

Distance to Food and Housing Prices

boseongyun

2024-10-10

Contents

Project: Distance to Food and Housing Prices	2
Project Description	2
Data Description	2
Model and Key Variable Description	2
Exploratory Data Analysis	3
Modeling	9
Conclusion	12
Sessioninfo	12

Project: Distance to Food and Housing Prices

Project Description

This project examines whether access to food, measured by proximity to grocery stores, impacts housing prices. We focus specifically on single family households in Connecticut to explore the relationship between housing prices and access to food resources.

Data Description

This analysis uses two primary datasets:

1. Housing data for Connecticut (in POLYGON format) and
2. A directory of stores authorized to accept SNAP benefits, with location coordinates (longitude and latitude).

The datasets are sourced from ATLAS. The housing data is spatially joined with store locations using the `st_nearest()` function, which calculates the nearest store to each housing location based on their geographic coordinates.

Model and Key Variable Description

We employ a linear model to investigate the relationship between housing prices and various property attributes:

1. **Rooms:** Includes bedrooms, bathrooms, half-baths, and total rooms. We anticipate a positive correlation between the number of rooms and housing prices.
2. **Size:** Property size, where larger properties are expected to have higher prices. We exclude farm properties by focusing on single-family households to avoid misclassification.
3. **Zone:** Different zoning regulations (e.g., residential, commercial) that affect property use, including permissions for commercial activities like restaurants, which can influence housing prices.
4. **Actual Year Built (AYB):** Reflects property age. We anticipate that older properties may be less expensive unless they have been renovated.
5. **Condition:** Property condition categories (e.g., Excellent, Poor) are expected to show variation in housing prices.

```

## Import Libraries and Data
library(sf)

## Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE

library(lme4)

## Loading required package: Matrix

# Read Data
dr <- readRDS("cleaned.rds")

```

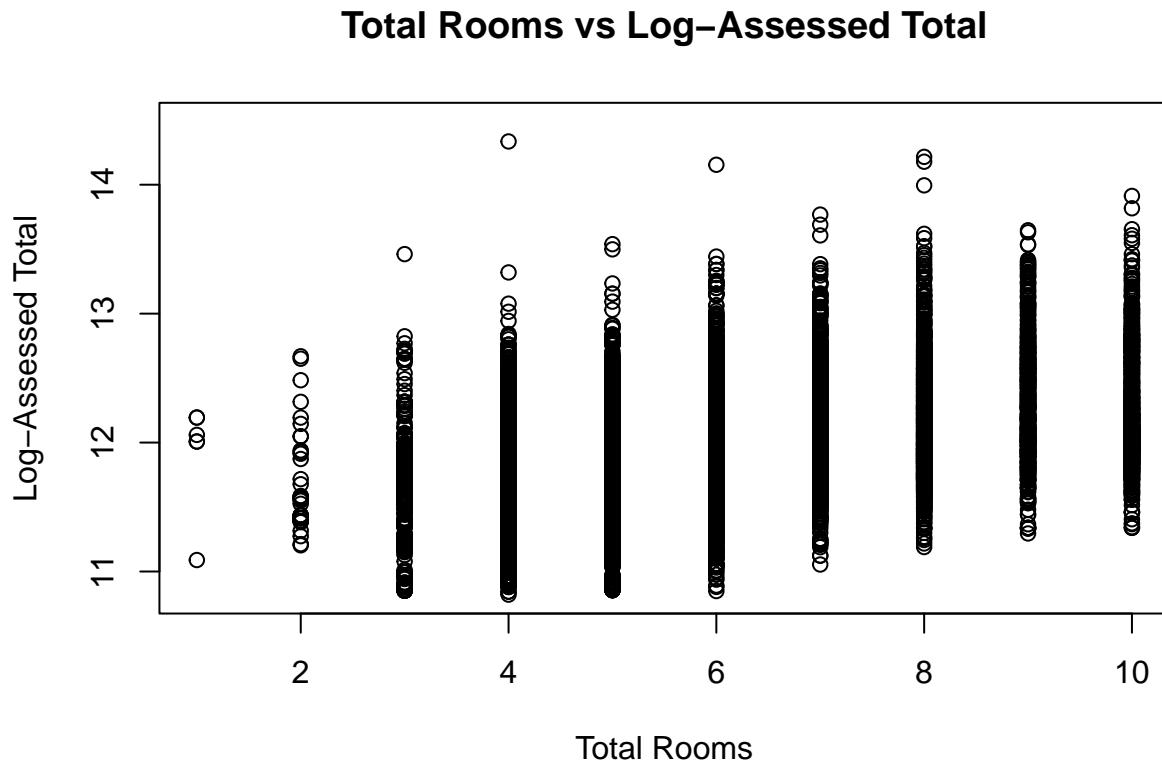
Exploratory Data Analysis

Scatter plot: Total Rooms vs. Log-Assessed Total

```

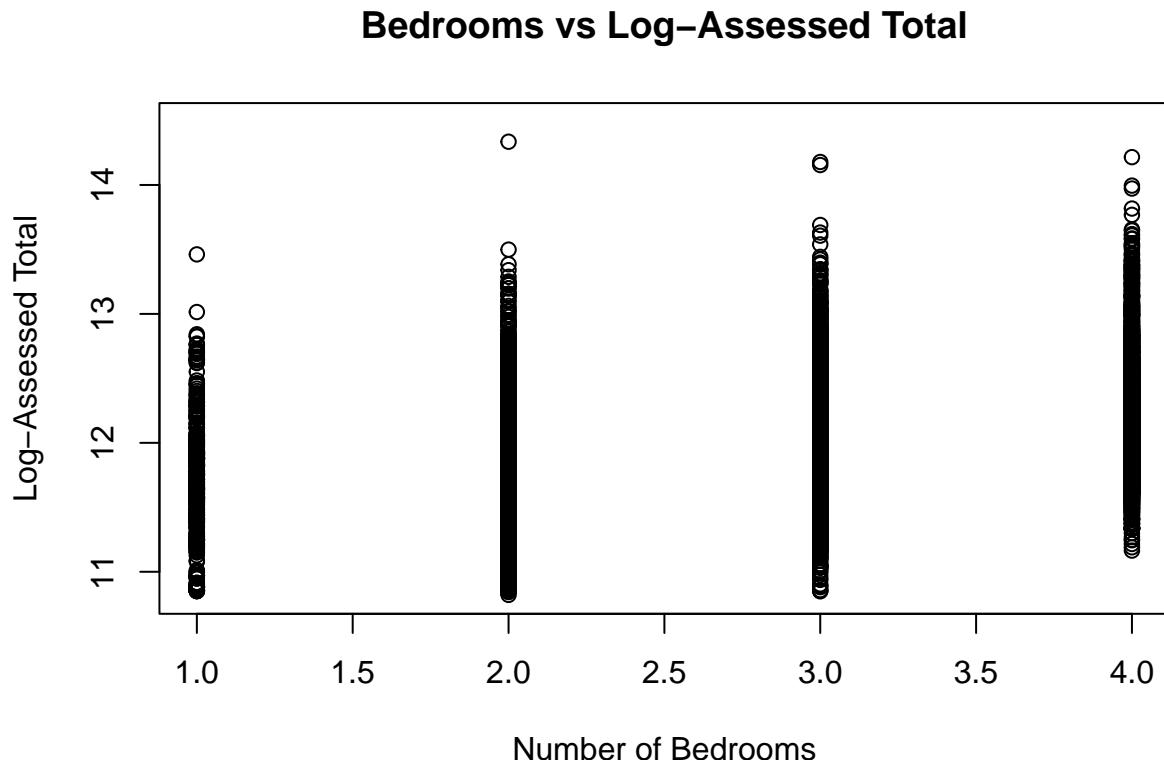
plot(dr$total_rooms, dr$at_log,
      main = "Total Rooms vs Log-Assessed Total",
      xlab = "Total Rooms",
      ylab = "Log-Assessed Total")

```



Scatter plot: Bedrooms vs. Log-Assessed Total

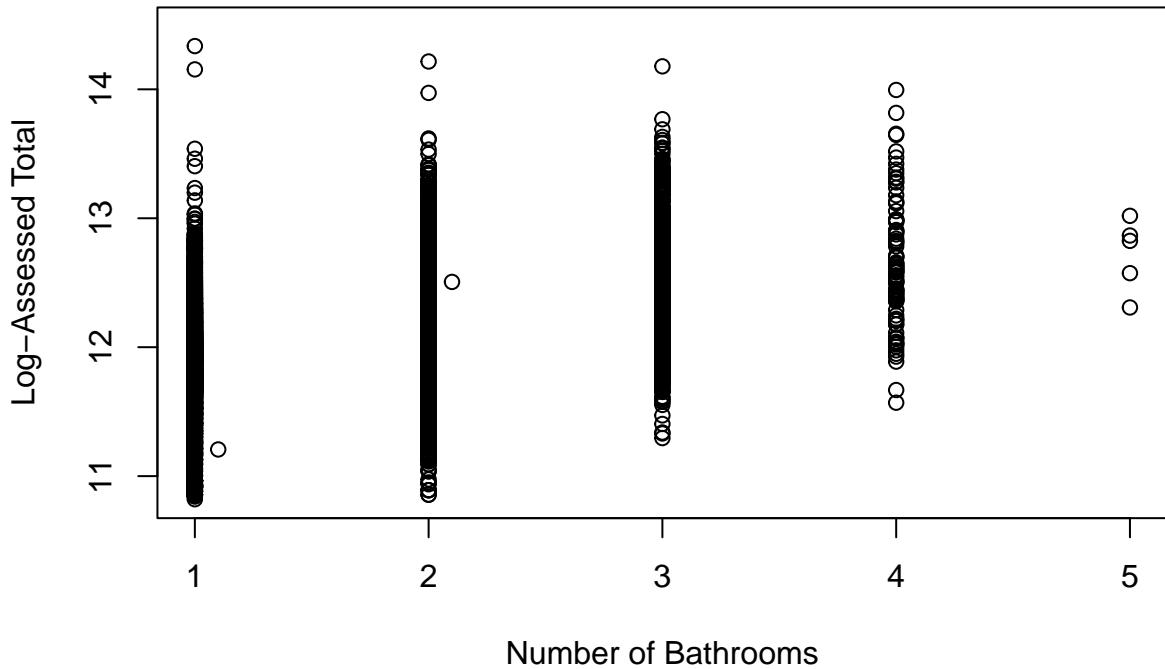
```
plot(dr$number_of_bedroom, dr$at_log,
  main = "Bedrooms vs Log-Assessed Total",
  xlab = "Number of Bedrooms",
  ylab = "Log-Assessed Total")
```



Scatter plot: Bathrooms vs. Log-Assessed Total

```
plot(dr$number_of_baths, dr$at_log,
  main = "Baths vs Log-Assessed Total",
  xlab = "Number of Bathrooms",
  ylab = "Log-Assessed Total")
```

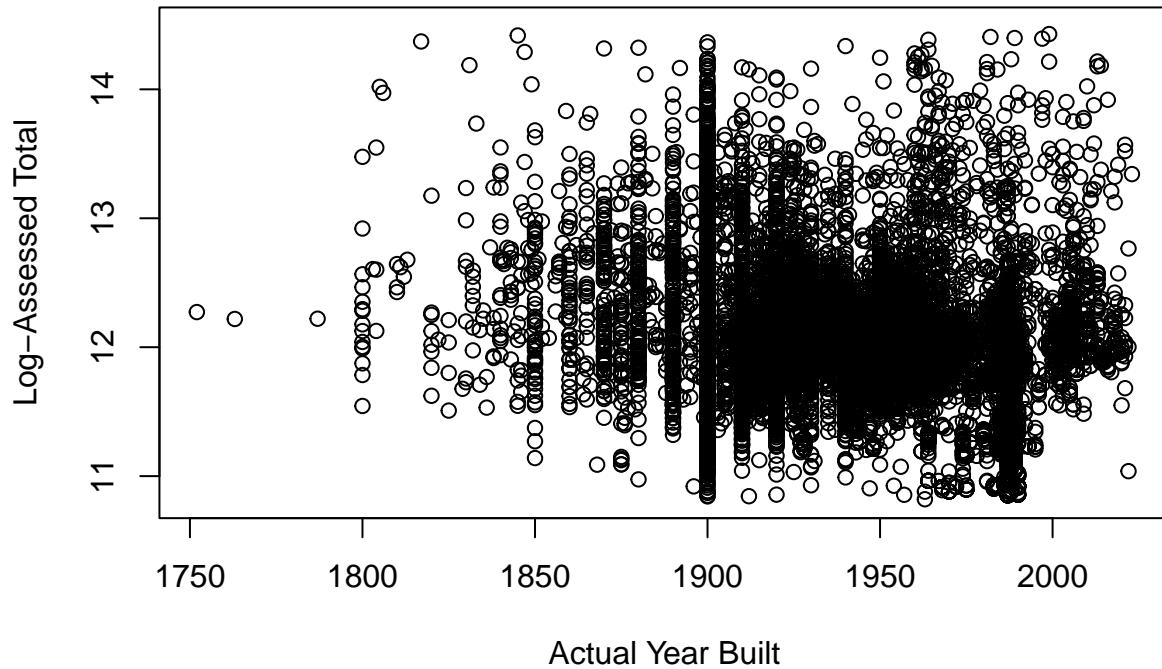
Baths vs Log-Assessed Total



Scatter plot: AYB (Actual Year Built) vs. Log-Assessed Total

```
plot(dr$ayb, dr$at_log,
  main = "AYB vs Log-Assessed Total",
  xlab = "Actual Year Built",
  ylab = "Log-Assessed Total")
```

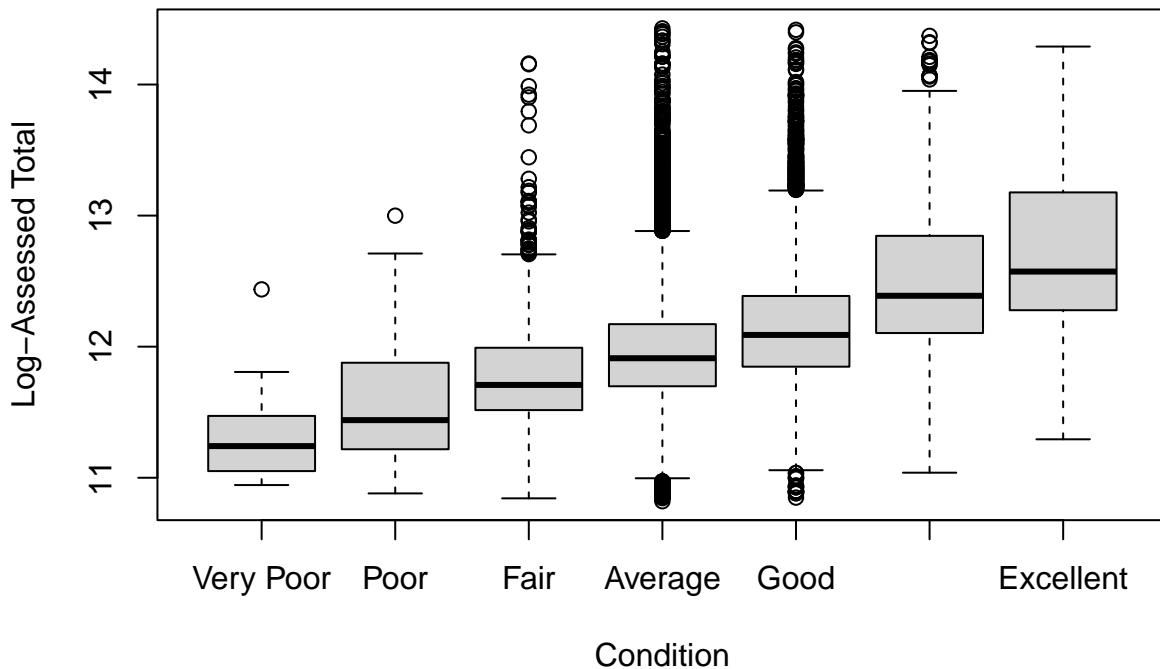
AYB vs Log-Assessed Total



Boxplot: Log-Assessed Total by Condition Description

```
boxplot(dr$at_log ~ dr$condition_description,  
       main = "Boxplot of Log-Assessed Total by Condition",  
       xlab = "Condition",  
       ylab = "Log-Assessed Total")
```

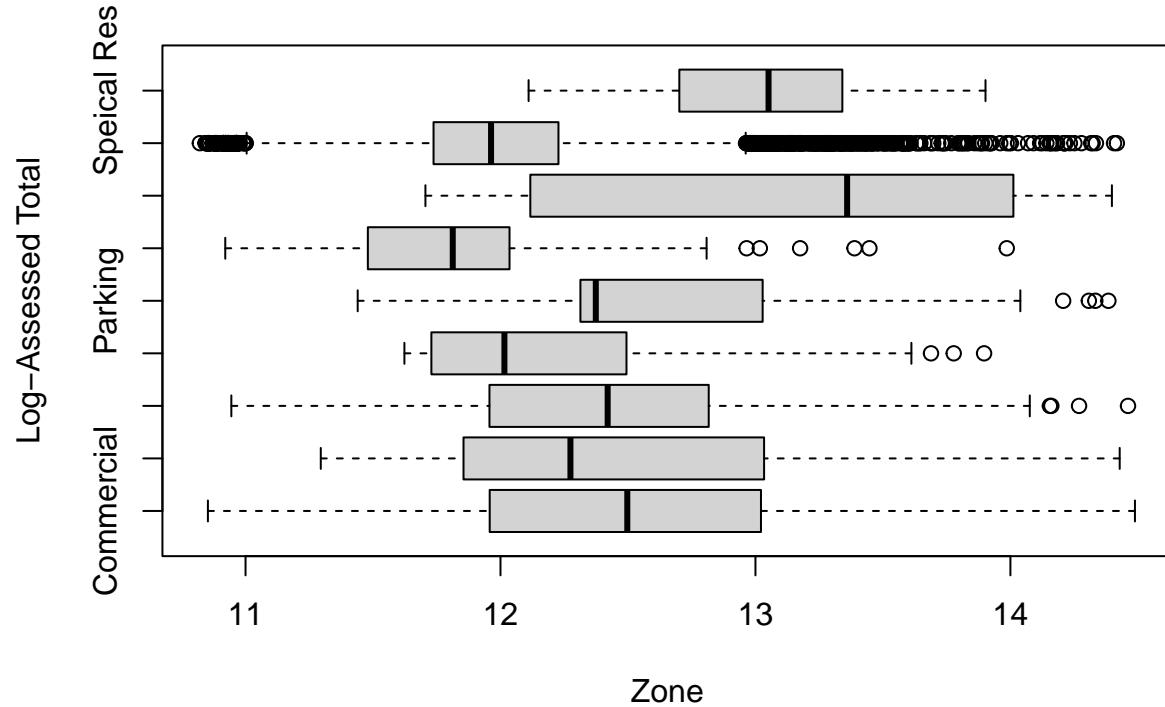
Boxplot of Log-Assessed Total by Condition



Boxplot: Log-Assessed Total by Zone

```
boxplot(dr$at_log ~ dr$zone,
       main = "Boxplot of Log-Assessed Total by Zone",
       xlab = "Zone",
       ylab = "Log-Assessed Total",
       horizontal = TRUE
)
```

Boxplot of Log-Assessed Total by Zone



Modeling

Linear models to explore relationships

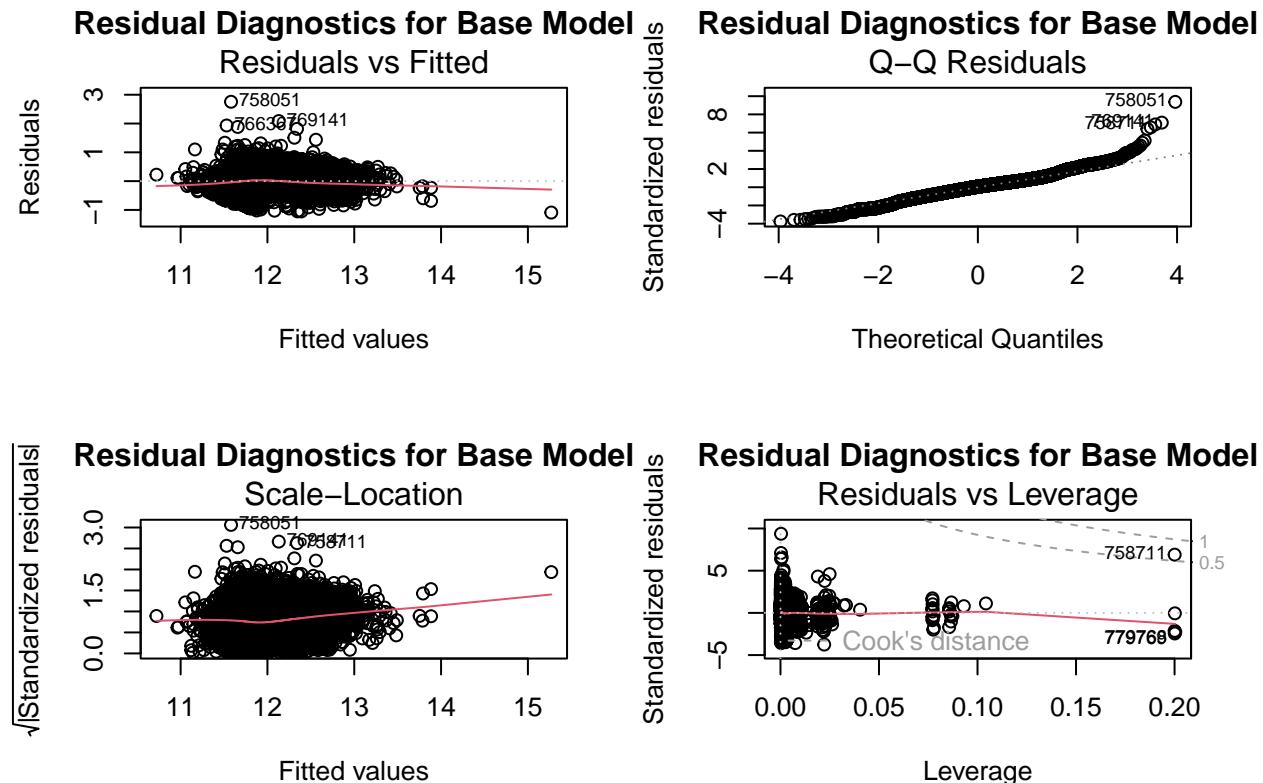
The base model explores the effect of rooms, condition, and zone on the log-transformed assessed total (at_log). The model includes bedrooms, bathrooms, and living area as predictors.

```
# Fit the linear model
mod_b <- lm(at_log ~ total_rooms + number_of_bedroom + number_of_baths +
             condition_description + living_area + zone, data = dr)

# Save and display the summary of the model
mod_b_summary <- summary(mod_b)
mod_b_summary # Display the summary of the model

## 
## Call:
## lm(formula = at_log ~ total_rooms + number_of_bedroom + number_of_baths +
##     condition_description + living_area + zone, data = dr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09114 -0.17712  0.01845  0.17101  2.75492
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.058e+01  8.361e-02 126.528 < 2e-16 ***
## total_rooms                -1.665e-02 2.837e-03 -5.868 4.51e-09 ***
## number_of_bedroom           1.917e-02 5.044e-03  3.800 0.000145 ***
## number_of_baths              6.202e-02 4.986e-03 12.439 < 2e-16 ***
## condition_descriptionPoor  1.678e-01 9.261e-02  1.811 0.070099 .
## condition_descriptionFair  3.769e-01 8.266e-02  4.559 5.18e-06 ***
## condition_descriptionAverage 6.081e-01 8.165e-02  7.448 1.00e-13 ***
## condition_descriptionGood  7.560e-01 8.171e-02  9.252 < 2e-16 ***
## condition_descriptionVery Good 9.465e-01 8.263e-02 11.455 < 2e-16 ***
## condition_descriptionExcellent 1.063e+00 8.585e-02 12.376 < 2e-16 ***
## living_area                 4.244e-04 6.634e-06 63.978 < 2e-16 ***
## zoneHistoric                2.574e-01 2.883e-02  8.926 < 2e-16 ***
## zoneIndustrial              -1.966e-01 4.343e-02 -4.527 6.04e-06 ***
## zoneOthers                  9.376e-03 4.877e-02  0.192 0.847544
## zoneParking                 2.921e-01 4.580e-02  6.377 1.87e-10 ***
## zonePlanned Development    -8.685e-02 2.049e-02 -4.239 2.26e-05 ***
## zonePublic Use               5.839e-01 1.325e-01  4.406 1.06e-05 ***
## zoneResidential              -6.827e-03 1.677e-02 -0.407 0.683991
## zoneSpeical Res              2.620e-01 8.700e-02  3.012 0.002603 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2939 on 13539 degrees of freedom
##   (1734 observations deleted due to missingness)
## Multiple R-squared:  0.52, Adjusted R-squared:  0.5194
## F-statistic: 814.9 on 18 and 13539 DF, p-value: < 2.2e-16
```

```
# Display residual diagnostics
par(mfrow = c(2, 2))
plot(mod_b, main = "Residual Diagnostics for Base Model")
```



```
# Reset the plotting layout to its default
par(mfrow = c(1, 1))
```

Mixed Effects Model

We treat zone as a random effect to model the variability across zones without estimating separate coefficients for each. This approach ensures generalizability and avoids overfitting. Treating it as a fixed effect would result in over 100 coefficients, leading to an overly complex model.

```
# Fit the mixed effects model with zone as a random effect
mod_b_mixed <- lmer(at_log ~ total_rooms + number_of_bedroom +
                      number_of_baths + condition_description +
                      (1 | zone) + living_area, data = dr)

# Display the summary of the mixed effects model
summary(mod_b_mixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: at_log ~ total_rooms + number_of_bedroom + number_of_baths +
##           condition_description + (1 | zone) + living_area
```

```

##      Data: dr
##
## REML criterion at convergence: 5390.2
##
## Scaled residuals:
##      Min     1Q Median     3Q    Max
## -3.7127 -0.6025  0.0619  0.5820  9.3726
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   zone     (Intercept) 0.04651  0.2157
##   Residual           0.08639  0.2939
##   Number of obs: 13558, groups: zone, 9
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                1.069e+01  1.105e-01 96.692
## total_rooms                 -1.669e-02 2.837e-03 -5.883
## number_of_bedroom            1.925e-02 5.043e-03  3.817
## number_of_baths               6.204e-02 4.986e-03 12.444
## condition_descriptionPoor    1.687e-01 9.261e-02  1.822
## condition_descriptionFair    3.775e-01 8.266e-02  4.567
## condition_descriptionAverage 6.090e-01 8.165e-02  7.458
## condition_descriptionGood    7.567e-01 8.171e-02  9.261
## condition_descriptionVery Good 9.473e-01 8.263e-02 11.465
## condition_descriptionExcellent 1.064e+00 8.584e-02 12.395
## living_area                  4.244e-04 6.633e-06 63.986
##
## Correlation of Fixed Effects:
##          (Intr) ttl_rm nmbr_f_bd nmbr_f_bt cndt_P cndt_F cndt_A cndt_G
## total_rooms -0.031
## nmbr_f_bdrm -0.017 -0.593
## nmbr_f_bths -0.018 -0.155 -0.021
## cndtn_dscrP -0.650  0.003 -0.016    0.002
## cndtn_dscrF -0.726  0.002 -0.013   -0.005    0.869
## cndtn_dscrA -0.736  0.011 -0.013   -0.004    0.880  0.986
## cndtn_dscrG -0.736  0.014 -0.012   -0.007    0.879  0.985  0.998
## cndtn_dscVG -0.727  0.018 -0.010   -0.022    0.869  0.975  0.987  0.986
## cndtn_dscrE -0.705  0.020 -0.002   -0.025    0.837  0.938  0.950  0.950
## living_area  0.011 -0.502 -0.031   -0.275    0.004  0.003 -0.003 -0.009
##          cnd_VG cndt_E
## total_rooms
## nmbr_f_bdrm
## nmbr_f_bths
## cndtn_dscrP
## cndtn_dscrF
## cndtn_dscrA
## cndtn_dscrG
## cndtn_dscVG
## cndtn_dscrE  0.939
## living_area -0.012 -0.011

```

Conclusion

Sessioninfo

```
sessionInfo()
```

```
## R version 4.4.1 (2024-06-14)
## Platform: aarch64-apple-darwin20
## Running under: macOS 15.0.1
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;  LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics   grDevices utils      datasets   methods    base
##
## other attached packages:
## [1] lme4_1.1-35.5 Matrix_1.7-0  sf_1.0-17
##
## loaded via a namespace (and not attached):
##  [1] nlme_3.1-164      cli_3.6.3        knitr_1.48       rlang_1.1.4
##  [5] xfun_0.47        highr_0.11       DBI_1.2.3        KernSmooth_2.23-24
##  [9] minqa_1.2.8      htmltools_0.5.8.1 e1071_1.7-14     rmarkdown_2.28
## [13] grid_4.4.1        evaluate_0.24.0  classInt_0.4-10 MASS_7.3-60.2
## [17] fastmap_1.2.0    yaml_2.3.10      compiler_4.4.1  Rcpp_1.0.13
## [21] rstudioapi_0.16.0 lattice_0.22-6   digest_0.6.37   nloptr_2.1.1
## [25] class_7.3-22    splines_4.4.1   magrittr_2.0.3  tools_4.4.1
## [29] proxy_0.4-27    boot_1.3-30     units_0.8-5
```