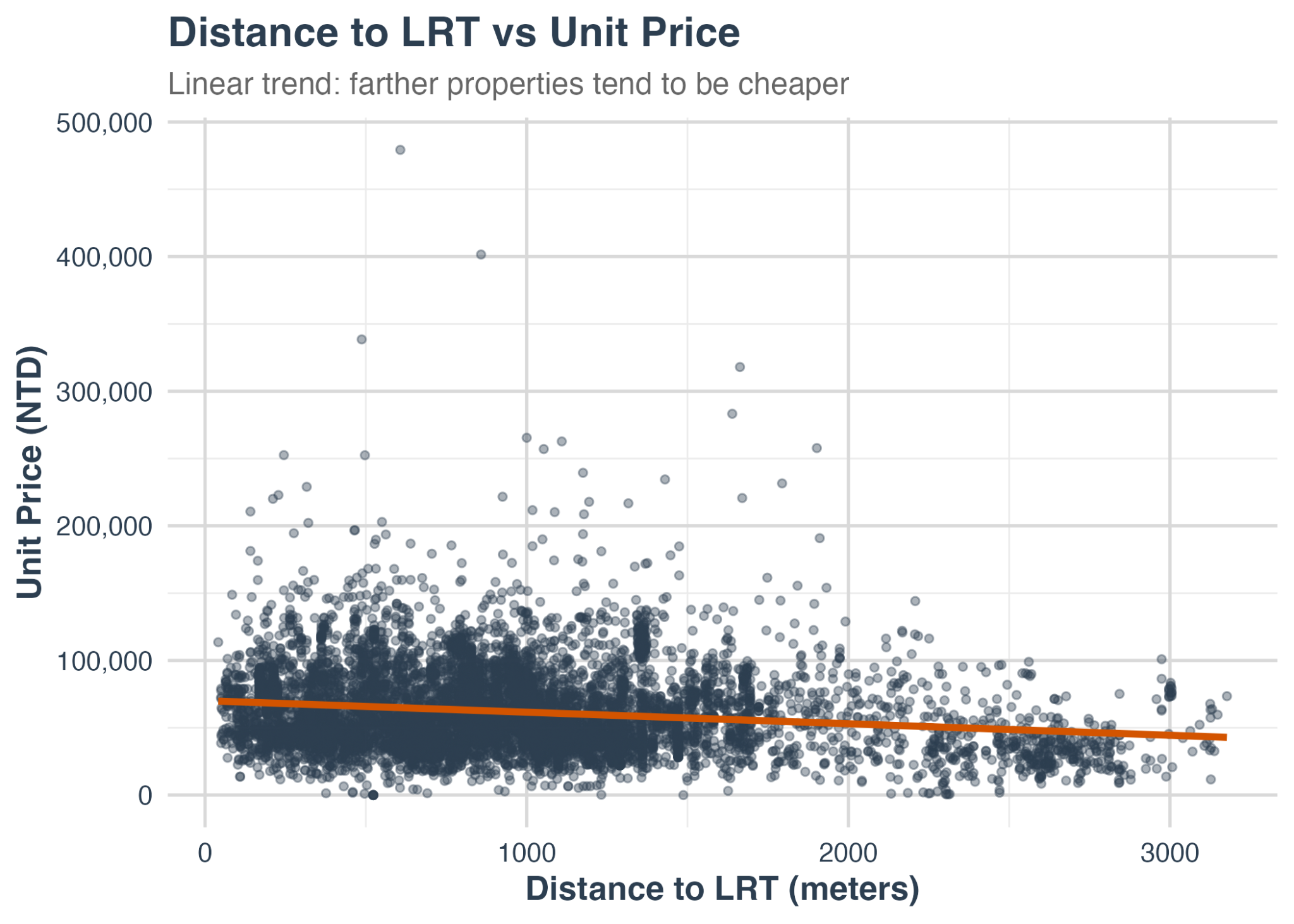
**Figure 2** illustrates the linear relationship between distance to the LRT station and unit price. Properties located farther from the LRT tend to have lower prices, consistent with our regression findings.

**Figure 3** compares unit prices by the presence of a building management organization. Interestingly, properties with management are associated with slightly lower unit prices, possibly due to larger area or different property types.

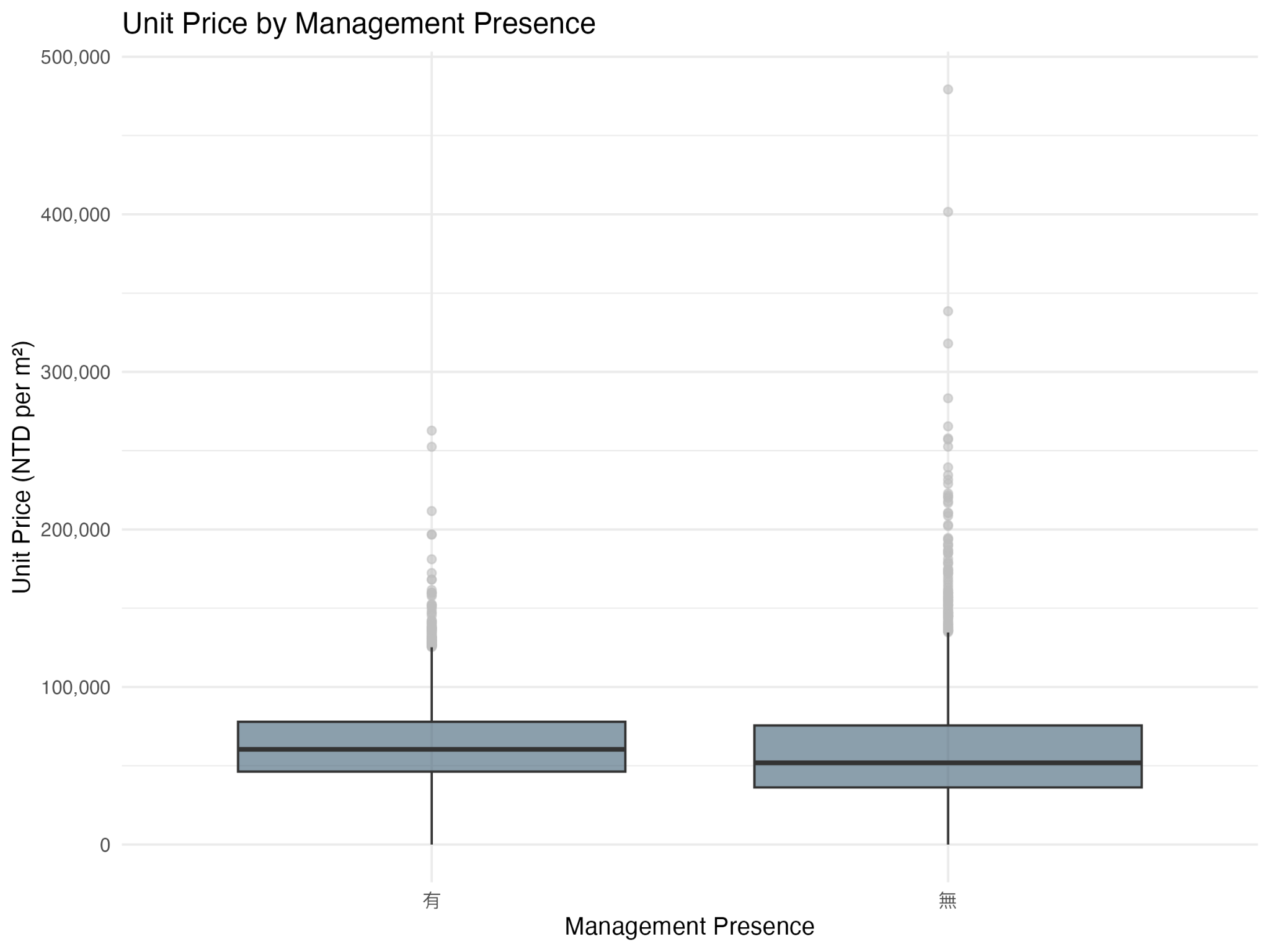
**Figure 4** presents unit prices grouped by parking availability. Homes with parking tend to have higher unit prices, supporting the idea that parking adds significant value to property.



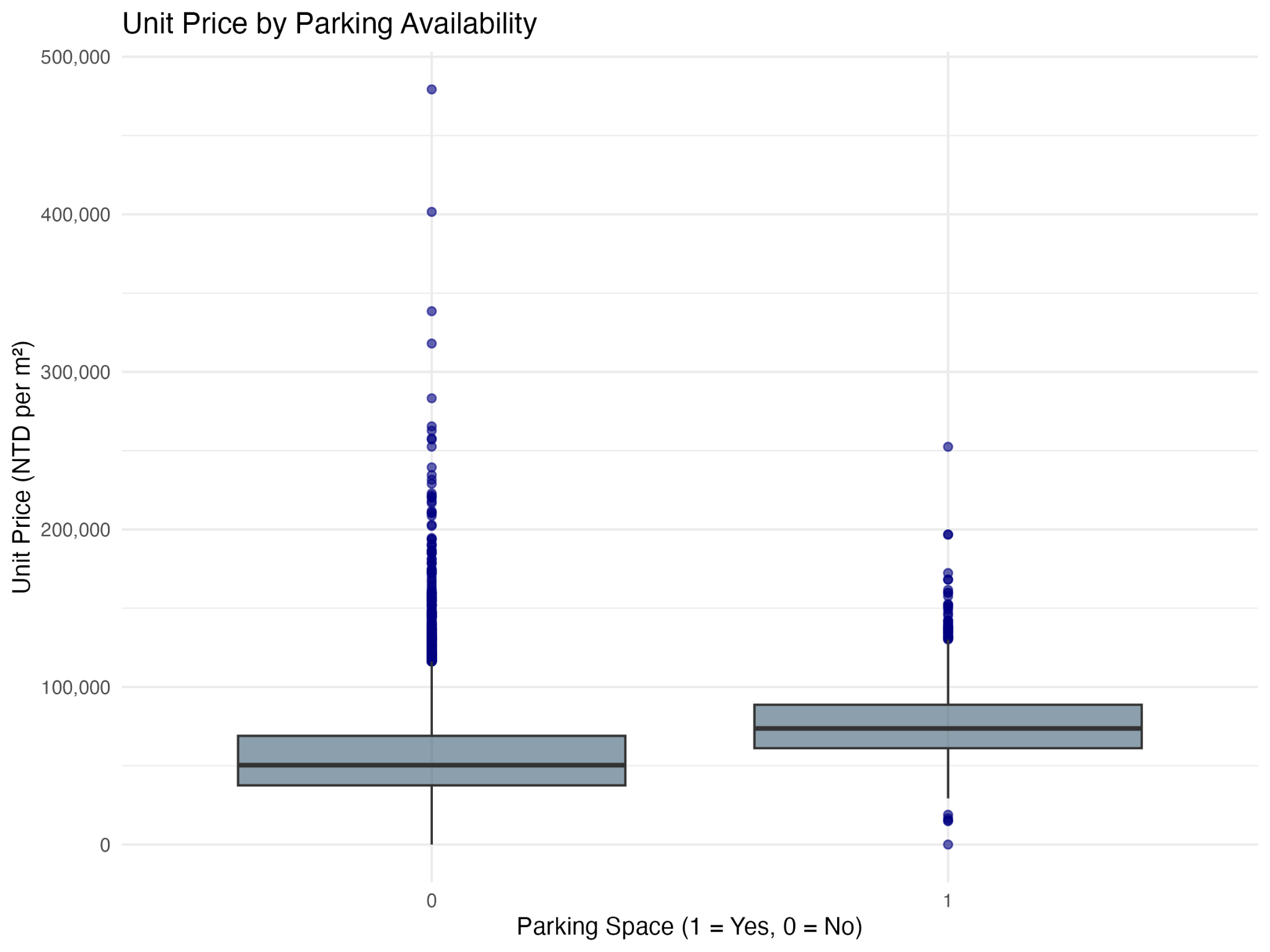
**Figure 1 Yearly Trend of Price per Square Meter**



**Figure 2 Distance to LRT vs Unit Price**



**Figure 3 Unit Price by Management Presence**



**Figure 4 Unit Price by Parking Availability**

1. **Conclusion**

This study investigated the determinants of housing prices in Qianzhen District, Kaohsiung, through a combination of statistical modeling and data visualization. Using real estate transaction data from 2017 to 2024, a multiple linear regression analysis was conducted to evaluate the effects of various structural and locational variables on the unit price per square meter.

The empirical findings revealed that building area and land area positively influence housing prices, while house age has a negative effect, consistent with economic theory and past literature. Parking availability emerged as a strong value-enhancing feature, reflecting buyer preferences for convenience and utility. Interestingly, the presence of a building management organization showed a slight negative association with unit price, potentially due to differences in property size or type.

Several visualizations were used to supplement the regression analysis and offer additional insights. Figure 1 highlighted a sharp increase in average unit prices after Year 109, possibly reflecting external market shocks or policy interventions. Figure 2 demonstrated a clear negative correlation between proximity to the LRT station and housing prices, reinforcing the significance of public transportation access. Figure 3 revealed that properties with management organizations were, on average, associated with slightly lower prices, which may be attributed to confounding factors such as floor area or building category. Figure 4 confirmed that homes with parking had higher unit prices, supporting the conclusion that parking is a critical pricing factor.

Overall, this research contributes to a deeper understanding of the price formation process in an urban housing market undergoing rapid redevelopment. The integration of regression models with graphical analysis provides a robust framework for evaluating property values. These findings have practical implications for urban planners, developers, investors, and policymakers aiming to promote transparent and evidence-based approaches to housing market regulation and development.

**References**  
Ministry of the Interior. (2024). *Real Estate Price Index*. https://pip.moi.gov.tw

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55.