

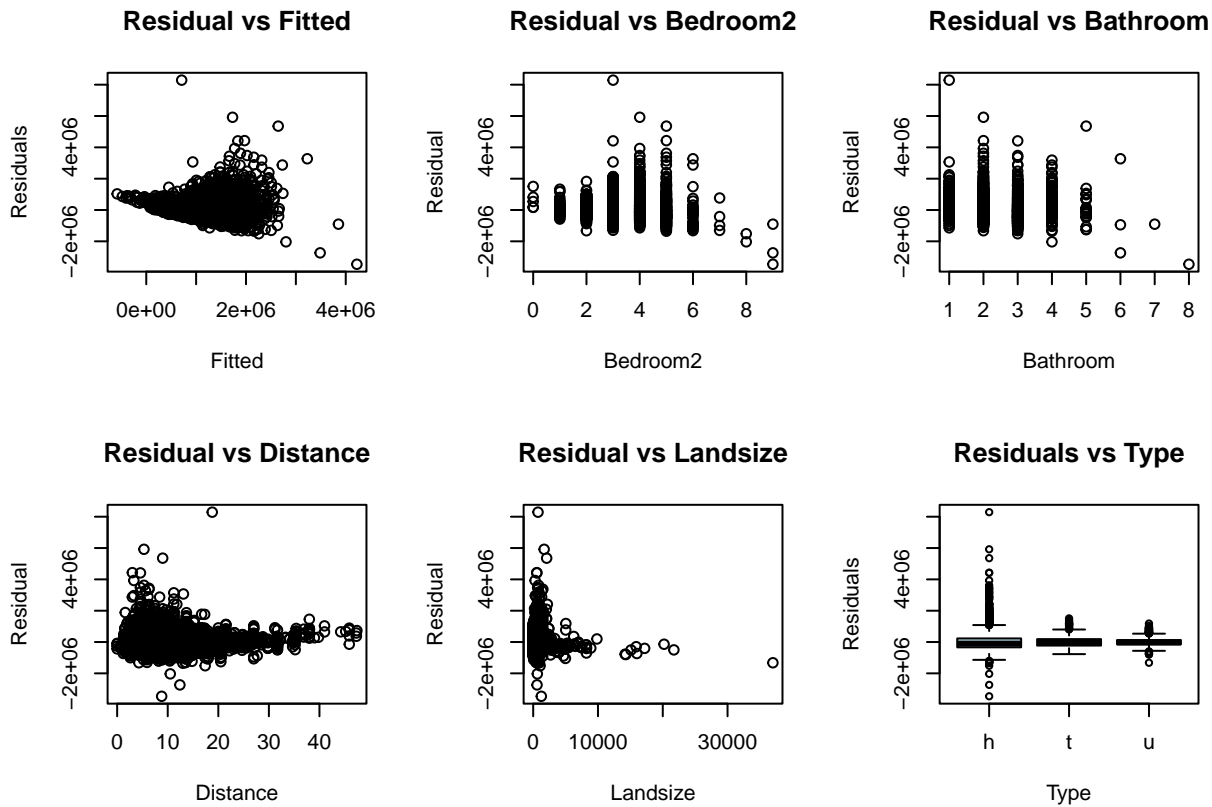
Preliminary Model Analysis

MAMBA OUT

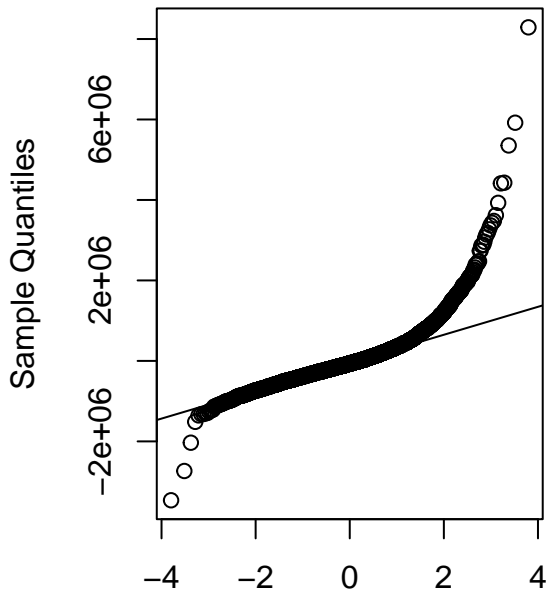
Introduction

A preliminary linear regression model was fitted using `Price` as the response variable, with five predictors: `Bedroom2`, `Bathroom`, `Distance`, `Landsize`, and `Type`. The model aimed to determine significant factors influencing housing prices in Melbourne.

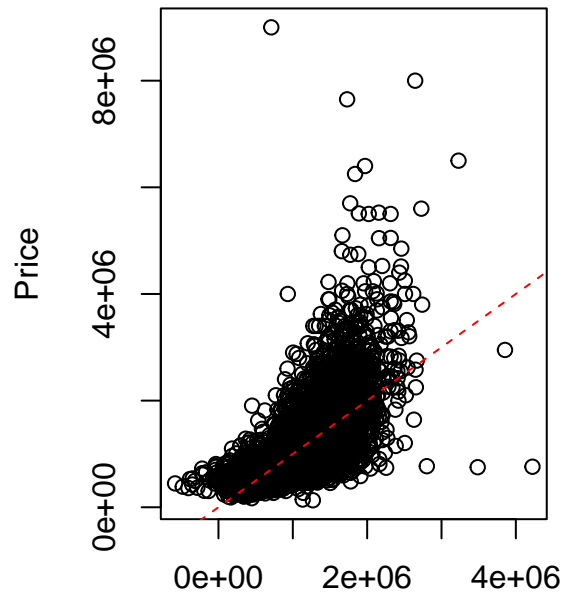
Residual Analysis and Diagnostic Plots



Normal Q-Q Plot



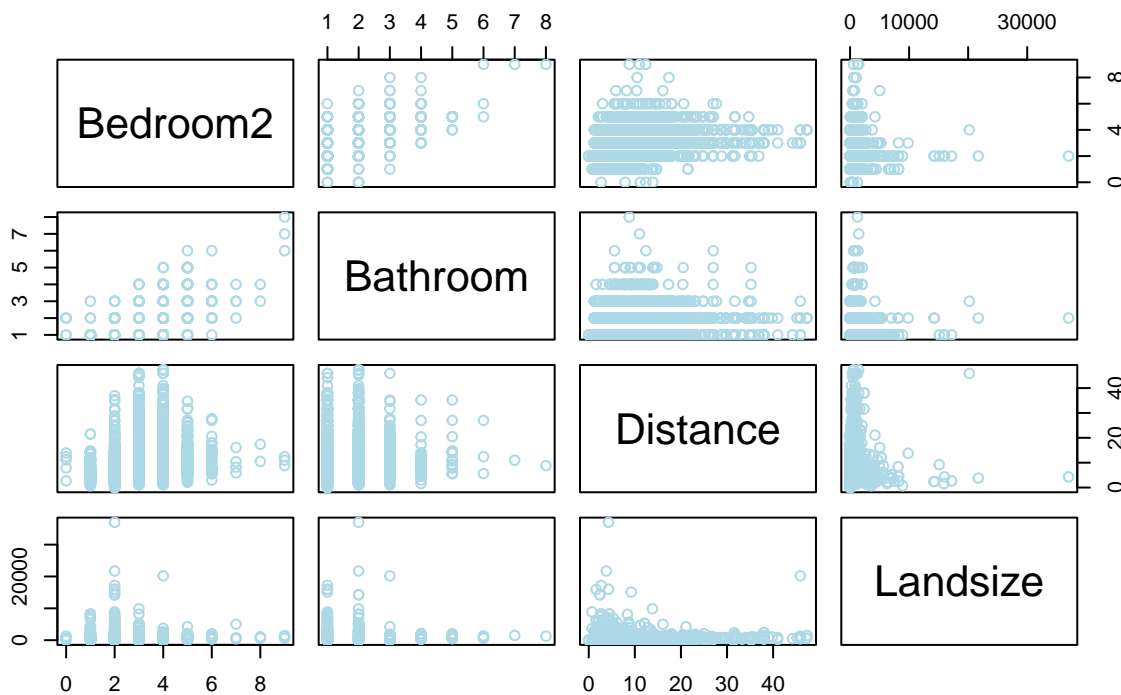
Response vs Fitted



Theoretical Quantiles

Fitted Values

Pairwise Plots of Predictors



Assumption Checks and Residual Analysis

1. **Linearity:** The **Residual vs Fitted** plot shows a noticeable pattern rather than random scatter. Residuals fan out at higher fitted values, indicating that the model does not fully capture the linear relationship between predictors and response.
2. **Homoscedasticity:** The residual spread in the **Residual vs Fitted** plot suggests heteroscedasticity, where variance increases with fitted values. This “funnel” shape indicates that residuals are not evenly distributed, violating the assumption of constant variance.
3. **Normality of Residuals:** The **Q-Q plot** deviates significantly from the diagonal line, especially in the tails. This implies non-normal distribution of residuals, particularly for extreme values, which affects the validity of statistical

inferences.

4. **Multicollinearity:** The **pairwise scatter plots** among predictors, especially between **Bedroom2** and **Bathroom**, indicate clustering and suggest strong correlations. This could lead to multicollinearity, making coefficients less reliable.

Comparison with Literature In the literature, predictors like **Bathroom** and **Landsize** are often significant in explaining house prices, typically yielding strong model fits. However, the current model shows deviations from linear regression assumptions, indicating that the Melbourne housing data may need more complexity to capture these relationships properly. Unlike typical results, residual plots indicate underfitting and issues like non-linearity and heteroscedasticity.

Conclusion The residual plots indicate that several assumptions of linear regression are violated, suggesting the model is not sufficiently capturing the complexity of housing price predictors. Improvements such as variable transformation, adding interaction terms, or using a non-linear model may be necessary to achieve a better fit.