STA9797_Group_Project_Code

Brandon Kokin, Ayrat Aymetov, Mohammed Saadman Chowdhury

2025-12-18

Module 1: Data Preprocessing

Loading data, cleaning, and preparatory EDA

```
# Loading data, data cleaning, and preparatory Exploratory Data Analysis
# Setup and basic preparation
AmesHousing <- read.csv("data/AmesHousing.csv")

dim(AmesHousing)  # 2930 rows and 82 columns, so very high dimensions make
the analysis a bit complex in nature

## [1] 2930   82

head(AmesHousing) # head not very useful for types; we'll rely on str() and
later cleaning

##   Order       PID MS.SubClass MS.Zoning Lot.Frontage Lot.Area Street Alley
## 1     1 526301100          20        RL          141    31770   Pave  <NA>
## 2     2 526350040          20        RH           80    11622   Pave  <NA>
## 3     3 526351010          20        RL           81    14267   Pave  <NA>
## 4     4 526353030          20        RL           93    11160   Pave  <NA>
## 5     5 527105010          60        RL           74    13830   Pave  <NA>
## 6     6 527105030          60        RL           78     9978   Pave  <NA>
##   Lot.Shape Land.Contour Utilities Lot.Config Land.Slope Neighborhood
## 1       IR1          Lvl    AllPub     Corner        Gtl        NAmes
## 2       Reg          Lvl    AllPub     Inside        Gtl        NAmes
## 3       IR1          Lvl    AllPub     Corner        Gtl        NAmes
## 4       Reg          Lvl    AllPub     Corner        Gtl        NAmes
## 5       IR1          Lvl    AllPub     Inside        Gtl      Gilbert
## 6       IR1          Lvl    AllPub     Inside        Gtl      Gilbert
##   Condition.1 Condition.2 Bldg.Type House.Style Overall.Qual Overall.Cond
## 1        Norm        Norm      1Fam      1Story            6            5
## 2       Feedr        Norm      1Fam      1Story            5            6
## 3        Norm        Norm      1Fam      1Story            6            6
## 4        Norm        Norm      1Fam      1Story            7            5
## 5        Norm        Norm      1Fam      2Story            5            5
## 6        Norm        Norm      1Fam      2Story            6            6
##   Year.Built Year.Remod.Add Roof.Style Roof.Matl Exterior.1st Exterior.2nd
## 1       1960           1960        Hip   CompShg      BrkFace      Plywood
## 2       1961           1961      Gable   CompShg      VinylSd      VinylSd
## 3       1958           1958        Hip   CompShg      Wd Sdng      Wd Sdng
## 4       1968           1968        Hip   CompShg      BrkFace      BrkFace
## 5       1997           1998      Gable   CompShg      VinylSd      VinylSd
## 6       1998           1998      Gable   CompShg      VinylSd      VinylSd
```

```
##   Mas.Vnr.Type Mas.Vnr.Area Exter.Qual Exter.Cond Foundation Bsmt.Qual
## 1        Stone          112         TA         TA     CBlock        TA
## 2         None            0         TA         TA     CBlock        TA
## 3      BrkFace          108         TA         TA     CBlock        TA
## 4         None            0         Gd         TA     CBlock        TA
## 5         None            0         TA         TA      PConc        Gd
## 6      BrkFace           20         TA         TA      PConc        TA
##   Bsmt.Cond Bsmt.Exposure BsmtFin.Type.1 BsmtFin.SF.1 BsmtFin.Type.2
## 1        Gd            Gd            BLQ          639            Unf
## 2        TA            No            Rec          468            LwQ
## 3        TA            No            ALQ          923            Unf
## 4        TA            No            ALQ         1065            Unf
## 5        TA            No            GLQ          791            Unf
## 6        TA            No            GLQ          602            Unf
##   BsmtFin.SF.2 Bsmt.Unf.SF Total.Bsmt.SF Heating Heating.QC Central.Air
## 1            0         441          1080    GasA         Fa           Y
## 2          144         270           882    GasA         TA           Y
## 3            0         406          1329    GasA         TA           Y
## 4            0        1045          2110    GasA         Ex           Y
## 5            0         137           928    GasA         Gd           Y
## 6            0         324           926    GasA         Ex           Y
##   Electrical X1st.Flr.SF X2nd.Flr.SF Low.Qual.Fin.SF Gr.Liv.Area
Bsmt.Full.Bath
## 1      SBrkr        1656           0               0        1656
1
## 2      SBrkr         896           0               0         896
0
## 3      SBrkr        1329           0               0        1329
0
## 4      SBrkr        2110           0               0        2110
1
## 5      SBrkr         928         701               0        1629
0
## 6      SBrkr         926         678               0        1604
0
##   Bsmt.Half.Bath Full.Bath Half.Bath Bedroom.AbvGr Kitchen.AbvGr
Kitchen.Qual
## 1              0         1         0             3             1
TA
## 2              0         1         0             2             1
TA
## 3              0         1         1             3             1
Gd
## 4              0         2         1             3             1
Ex
## 5              0         2         1             3             1
TA
## 6              0         2         1             3             1
Gd
##   TotRms.AbvGrd Functional Fireplaces Fireplace.Qu Garage.Type
```

```
Garage.Yr.Blt
## 1            7        Typ          2           Gd      Attchd
1960
## 2            5        Typ          0         <NA>      Attchd
1961
## 3            6        Typ          0         <NA>      Attchd
1958
## 4            8        Typ          2           TA      Attchd
1968
## 5            6        Typ          1           TA      Attchd
1997
## 6            7        Typ          1           Gd      Attchd
1998
##   Garage.Finish Garage.Cars Garage.Area Garage.Qual Garage.Cond
Paved.Drive
## 1           Fin           2         528          TA          TA
P
## 2           Unf           1         730          TA          TA
Y
## 3           Unf           1         312          TA          TA
Y
## 4           Fin           2         522          TA          TA
Y
## 5           Fin           2         482          TA          TA
Y
## 6           Fin           2         470          TA          TA
Y
##   Wood.Deck.SF Open.Porch.SF Enclosed.Porch X3Ssn.Porch Screen.Porch
Pool.Area
## 1          210            62              0           0            0
0
## 2          140             0              0           0          120
0
## 3          393            36              0           0            0
0
## 4            0             0              0           0            0
0
## 5          212            34              0           0            0
0
## 6          360            36              0           0            0
0
##   Pool.QC Fence Misc.Feature Misc.Val Mo.Sold Yr.Sold Sale.Type
Sale.Condition
## 1    <NA>  <NA>         <NA>        0       5    2010       WD
Normal
## 2    <NA> MnPrv         <NA>        0       6    2010       WD
Normal
## 3    <NA>  <NA>         Gar2    12500       6    2010       WD
Normal
## 4    <NA>  <NA>         <NA>        0       4    2010       WD
```

```
Normal
## 5    <NA> MnPrv         <NA>        0       3    2010       WD
Normal
## 6    <NA>  <NA>         <NA>        0       6    2010       WD
Normal
##   SalePrice
## 1    215000
## 2    105000
## 3    172000
## 4    244000
## 5    189900
## 6    195500
```

```
str(AmesHousing)  # SalePrice is int; OK (R will treat as numeric in
modeling)
```

```
## 'data.frame':    2930 obs. of  82 variables:
##  $ Order          : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ PID            : int  526301100 526350040 526351010 526353030 527105010
527105030 527127150 527145080 527146030 527162130 ...
##  $ MS.SubClass    : int  20 20 20 20 60 60 120 120 120 60 ...
##  $ MS.Zoning      : chr  "RL" "RH" "RL" "RL" ...
##  $ Lot.Frontage   : int  141 80 81 93 74 78 41 43 39 60 ...
##  $ Lot.Area       : int  31770 11622 14267 11160 13830 9978 4920 5005 5389
7500 ...
##  $ Street         : chr  "Pave" "Pave" "Pave" "Pave" ...
##  $ Alley          : chr  NA NA NA NA ...
##  $ Lot.Shape      : chr  "IR1" "Reg" "IR1" "Reg" ...
##  $ Land.Contour   : chr  "Lvl" "Lvl" "Lvl" "Lvl" ...
##  $ Utilities      : chr  "AllPub" "AllPub" "AllPub" "AllPub" ...
##  $ Lot.Config     : chr  "Corner" "Inside" "Corner" "Corner" ...
##  $ Land.Slope     : chr  "Gtl" "Gtl" "Gtl" "Gtl" ...
##  $ Neighborhood   : chr  "NAmes" "NAmes" "NAmes" "NAmes" ...
##  $ Condition.1    : chr  "Norm" "Feedr" "Norm" "Norm" ...
##  $ Condition.2    : chr  "Norm" "Norm" "Norm" "Norm" ...
##  $ Bldg.Type      : chr  "1Fam" "1Fam" "1Fam" "1Fam" ...
##  $ House.Style    : chr  "1Story" "1Story" "1Story" "1Story" ...
##  $ Overall.Qual   : int  6 5 6 7 5 6 8 8 8 7 ...
##  $ Overall.Cond   : int  5 6 6 5 5 6 5 5 5 5 ...
##  $ Year.Built     : int  1960 1961 1958 1968 1997 1998 2001 1992 1995 1999
...
##  $ Year.Remod.Add : int  1960 1961 1958 1968 1998 1998 2001 1992 1996 1999
...
##  $ Roof.Style     : chr  "Hip" "Gable" "Hip" "Hip" ...
##  $ Roof.Matl      : chr  "CompShg" "CompShg" "CompShg" "CompShg" ...
##  $ Exterior.1st   : chr  "BrkFace" "VinylSd" "Wd Sdng" "BrkFace" ...
##  $ Exterior.2nd   : chr  "Plywood" "VinylSd" "Wd Sdng" "BrkFace" ...
##  $ Mas.Vnr.Type   : chr  "Stone" "None" "BrkFace" "None" ...
##  $ Mas.Vnr.Area   : int  112 0 108 0 0 20 0 0 0 0 ...
##  $ Exter.Qual     : chr  "TA" "TA" "TA" "Gd" ...
```

```
##  $ Exter.Cond     : chr  "TA" "TA" "TA" "TA" ...
##  $ Foundation     : chr  "CBlock" "CBlock" "CBlock" "CBlock" ...
##  $ Bsmt.Qual      : chr  "TA" "TA" "TA" "TA" ...
##  $ Bsmt.Cond      : chr  "Gd" "TA" "TA" "TA" ...
##  $ Bsmt.Exposure  : chr  "Gd" "No" "No" "No" ...
##  $ BsmtFin.Type.1 : chr  "BLQ" "Rec" "ALQ" "ALQ" ...
##  $ BsmtFin.SF.1   : int  639 468 923 1065 791 602 616 263 1180 0 ...
##  $ BsmtFin.Type.2 : chr  "Unf" "LwQ" "Unf" "Unf" ...
##  $ BsmtFin.SF.2   : int  0 144 0 0 0 0 0 0 0 0 ...
##  $ Bsmt.Unf.SF    : int  441 270 406 1045 137 324 722 1017 415 994 ...
##  $ Total.Bsmt.SF  : int  1080 882 1329 2110 928 926 1338 1280 1595 994 ...
##  $ Heating        : chr  "GasA" "GasA" "GasA" "GasA" ...
##  $ Heating.QC     : chr  "Fa" "TA" "TA" "Ex" ...
##  $ Central.Air    : chr  "Y" "Y" "Y" "Y" ...
##  $ Electrical     : chr  "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...
##  $ X1st.Flr.SF    : int  1656 896 1329 2110 928 926 1338 1280 1616 1028
...
##  $ X2nd.Flr.SF    : int  0 0 0 0 701 678 0 0 0 776 ...
##  $ Low.Qual.Fin.SF: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Gr.Liv.Area    : int  1656 896 1329 2110 1629 1604 1338 1280 1616 1804
...
##  $ Bsmt.Full.Bath : int  1 0 0 1 0 0 1 0 1 0 ...
##  $ Bsmt.Half.Bath : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Full.Bath      : int  1 1 1 2 2 2 2 2 2 2 ...
##  $ Half.Bath      : int  0 0 1 1 1 1 0 0 0 1 ...
##  $ Bedroom.AbvGr  : int  3 2 3 3 3 3 2 2 2 3 ...
##  $ Kitchen.AbvGr  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Kitchen.Qual   : chr  "TA" "TA" "Gd" "Ex" ...
##  $ TotRms.AbvGrd  : int  7 5 6 8 6 7 6 5 5 7 ...
##  $ Functional     : chr  "Typ" "Typ" "Typ" "Typ" ...
##  $ Fireplaces     : int  2 0 0 2 1 1 0 0 1 1 ...
##  $ Fireplace.Qu   : chr  "Gd" NA NA "TA" ...
##  $ Garage.Type    : chr  "Attchd" "Attchd" "Attchd" "Attchd" ...
##  $ Garage.Yr.Blt  : int  1960 1961 1958 1968 1997 1998 2001 1992 1995 1999
...
##  $ Garage.Finish  : chr  "Fin" "Unf" "Unf" "Fin" ...
##  $ Garage.Cars    : int  2 1 1 2 2 2 2 2 2 2 ...
##  $ Garage.Area    : int  528 730 312 522 482 470 582 506 608 442 ...
##  $ Garage.Qual    : chr  "TA" "TA" "TA" "TA" ...
##  $ Garage.Cond    : chr  "TA" "TA" "TA" "TA" ...
##  $ Paved.Drive    : chr  "P" "Y" "Y" "Y" ...
##  $ Wood.Deck.SF   : int  210 140 393 0 212 360 0 0 237 140 ...
##  $ Open.Porch.SF  : int  62 0 36 0 34 36 0 82 152 60 ...
##  $ Enclosed.Porch : int  0 0 0 0 0 0 170 0 0 0 ...
##  $ X3Ssn.Porch    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Screen.Porch   : int  0 120 0 0 0 0 0 144 0 0 ...
##  $ Pool.Area      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Pool.QC        : chr  NA NA NA NA ...
##  $ Fence          : chr  NA "MnPrv" NA NA ...
##  $ Misc.Feature   : chr  NA NA "Gar2" NA ...
```

```
##  $ Misc.Val       : int  0 0 12500 0 0 0 0 0 0 0 ...
##  $ Mo.Sold        : int  5 6 6 4 3 6 4 1 3 6 ...
##  $ Yr.Sold        : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010
...
##  $ Sale.Type      : chr  "WD " "WD " "WD " "WD " ...
##  $ Sale.Condition : chr  "Normal" "Normal" "Normal" "Normal" ...
##  $ SalePrice      : int  215000 105000 172000 244000 189900 195500 213500
191500 236500 189000 ...
```

Notice our hypothesized primary and secondary sources of variation based on theory Primary : Neighborhood, Overall Quality, Above-ground living area Secondary: Basement Area, Garage Space, Architectural Styl

Note Overall_Qual is actually Ordinal so we elect to change the data type Note Overall_Cond is actually Ordinal so we elect to change the data type Note Exter_Cond is actually Ordinal so we elect to change the data type and can also potentially reinput it as a source of variation Note Bsmt_Cond is actually Ordinal so we elect to change the data type and can also potentially reinput it as a source of variation we may elect to create a new dataframe with all these to make our analysis smoother, not decided

```r
# Quick numeric sanity check on the response
summary(AmesHousing$SalePrice)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   12789  129500  160000  180796  213500  755000
```

Theory-driven sources of variation

```r
# Primary sources of variation
primary_vars <- c("Neighborhood", "Overall.Qual", "Gr.Liv.Area")

# Secondary sources of variation
secondary_vars <- c("Total.Bsmt.SF", "Garage.Cars", "House.Style")

# Ordinal variables flagged for later recoding
ordinal_vars <- c("Overall.Qual", "Overall.Cond", "Exter.Cond", "Bsmt.Cond")
```

Factor level counts

```r
factor_levels <- c()

for (var_name in names(AmesHousing)) {
  x <- AmesHousing[[var_name]]

  # count levels for factor OR character
  if (is.factor(x)) {
    factor_levels[var_name] <- nlevels(x)
  } else if (is.character(x)) {
    factor_levels[var_name] <- length(unique(x))
  }
```

```
}

factor_levels <- sort(factor_levels, decreasing = TRUE)
factor_levels

##    Neighborhood    Exterior.2nd    Exterior.1st       Sale.Type     Condition.1
##              28              17              16              10               9
##     Condition.2     House.Style       Roof.Matl BsmtFin.Type.1 BsmtFin.Type.2
##               8               8               8               8               8
##      Functional       MS.Zoning       Bsmt.Qual       Bsmt.Cond     Garage.Type
##               8               7               7               7               7
##     Garage.Qual     Garage.Cond      Roof.Style    Mas.Vnr.Type      Foundation
##               7               7               6               6               6
##   Bsmt.Exposure         Heating      Electrical    Fireplace.Qu    Misc.Feature
##               6               6               6               6               6
##  Sale.Condition      Lot.Config       Bldg.Type      Exter.Cond      Heating.QC
##               6               5               5               5               5
##    Kitchen.Qual   Garage.Finish         Pool.QC           Fence       Lot.Shape
##               5               5               5               5               4
##    Land.Contour      Exter.Qual           Alley       Utilities      Land.Slope
##               4               4               3               3               3
##     Paved.Drive          Street     Central.Air
##               3               2               2

# Checks
length(AmesHousing$Neighborhood)

## [1] 2930

length(unique(AmesHousing$Neighborhood))

## [1] 28

head(factor_levels)

## Neighborhood Exterior.2nd Exterior.1st    Sale.Type  Condition.1
Condition.2
##           28           17           16           10            9
8

which.max(factor_levels) # Neighborhood

## Neighborhood
##            1
```

Neighborhood has substantially higher cardinality than other categorical variables, confirming it as a major structural source of variation and motivating careful treatment in regression modeling.

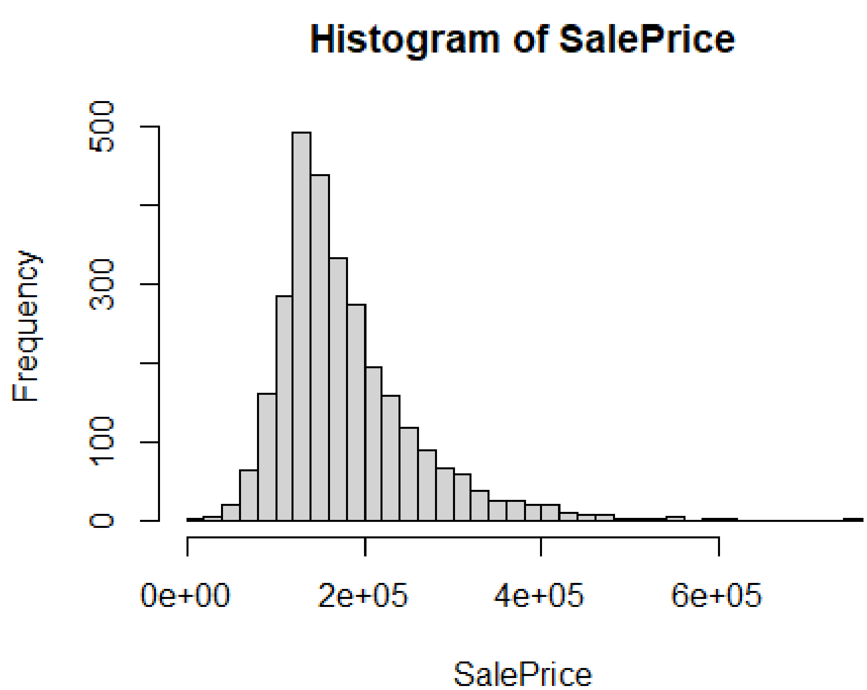
Ordinal recoding (quality variables)

1–10 scales: keep as numeric (already ordinal) Overall.Qual, Overall.Cond → DO NOTHING (correct as-is) Quality-grade variables: recode to ordered factors Ames quality order (worst → best)

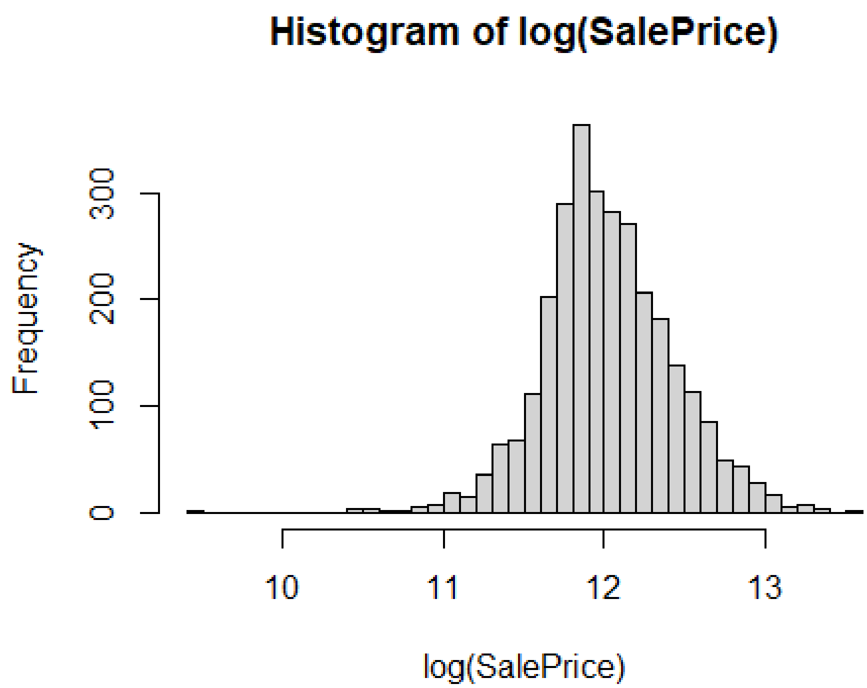
```
qual_levels <- c("Po", "Fa", "TA", "Gd", "Ex")

AmesHousing$Exter.Cond <- factor(
AmesHousing$Exter.Cond,
levels = qual_levels,
ordered = TRUE
)

AmesHousing$Bsmt.Cond <- factor(
AmesHousing$Bsmt.Cond,
levels = qual_levels,
ordered = TRUE
)
```

SalePrice distribution checks and transformation

```
hist(AmesHousing$SalePrice,
     breaks = 50,
     main = "Histogram of SalePrice",
     xlab = "SalePrice")                          # clearly this is skewed
```



```r
# Log-scale version
hist(log(AmesHousing$SalePrice),
     breaks = 50,
     main = "Histogram of log(SalePrice)",
     xlab = "log(SalePrice)")                    # much better for normal
approximations
```



Justification for log(SalePrice) transformation

```r
# Coefficient of variation comparison
sd(AmesHousing$SalePrice) / mean(AmesHousing$SalePrice)

## [1] 0.4418608

sd(log(AmesHousing$SalePrice)) / mean(log(AmesHousing$SalePrice))

## [1] 0.03390633
```

The raw SalePrice distribution is strongly right-skewed, with increasing variance at higher price levels, violating the constant variance assumption of OLS regression. A logarithmic transformation reduces skewness and stabilizes variance, yielding a distribution more appropriate for linear modeling. Accordingly, all subsequent analyses use log(SalePrice) as the response variable.

Preparing data for Analysis procedures

```r
AmesHousing <- AmesHousing |> mutate(logSalePrice = log10(SalePrice))
```

Response missingness and structural missingness handling

Observations with missing SalePrice cannot be used in supervised modeling (OLS, ANOVA, logistic regression). Such rows are removed if present.

```
AmesHousing <- AmesHousing |>
  filter(!is.na(SalePrice))

sum(is.na(AmesHousing$SalePrice))  # should be 0

## [1] 0
```

Handle structural missingness (do NOT drop observations)

For Ames housing data, missing basement or garage values indicate the absence of that feature, not missing data. Replacing NA with 0 preserves information and avoids biased row deletion in downstream OLS analysis.

```
AmesHousing <- AmesHousing |>
mutate(
Total.Bsmt.SF = ifelse(is.na(Total.Bsmt.SF), 0, Total.Bsmt.SF),
Garage.Cars = ifelse(is.na(Garage.Cars), 0, Garage.Cars)
)

sapply(
AmesHousing[c("Total.Bsmt.SF", "Garage.Cars")],
function(x) sum(is.na(x))
)

## Total.Bsmt.SF   Garage.Cars
##             0             0
```

Module 2: OLS Regression — Primary + Secondary Predictors

Model specification

```
AmesHousing$Neighborhood <- factor(AmesHousing$Neighborhood)

ols_model <- lm(
  logSalePrice ~ Neighborhood +
    Overall.Qual +
    Gr.Liv.Area +
    Total.Bsmt.SF +
    Garage.Cars +
    House.Style,
  data = AmesHousing
)

#Model summary
summary(ols_model)
```

```
##
## Call:
## lm(formula = logSalePrice ~ Neighborhood + Overall.Qual + Gr.Liv.Area +
##     Total.Bsmt.SF + Garage.Cars + House.Style, data = AmesHousing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82872 -0.03055  0.00489  0.03882  0.30770
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          4.679e+00  1.768e-02 264.634  < 2e-16 ***
## NeighborhoodBlueste -4.790e-02  2.596e-02  -1.845 0.065077 .
## NeighborhoodBrDale  -9.752e-02  1.905e-02  -5.119 3.27e-07 ***
## NeighborhoodBrkSide -3.631e-02  1.553e-02  -2.338 0.019469 *
## NeighborhoodClearCr  5.899e-02  1.717e-02   3.435 0.000601 ***
## NeighborhoodCollgCr  2.940e-02  1.399e-02   2.102 0.035656 *
## NeighborhoodCrawfor  5.230e-02  1.519e-02   3.442 0.000585 ***
## NeighborhoodEdwards -4.034e-02  1.468e-02  -2.748 0.006036 **
## NeighborhoodGilbert  2.337e-02  1.460e-02   1.601 0.109418
## NeighborhoodGreens  -1.725e-03  2.811e-02  -0.061 0.951077
## NeighborhoodGrnHill  2.232e-01  5.133e-02   4.348 1.42e-05 ***
## NeighborhoodIDOTRR  -1.036e-01  1.579e-02  -6.559 6.41e-11 ***
## NeighborhoodLandmrk -4.287e-02  7.127e-02  -0.601 0.547577
## NeighborhoodMeadowV -7.948e-02  1.843e-02  -4.313 1.66e-05 ***
## NeighborhoodMitchel  1.003e-02  1.505e-02   0.667 0.505008
## NeighborhoodNAmes   -5.017e-03  1.396e-02  -0.359 0.719341
## NeighborhoodNoRidge  5.739e-02  1.602e-02   3.582 0.000346 ***
## NeighborhoodNPkVill -4.866e-02  1.985e-02  -2.452 0.014276 *
## NeighborhoodNridgHt  6.804e-02  1.444e-02   4.714 2.55e-06 ***
## NeighborhoodNWAmes   1.103e-03  1.474e-02   0.075 0.940381
## NeighborhoodOldTown -7.542e-02  1.456e-02  -5.179 2.39e-07 ***
## NeighborhoodSawyer  -2.650e-03  1.478e-02  -0.179 0.857695
## NeighborhoodSawyerW  4.965e-03  1.481e-02   0.335 0.737457
## NeighborhoodSomerst  3.787e-02  1.430e-02   2.648 0.008149 **
## NeighborhoodStoneBr  6.614e-02  1.658e-02   3.989 6.80e-05 ***
## NeighborhoodSWISU   -4.803e-02  1.743e-02  -2.755 0.005906 **
## NeighborhoodTimber   4.898e-02  1.563e-02   3.135 0.001737 **
## NeighborhoodVeenker  4.406e-02  1.956e-02   2.253 0.024322 *
## Overall.Qual         4.677e-02  1.612e-03  29.020  < 2e-16 ***
## Gr.Liv.Area          1.014e-04  4.809e-06  21.080  < 2e-16 ***
## Total.Bsmt.SF        4.897e-05  4.895e-06  10.003  < 2e-16 ***
## Garage.Cars          2.853e-02  2.390e-03  11.939  < 2e-16 ***
## House.Style1.5Unf   -2.631e-02  1.688e-02  -1.559 0.119121
## House.Style1Story    3.656e-03  5.405e-03   0.677 0.498771
## House.Style2.5Fin   -1.390e-02  2.615e-02  -0.532 0.594903
## House.Style2.5Unf   -1.749e-02  1.512e-02  -1.157 0.247344
## House.Style2Story   -1.602e-03  5.469e-03  -0.293 0.769650
## House.StyleSFoyer    3.391e-02  9.499e-03   3.570 0.000362 ***
## House.StyleSLvl      1.645e-02  7.906e-03   2.080 0.037568 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06991 on 2891 degrees of freedom
## Multiple R-squared:  0.846,  Adjusted R-squared:  0.844
## F-statistic: 418.1 on 38 and 2891 DF,  p-value: < 2.2e-16
```

The OLS model explains a large proportion of variability in logSalePrice (Adjusted $R^2 \approx$ 0.84). Key structural predictors such as Overall.Qual and Gr.Liv.Area are highly statistically significant, supporting the use of this model as a baseline for defining relative over- and under-performance.

Fitted values and residuals

```
AmesHousing$fitted_logPrice <- fitted(ols_model)
AmesHousing$residuals_log   <- resid(ols_model)
```

The OLS results indicate that logSalePrice is strongly explained by structural characteristics and location. Overall.Qual and Gr.Liv.Area exhibit the largest t-statistics, confirming them as dominant drivers of housing prices.

Neighborhood effects remain significant even after controlling for quality and size, indicating location-specific price premia beyond physical attributes.

Several neighborhood indicators are statistically significant, with both positive and negative effects relative to the reference category (Blmngtn). This suggests meaningful spatial heterogeneity in housing prices that is not fully explained by observable structural features.

Most House.Style categories are not statistically significant once size, quality, and neighborhood are accounted for, indicating that architectural style contributes limited additional explanatory power beyond core structural characteristics.

Reference neighborhood

```
levels(AmesHousing$Neighborhood)

##  [1] "Blmngtn" "Blueste" "BrDale"  "BrkSide" "ClearCr" "CollgCr" "Crawfor"
##  [8] "Edwards" "Gilbert" "Greens"  "GrnHill" "IDOTRR"  "Landmrk" "MeadowV"
## [15] "Mitchel" "NAmes"   "NoRidge" "NPkVill" "NridgHt" "NWAmes"  "OldTown"
## [22] "Sawyer"  "SawyerW" "Somerst" "StoneBr" "SWISU"   "Timber"  "Veenker"
```

Blmngtn is the reference neighborhood in the OLS model, and all neighborhood coefficients are interpreted relative to this baseline.

```
# Multicollinearity check (VIF)
vif(ols_model)

##                  GVIF Df GVIF^(1/(2*Df))
## Neighborhood 6.001298 27        1.033741
## Overall.Qual 3.098740  1        1.760324
## Gr.Liv.Area  3.542154  1        1.882061
```

```
## Total.Bsmt.SF 2.792847  1        1.671181
## Garage.Cars   1.983125  1        1.408235
## House.Style   4.989281  7        1.121656
```

The Variance Inflation Factors (GVIF-adjusted for terms with multiple degrees of freedom) all come in below standard cutoffs, so there's no sign of serious multicollinearity issues between the predictors. This means the coefficient estimates from the OLS model should be stable and we can interpret them reliably.

Module 2B: Interaction Check (Neighborhood × Overall.Qual)

```
ols_interaction <- lm(
  logSalePrice ~ Neighborhood * Overall.Qual +
    Gr.Liv.Area +
    Total.Bsmt.SF +
    Garage.Cars +
    House.Style,
  data = AmesHousing
)

summary(ols_interaction)

##
## Call:
## lm(formula = logSalePrice ~ Neighborhood * Overall.Qual + Gr.Liv.Area +
##     Total.Bsmt.SF + Garage.Cars + House.Style, data = AmesHousing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74165 -0.03022  0.00376  0.03693  0.30249
##
## Coefficients: (3 not defined because of singularities)
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        4.635e+00  2.633e-01  17.601  < 2e-16
***
## NeighborhoodBlueste               -1.600e-01  3.067e-01  -0.522  0.60185
## NeighborhoodBrDale                 2.684e-01  3.091e-01   0.868  0.38526
## NeighborhoodBrkSide               -6.465e-02  2.652e-01  -0.244  0.80746
## NeighborhoodClearCr                3.649e-02  2.711e-01   0.135  0.89291
## NeighborhoodCollgCr                7.451e-02  2.646e-01   0.282  0.77826
## NeighborhoodCrawfor                3.037e-02  2.663e-01   0.114  0.90920
## NeighborhoodEdwards                1.760e-01  2.642e-01   0.666  0.50521
## NeighborhoodGilbert                1.752e-01  2.677e-01   0.654  0.51289
## NeighborhoodGreens                -4.874e-03  4.176e-02  -0.117  0.90710
## NeighborhoodGrnHill                2.256e-01  5.032e-02   4.484 7.63e-06
***
## NeighborhoodIDOTRR                -2.490e-01  2.648e-01  -0.940  0.34714
## NeighborhoodLandmrk               -3.533e-02  8.121e-02  -0.435  0.66353
## NeighborhoodMeadowV               -1.175e-01  2.766e-01  -0.425  0.67118
## NeighborhoodMitchel                6.952e-02  2.655e-01   0.262  0.79347
```

```
## NeighborhoodNAmes                  7.614e-02  2.642e-01   0.288  0.77322    
## NeighborhoodNoRidge                1.607e-01  2.762e-01   0.582  0.56087    
## NeighborhoodNPkVill                1.810e-01  3.332e-01   0.543  0.58694    
## NeighborhoodNridgHt               -5.761e-03  2.672e-01  -0.022  0.98280    
## NeighborhoodNWAmes                 1.668e-01  2.688e-01   0.620  0.53500    
## NeighborhoodOldTown               -7.912e-02  2.639e-01  -0.300  0.76437    
## NeighborhoodSawyer                 4.050e-02  2.675e-01   0.151  0.87967    
## NeighborhoodSawyerW               -4.257e-02  2.674e-01  -0.159  0.87351    
## NeighborhoodSomerst                1.140e-01  2.669e-01   0.427  0.66938    
## NeighborhoodStoneBr               -2.153e-01  2.864e-01  -0.752  0.45220    
## NeighborhoodSWISU                  9.919e-03  2.693e-01   0.037  0.97063    
## NeighborhoodTimber                 6.731e-02  2.679e-01   0.251  0.80160    
## NeighborhoodVeenker                2.396e-01  2.785e-01   0.860  0.38973    
## Overall.Qual                       5.173e-02  3.681e-02   1.405  0.16008    
## Gr.Liv.Area                        1.032e-04  4.779e-06  21.592  < 2e-16 
***
## Total.Bsmt.SF                      5.335e-05  4.910e-06  10.865  < 2e-16 
***
## Garage.Cars                        2.801e-02  2.354e-03  11.901  < 2e-16 
***
## House.Style1.5Unf                 -2.730e-02  1.650e-02  -1.655  0.09812 .  
## House.Style1Story                  5.374e-03  5.343e-03   1.006  0.31461    
## House.Style2.5Fin                 -2.294e-02  2.574e-02  -0.891  0.37297    
## House.Style2.5Unf                 -2.625e-02  1.503e-02  -1.747  0.08072 .  
## House.Style2Story                  1.204e-03  5.392e-03   0.223  0.82328    
## House.StyleSFoyer                  4.232e-02  9.338e-03   4.532 6.09e-06 
***
## House.StyleSLvl                    2.304e-02  7.788e-03   2.958  0.00312 ** 
## NeighborhoodBlueste:Overall.Qual  1.772e-02  4.372e-02   0.405  0.68528    
## NeighborhoodBrDale:Overall.Qual  -6.224e-02  4.637e-02  -1.342  0.17958    
## NeighborhoodBrkSide:Overall.Qual  8.143e-03  3.734e-02   0.218  0.82737    
## NeighborhoodClearCr:Overall.Qual  4.590e-03  3.832e-02   0.120  0.90467    
## NeighborhoodCollgCr:Overall.Qual -6.452e-03  3.702e-02  -0.174  0.86165    
## NeighborhoodCrawfor:Overall.Qual  4.297e-03  3.734e-02   0.115  0.90841    
## NeighborhoodEdwards:Overall.Qual -4.059e-02  3.706e-02  -1.095  0.27343    
## NeighborhoodGilbert:Overall.Qual -2.282e-02  3.755e-02  -0.608  0.54345    
## NeighborhoodGreens:Overall.Qual          NA         NA      NA       NA    
## NeighborhoodGrnHill:Overall.Qual         NA         NA      NA       NA    
## NeighborhoodIDOTRR:Overall.Qual   3.387e-02  3.727e-02   0.909  0.36354    
## NeighborhoodLandmrk:Overall.Qual         NA         NA      NA       NA    
## NeighborhoodMeadowV:Overall.Qual  1.189e-02  4.151e-02   0.286  0.77457    
## NeighborhoodMitchel:Overall.Qual -9.417e-03  3.732e-02  -0.252  0.80081    
## NeighborhoodNAmes:Overall.Qual   -1.344e-02  3.705e-02  -0.363  0.71686    
## NeighborhoodNoRidge:Overall.Qual -1.390e-02  3.829e-02  -0.363  0.71670    
## NeighborhoodNPkVill:Overall.Qual -3.580e-02  4.912e-02  -0.729  0.46610    
## NeighborhoodNridgHt:Overall.Qual  7.932e-03  3.723e-02   0.213  0.83128    
## NeighborhoodNWAmes:Overall.Qual  -2.567e-02  3.780e-02  -0.679  0.49708    
## NeighborhoodOldTown:Overall.Qual  3.039e-03  3.698e-02   0.082  0.93451    
## NeighborhoodSawyer:Overall.Qual  -6.372e-03  3.798e-02  -0.168  0.86678    
## NeighborhoodSawyerW:Overall.Qual  8.127e-03  3.754e-02   0.217  0.82860    
```

```
## NeighborhoodSomerst:Overall.Qual -1.051e-02  3.727e-02  -0.282  0.77795
## NeighborhoodStoneBr:Overall.Qual  3.260e-02  3.918e-02   0.832  0.40542
## NeighborhoodSWISU:Overall.Qual   -8.644e-03  3.822e-02  -0.226  0.82110
## NeighborhoodTimber:Overall.Qual  -2.781e-03  3.743e-02  -0.074  0.94079
## NeighborhoodVeenker:Overall.Qual -2.750e-02  3.886e-02  -0.708  0.47931
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06815 on 2867 degrees of freedom
## Multiple R-squared:  0.8549, Adjusted R-squared:  0.8518
## F-statistic: 272.4 on 62 and 2867 DF,  p-value: < 2.2e-16
```

Baseline OLS model retained for defining expected price. Interaction terms were explored but not adopted due to limited interpretability and model stability concerns.

Module 2C: ANOVA (Type II) for OLS Model

```
Anova(ols_model, type = 2)
```

```
## Anova Table (Type II tests)
##
## Response: logSalePrice
##                Sum Sq   Df  F value    Pr(>F)
## Neighborhood   3.3584   27  25.4507 < 2.2e-16 ***
## Overall.Qual   4.1159    1 842.1629 < 2.2e-16 ***
## Gr.Liv.Area    2.1718    1 444.3864 < 2.2e-16 ***
## Total.Bsmt.SF  0.4891    1 100.0667 < 2.2e-16 ***
## Garage.Cars    0.6966    1 142.5382 < 2.2e-16 ***
## House.Style    0.1200    7   3.5086 0.0009386 ***
## Residuals     14.1290 2891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA (Type II) Interpretation

The Type II ANOVA assesses the marginal contribution of each predictor after accounting for all other variables in the model.

Neighborhood is highly statistically significant ($p < 2e-16$), confirming strong location-based differences in housing prices.

Overall.Qual has the largest F-statistic, indicating it is the single most important predictor of log(SalePrice).

Gr.Liv.Area, Total.Bsmt.SF, and Garage.Cars are all highly significant, supporting the role of size and functional space in determining prices.

House.Style is statistically significant as a group ($p < 0.001$), although individual style coefficients may be weak, indicating that architectural style contributes modest but non-negligible variation when cons

Module 2D: One-Way ANOVA (Neighborhood Only)

```
anova_neighborhood <- aov(
  logSalePrice ~ Neighborhood,
  data = AmesHousing
)

summary(anova_neighborhood)

##                Df Sum Sq Mean Sq F value Pr(>F)
## Neighborhood   27  53.46  1.9798   149.9 <2e-16 ***
## Residuals    2902  38.32  0.0132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

One-Way ANOVA Interpretation (Neighborhood Only)

The one-way ANOVA tests whether mean log(SalePrice) differs across neighborhoods without controlling for other structural variables.

The F-statistic is very large ($F \approx 150$) with $p < 2e-16$, providing overwhelming evidence that average housing prices differ significantly across neighborhoods.

This confirms Neighborhood as a dominant source of variation in housing prices and motivates its inclusion in the multivariable OLS model.

Module 2E: Post-hoc Comparisons (Tukey HSD)

```
tukey_neighborhood <- TukeyHSD(anova_neighborhood)
tukey_neighborhood

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = logSalePrice ~ Neighborhood, data = AmesHousing)
##
## $Neighborhood
##                         diff          lwr           upr     p adj
## Blueste-Blmngtn -0.140127078 -0.2975316466  1.727749e-02 0.1715389
## BrDale-Blmngtn  -0.268390632 -0.3806645562 -1.561167e-01 0.0000000
## BrkSide-Blmngtn -0.211962270 -0.3025737262 -1.213508e-01 0.0000000
## ClearCr-Blmngtn  0.016966034 -0.0863256707  1.202577e-01 1.0000000
## CollgCr-Blmngtn  0.001167390 -0.0837078297  8.604261e-02 1.0000000
## Crawfor-Blmngtn  0.007412722 -0.0836504070  9.847585e-02 1.0000000
## Edwards-Blmngtn -0.196532692 -0.2829102963 -1.101551e-01 0.0000000
## Gilbert-Blmngtn -0.014498654 -0.1018284037  7.283110e-02 1.0000000
## Greens-Blmngtn  -0.005083857 -0.1763737706  1.662061e-01 1.0000000
## GrnHill-Blmngtn  0.150888337 -0.1618428293  4.636195e-01 0.9931965
## IDOTRR-Blmngtn  -0.306115909 -0.3982195388 -2.140123e-01 0.0000000
## Landmrk-Blmngtn -0.152511984 -0.5873470243  2.823231e-01 0.9999717
## MeadowV-Blmngtn -0.316566729 -0.4235907716 -2.095427e-01 0.0000000
```

```
## Mitchel-Blmngtn -0.092293750 -0.1824129405 -2.174560e-03 0.0367354
## NAmes-Blmngtn   -0.136886232 -0.2201457942 -5.362667e-02 0.0000004
## NoRidge-Blmngtn  0.214758424  0.1194099038  3.101069e-01 0.0000000
## NPkVill-Blmngtn -0.141858535 -0.2620979086 -2.161916e-02 0.0037950
## NridgHt-Blmngtn  0.199826403  0.1125348227  2.871180e-01 0.0000000
## NWAmes-Blmngtn  -0.022527954 -0.1114866588  6.643075e-02 1.0000000
## OldTown-Blmngtn -0.219892541 -0.3052383527 -1.345467e-01 0.0000000
## Sawyer-Blmngtn  -0.160044899 -0.2479600193 -7.212978e-02 0.0000000
## SawyerW-Blmngtn -0.040168379 -0.1295022683  4.916551e-02 0.9976499
## Somerst-Blmngtn  0.059330902 -0.0274051180  1.460669e-01 0.7179487
## StoneBr-Blmngtn  0.192290228  0.0917930213  2.927874e-01 0.0000000
## SWISU-Blmngtn   -0.170609027 -0.2722132389 -6.900481e-02 0.0000002
## Timber-Blmngtn   0.086611952 -0.0085491117  1.817730e-01 0.1401778
## Veenker-Blmngtn  0.091172655 -0.0276834751  2.100288e-01 0.4690941
## BrDale-Blueste  -0.128263553 -0.2842812703  2.775416e-02 0.3156909
## BrkSide-Blueste -0.071835192 -0.2130673759  6.939699e-02 0.9858501
## ClearCr-Blueste  0.157093112  0.0074092116  3.067770e-01 0.0262132
## CollgCr-Blueste  0.141294468  0.0036721708  2.789168e-01 0.0355304
## Crawfor-Blueste  0.147539800  0.0060174087  2.890622e-01 0.0288022
## Edwards-Blueste -0.056405613 -0.1949595178  8.214829e-02 0.9995490
## Gilbert-Blueste  0.125628425 -0.0135210601  2.647779e-01 0.1511659
## Greens-Blueste   0.135043222 -0.0676297377  3.377162e-01 0.7656704
## GrnHill-Blueste  0.291015415 -0.0399481420  6.219790e-01 0.1908025
## IDOTRR-Blueste  -0.165988831 -0.3081829674 -2.379469e-02 0.0046278
## Landmrk-Blueste -0.012384906 -0.4605116804  4.357419e-01 1.0000000
## MeadowV-Blueste -0.176439651 -0.3287230693 -2.415623e-02 0.0053078
## Mitchel-Blueste  0.047833328 -0.0930835351  1.887502e-01 0.9999852
## NAmes-Blueste    0.003240847 -0.1333909528  1.398726e-01 1.0000000
## NoRidge-Blueste  0.354885502  0.2105683717  4.992026e-01 0.0000000
## NPkVill-Blueste -0.001731457 -0.1635758290  1.601129e-01 1.0000000
## NridgHt-Blueste  0.339953481  0.2008279485  4.790790e-01 0.0000000
## NWAmes-Blueste   0.117599124 -0.0225784240  2.577767e-01 0.2748186
## OldTown-Blueste -0.079765462 -0.2176784839  5.814756e-02 0.9340996
## Sawyer-Blueste  -0.019917820 -0.1594354261  1.195998e-01 1.0000000
## SawyerW-Blueste  0.099958700 -0.0404572457  2.403746e-01 0.6370547
## Somerst-Blueste  0.199457980  0.0606803486  3.382356e-01 0.0000357
## StoneBr-Blueste  0.332417307  0.1846479509  4.801867e-01 0.0000000
## SWISU-Blueste   -0.030481949 -0.1790063902  1.180425e-01 1.0000000
## Timber-Blueste   0.226739030  0.0825456812  3.709324e-01 0.0000021
## Veenker-Blueste  0.231299733  0.0704803516  3.921191e-01 0.0000350
## BrkSide-BrDale   0.056428362 -0.0317519415  1.446087e-01 0.8291153
## ClearCr-BrDale   0.285356666  0.1841909298  3.865224e-01 0.0000000
## CollgCr-BrDale   0.269558022  0.1872832821  3.518328e-01 0.0000000
## Crawfor-BrDale   0.275803354  0.1871589893  3.644477e-01 0.0000000
## Edwards-BrDale   0.071857940 -0.0119658063  1.556817e-01 0.2341703
## Gilbert-BrDale   0.253891978  0.1690874080  3.386965e-01 0.0000000
## Greens-BrDale    0.263306775  0.0932904097  4.333231e-01 0.0000035
## GrnHill-BrDale   0.419278968  0.1072435344  7.313144e-01 0.0002193
## IDOTRR-BrDale   -0.037725278 -0.1274382000  5.198764e-02 0.9992130
## Landmrk-BrDale   0.115878647 -0.3184562947  5.502136e-01 0.9999999
```

```
## MeadowV-BrDale  -0.048176098 -0.1531497872  5.679759e-02 0.9967699
## Mitchel-BrDale   0.176096881  0.0884224934  2.637713e-01 0.0000000
## NAmes-BrDale     0.131504400  0.0508974238  2.121114e-01 0.0000006
## NoRidge-BrDale   0.483149056  0.3901078273  5.761903e-01 0.0000000
## NPkVill-BrDale   0.126532096  0.0081140371  2.449502e-01 0.0201768
## NridgHt-BrDale   0.468217035  0.3834517714  5.529823e-01 0.0000000
## NWAmes-BrDale    0.245862677  0.1593815756  3.323438e-01 0.0000000
## OldTown-BrDale   0.048498091 -0.0342620292  1.312582e-01 0.9244535
## Sawyer-BrDale    0.108345733  0.0229384833  1.937530e-01 0.0008550
## SawyerW-BrDale   0.228222253  0.1413552645  3.150892e-01 0.0000000
## Somerst-BrDale   0.327721533  0.2435284994  4.119146e-01 0.0000000
## StoneBr-BrDale   0.460680860  0.3623700341  5.589917e-01 0.0000000
## SWISU-BrDale     0.097781605 -0.0016605681  1.972238e-01 0.0614285
## Timber-BrDale    0.355002583  0.2621534701  4.478517e-01 0.0000000
## Veenker-BrDale   0.359563287  0.2425499989  4.765766e-01 0.0000000
## ClearCr-BrkSide  0.228928304  0.1525116303  3.053450e-01 0.0000000
## CollgCr-BrkSide  0.213129660  0.1644045670  2.618548e-01 0.0000000
## Crawfor-BrkSide  0.219374992  0.1605292169  2.782208e-01 0.0000000
## Edwards-BrkSide  0.015429578 -0.0358678130  6.672697e-02 0.9999988
## Gilbert-BrkSide  0.197463616  0.1445786752  2.503486e-01 0.0000000
## Greens-BrkSide   0.206878413  0.0503199054  3.634369e-01 0.0003313
## GrnHill-BrkSide  0.362850606  0.0579389571  6.677623e-01 0.0032186
## IDOTRR-BrkSide  -0.094153639 -0.1545970868 -3.371019e-02 0.0000029
## Landmrk-BrkSide  0.059450286 -0.3697953828  4.886960e-01 1.0000000
## MeadowV-BrkSide -0.104604459 -0.1859953520 -2.321357e-02 0.0006374
## Mitchel-BrkSide  0.119668520  0.0622943076  1.770427e-01 0.0000000
## NAmes-BrkSide    0.075076038  0.0292231592  1.209289e-01 0.0000005
## NoRidge-BrkSide  0.426720694  0.3614392106  4.920022e-01 0.0000000
## NPkVill-BrkSide  0.070103735 -0.0280178159  1.682253e-01 0.6292919
## NridgHt-BrkSide  0.411788673  0.3589667860  4.646106e-01 0.0000000
## NWAmes-BrkSide   0.189434315  0.1339006966  2.449679e-01 0.0000000
## OldTown-BrkSide -0.007930271 -0.0574705511  4.161001e-02 1.0000000
## Sawyer-BrkSide   0.051917371 -0.0019287074  1.057634e-01 0.0771547
## SawyerW-BrkSide  0.171793891  0.1156612307  2.279266e-01 0.0000000
## Somerst-BrkSide  0.271293171  0.2193945314  3.231918e-01 0.0000000
## StoneBr-BrkSide  0.404252498  0.3316575922  4.768474e-01 0.0000000
## SWISU-BrkSide    0.041353243 -0.0327665765  1.154731e-01 0.9555961
## Timber-BrkSide   0.298574221  0.2335668394  3.635816e-01 0.0000000
## Veenker-BrkSide  0.303134925  0.2067133951  3.995565e-01 0.0000000
## CollgCr-ClearCr -0.015798644 -0.0853175253  5.372024e-02 1.0000000
## Crawfor-ClearCr -0.009553312 -0.0865050203  6.739840e-02 1.0000000
## Edwards-ClearCr -0.213498725 -0.2848441005 -1.421534e-01 0.0000000
## Gilbert-ClearCr -0.031464688 -0.1039599093  4.103053e-02 0.9986678
## Greens-ClearCr  -0.022049890 -0.1862732340  1.421735e-01 1.0000000
## GrnHill-ClearCr  0.133922303 -0.1749949258  4.428395e-01 0.9986929
## IDOTRR-ClearCr  -0.323081943 -0.4012621870 -2.449017e-01 0.0000000
## Landmrk-ClearCr -0.169478018 -0.6015782195  2.626222e-01 0.9997671
## MeadowV-ClearCr -0.333532763 -0.4288386790 -2.382268e-01 0.0000000
## Mitchel-ClearCr -0.109259784 -0.1850921018 -3.342747e-02 0.0000332
## NAmes-ClearCr   -0.153852266 -0.2213891200 -8.631541e-02 0.0000000
```

```
## NoRidge-ClearCr  0.197792390  0.1158142695  2.797705e-01 0.0000000
## NPkVill-ClearCr -0.158824569 -0.2687635470 -4.888559e-02 0.0000308
## NridgHt-ClearCr  0.182860369  0.1104111322  2.553096e-01 0.0000000
## NWAmes-ClearCr  -0.039493988 -0.1139434545  3.495548e-02 0.9756547
## OldTown-ClearCr -0.236858575 -0.3069512245 -1.667659e-01 0.0000000
## Sawyer-ClearCr  -0.177010933 -0.2502102521 -1.038116e-01 0.0000000
## SawyerW-ClearCr -0.057134413 -0.1320317808  1.776296e-02 0.4818487
## Somerst-ClearCr  0.042364868 -0.0294140222  1.141438e-01 0.9187423
## StoneBr-ClearCr  0.175324194  0.0874108248  2.632376e-01 0.0000000
## SWISU-ClearCr   -0.187575061 -0.2767517833 -9.839834e-02 0.0000000
## Timber-ClearCr   0.069645918 -0.0121140964  1.514059e-01 0.2457574
## Veenker-ClearCr  0.074206621 -0.0342177839  1.826310e-01 0.7169191
## Crawfor-CollgCr  0.006245332 -0.0433146532  5.580532e-02 1.0000000
## Edwards-CollgCr -0.197700082 -0.2380087562 -1.573914e-01 0.0000000
## Gilbert-CollgCr -0.015666044 -0.0579766083  2.664452e-02 0.9999196
## Greens-CollgCr  -0.006251247 -0.1595611722  1.470587e-01 1.0000000
## GrnHill-CollgCr  0.149720946 -0.1535355139  4.529774e-01 0.9905743
## IDOTRR-CollgCr  -0.307283299 -0.3587301382 -2.558365e-01 0.0000000
## Landmrk-CollgCr -0.153679374 -0.5817508762  2.743921e-01 0.9999554
## MeadowV-CollgCr -0.317734119 -0.3926863994 -2.427818e-01 0.0000000
## Mitchel-CollgCr -0.093461140 -0.1412645618 -4.565772e-02 0.0000000
## NAmes-CollgCr   -0.138053622 -0.1711572983 -1.049499e-01 0.0000000
## NoRidge-CollgCr  0.213591034  0.1565380853  2.706440e-01 0.0000000
## NPkVill-CollgCr -0.143025925 -0.2358763726 -5.017548e-02 0.0000042
## NridgHt-CollgCr  0.198659013  0.1564272876  2.408907e-01 0.0000000
## NWAmes-CollgCr  -0.023695344 -0.0692732966  2.188261e-02 0.9811466
## OldTown-CollgCr -0.221059931 -0.2591073520 -1.830125e-01 0.0000000
## Sawyer-CollgCr  -0.161212289 -0.2047182317 -1.177063e-01 0.0000000
## SawyerW-CollgCr -0.041335769 -0.0876417345  4.970197e-03 0.1674573
## Somerst-CollgCr  0.058163512  0.0170924057  9.923462e-02 0.0000544
## StoneBr-CollgCr  0.191122838  0.1258282193  2.564175e-01 0.0000000
## SWISU-CollgCr   -0.171776417 -0.2387623448 -1.047905e-01 0.0000000
## Timber-CollgCr   0.085444562  0.0287054515  1.421837e-01 0.0000087
## Veenker-CollgCr  0.090005265 -0.0010467987  1.810573e-01 0.0576514
## Edwards-Crawfor -0.203945414 -0.2560364861 -1.518543e-01 0.0000000
## Gilbert-Crawfor -0.021911376 -0.0755665196  3.174377e-02 0.9995238
## Greens-Crawfor  -0.012496579 -0.1693169341  1.443238e-01 1.0000000
## GrnHill-Crawfor  0.143475614 -0.1615705646  4.485218e-01 0.9953171
## IDOTRR-Crawfor  -0.313528631 -0.3746471049 -2.524102e-01 0.0000000
## Landmrk-Crawfor -0.159924706 -0.5892659475  2.694165e-01 0.9999100
## MeadowV-Crawfor -0.323979451 -0.4058728874 -2.420860e-01 0.0000000
## Mitchel-Crawfor -0.099706472 -0.1577913902 -4.162155e-02 0.0000001
## NAmes-Crawfor   -0.144298954 -0.1910380593 -9.755985e-02 0.0000000
## NoRidge-Crawfor  0.207345702  0.1414387255  2.732527e-01 0.0000000
## NPkVill-Crawfor -0.149271257 -0.2478100626 -5.073245e-02 0.0000072
## NridgHt-Crawfor  0.192413681  0.1388206850  2.460067e-01 0.0000000
## NWAmes-Crawfor  -0.029940677 -0.0862082542  2.632690e-02 0.9747076
## OldTown-Crawfor -0.227305263 -0.2776669233 -1.769436e-01 0.0000000
## Sawyer-Crawfor  -0.167457621 -0.2220603465 -1.128549e-01 0.0000000
## SawyerW-Crawfor -0.047581101 -0.1044399880  9.277786e-03 0.2797443
```

```
## Somerst-Crawfor  0.051918179 -0.0007650846  1.046014e-01 0.0598234
## StoneBr-Crawfor  0.184877506  0.1117196096  2.580354e-01 0.0000000
## SWISU-Crawfor   -0.178021749 -0.2526930629 -1.033504e-01 0.0000000
## Timber-Crawfor   0.079199229  0.0135637422  1.448347e-01 0.0024521
## Veenker-Crawfor  0.083759933 -0.0130861764  1.806060e-01 0.2184794
## Gilbert-Edwards  0.182034038  0.1367850430  2.272830e-01 0.0000000
## Greens-Edwards   0.191448835  0.0373020837  3.455956e-01 0.0013545
## GrnHill-Edwards  0.347421028  0.0437406560  6.511014e-01 0.0066573
## IDOTRR-Edwards  -0.109583218 -0.1634725941 -5.569384e-02 0.0000000
## Landmrk-Edwards  0.044020707 -0.3843512086  4.723926e-01 1.0000000
## MeadowV-Edwards -0.120034038 -0.1966834470 -4.338463e-02 0.0000024
## Mitchel-Edwards  0.104238941  0.0538161809  1.546617e-01 0.0000000
## NAmes-Edwards    0.059646460  0.0228613941  9.643153e-02 0.0000007
## NoRidge-Edwards  0.411291116  0.3520262304  4.705560e-01 0.0000000
## NPkVill-Edwards  0.054674156 -0.0395516003  1.488999e-01 0.9318195
## NridgHt-Edwards  0.396359095  0.3511838105  4.415344e-01 0.0000000
## NWAmes-Edwards   0.174004737  0.1256866517  2.223228e-01 0.0000000
## OldTown-Edwards -0.023359849 -0.0646502104  1.793051e-02 0.9480947
## Sawyer-Edwards   0.036487793 -0.0098808899  8.285648e-02 0.4109464
## SawyerW-Edwards  0.156364313  0.1073589037  2.053697e-01 0.0000000
## Somerst-Edwards  0.255863593  0.2117713785  2.999558e-01 0.0000000
## StoneBr-Edwards  0.388822920  0.3215869554  4.560589e-01 0.0000000
## SWISU-Edwards    0.025923665 -0.0429559564  9.480329e-02 0.9998907
## Timber-Edwards   0.283144643  0.2241818221  3.421075e-01 0.0000000
## Veenker-Edwards  0.287705347  0.1952512176  3.801595e-01 0.0000000
## Greens-Gilbert   0.009414797 -0.1452675082  1.640971e-01 1.0000000
## GrnHill-Gilbert  0.165386990 -0.1385655769  4.693396e-01 0.9668271
## IDOTRR-Gilbert  -0.291617256 -0.3470199589 -2.362146e-01 0.0000000
## Landmrk-Gilbert -0.138013331 -0.5665782533  2.905516e-01 0.9999949
## MeadowV-Gilbert -0.302068076 -0.3797889003 -2.243473e-01 0.0000000
## Mitchel-Gilbert -0.077795097 -0.1298320969 -2.575810e-02 0.0000109
## NAmes-Gilbert   -0.122387578 -0.1613559717 -8.341918e-02 0.0000000
## NoRidge-Gilbert  0.229257078  0.1686128590  2.899013e-01 0.0000000
## NPkVill-Gilbert -0.127359882 -0.2224592394 -3.226052e-02 0.0002383
## NridgHt-Gilbert  0.214325057  0.1673548435  2.612953e-01 0.0000000
## NWAmes-Gilbert  -0.008029301 -0.0580296210  4.197102e-02 1.0000000
## OldTown-Gilbert -0.205393887 -0.2486407201 -1.621471e-01 0.0000000
## Sawyer-Gilbert  -0.145546245 -0.1936653616 -9.742713e-02 0.0000000
## SawyerW-Gilbert -0.025669725 -0.0763345528  2.499510e-02 0.9865676
## Somerst-Gilbert  0.073829555  0.0279000654  1.197590e-01 0.0000010
## StoneBr-Gilbert  0.206788882  0.1383340095  2.752438e-01 0.0000000
## SWISU-Gilbert   -0.156110373 -0.2261803157 -8.604043e-02 0.0000000
## Timber-Gilbert   0.101110605  0.0407615463  1.614597e-01 0.0000002
## Veenker-Gilbert  0.105671309  0.0123269966  1.990156e-01 0.0079944
## GrnHill-Greens   0.155972193 -0.1818160726  4.937605e-01 0.9964536
## IDOTRR-Greens   -0.301032053 -0.4584588894 -1.436052e-01 0.0000000
## Landmrk-Greens  -0.147428128 -0.6006186421  3.057624e-01 0.9999938
## MeadowV-Greens  -0.311482873 -0.4780790191 -1.448867e-01 0.0000000
## Mitchel-Greens  -0.087209894 -0.2434840080  6.906422e-02 0.9554757
## NAmes-Greens    -0.131802375 -0.2842237814  2.061903e-02 0.2187872
```

```
## NoRidge-Greens   0.219842281  0.0604952672  3.791893e-01 0.0001108
## NPkVill-Greens  -0.136774679 -0.3121532328  3.860388e-02 0.4312552
## NridgHt-Greens   0.204910259  0.0502495006  3.595710e-01 0.0003107
## NWAmes-Greens   -0.017444098 -0.1730518786  1.381637e-01 1.0000000
## OldTown-Greens  -0.214808684 -0.3683796383 -6.123773e-02 0.0000765
## Sawyer-Greens   -0.154961042 -0.3099745860  5.250176e-05 0.0502123
## SawyerW-Greens  -0.035084522 -0.1909070949  1.207381e-01 1.0000000
## Somerst-Greens   0.064414758 -0.0899331198  2.187626e-01 0.9993081
## StoneBr-Greens   0.197374085  0.0348938739  3.598543e-01 0.0021440
## SWISU-Greens    -0.165525170 -0.3286924043 -2.357936e-03 0.0416337
## Timber-Greens    0.091695808 -0.0675391073  2.509307e-01 0.9370877
## Veenker-Greens   0.096256511 -0.0781765989  2.706896e-01 0.9608632
## IDOTRR-GrnHill  -0.457004246 -0.7623626522 -1.516458e-01 0.0000106
## Landmrk-GrnHill -0.303400321 -0.8266996519  2.198990e-01 0.9323886
## MeadowV-GrnHill -0.467455066 -0.7776402096 -1.572699e-01 0.0000085
## Mitchel-GrnHill -0.243182087 -0.5479478104  6.158364e-02 0.3798529
## NAmes-GrnHill   -0.287774568 -0.5905828121  1.503368e-02 0.0905024
## NoRidge-GrnHill  0.063870087 -0.2424826800  3.702229e-01 1.0000000
## NPkVill-GrnHill -0.292746872 -0.6077360537  2.224231e-02 0.1142707
## NridgHt-GrnHill  0.048938066 -0.2550035363  3.528797e-01 1.0000000
## NWAmes-GrnHill  -0.173416291 -0.4778408775  1.310083e-01 0.9440197
## OldTown-GrnHill -0.370780877 -0.6741693830 -6.739237e-02 0.0018956
## Sawyer-GrnHill  -0.310933235 -0.6150545042 -6.811966e-03 0.0375725
## SawyerW-GrnHill -0.191056715 -0.4955911495  1.134777e-01 0.8555829
## Somerst-GrnHill -0.091557435 -0.3953399475  2.122251e-01 0.9999987
## StoneBr-GrnHill  0.041401892 -0.2665922096  3.493960e-01 1.0000000
## SWISU-GrnHill   -0.321497363 -0.6298544514 -1.314028e-02 0.0287656
## Timber-GrnHill  -0.064276385 -0.3705708605  2.420181e-01 1.0000000
## Veenker-GrnHill -0.059715682 -0.3741794434  2.547481e-01 1.0000000
## Landmrk-IDOTRR   0.153603925 -0.2759592094  5.831671e-01 0.9999587
## MeadowV-IDOTRR  -0.010450820 -0.0934997210  7.259808e-02 1.0000000
## Mitchel-IDOTRR   0.213822159  0.1541192045  2.735251e-01 0.0000000
## NAmes-IDOTRR     0.169229678  0.1204943816  2.179650e-01 0.0000000
## NoRidge-IDOTRR   0.520874333  0.4535370135  5.882117e-01 0.0000000
## NPkVill-IDOTRR   0.164257374  0.0647562120  2.637585e-01 0.0000004
## NridgHt-IDOTRR   0.505942312  0.4505997943  5.612848e-01 0.0000000
## NWAmes-IDOTRR    0.283587955  0.2256515639  3.415243e-01 0.0000000
## OldTown-IDOTRR   0.086223369  0.0340038134  1.384429e-01 0.0000004
## Sawyer-IDOTRR    0.146071011  0.0897501197  2.023919e-01 0.0000000
## SawyerW-IDOTRR   0.265947530  0.2074366928  3.244584e-01 0.0000000
## Somerst-IDOTRR   0.365446811  0.3109847928  4.199088e-01 0.0000000
## StoneBr-IDOTRR   0.498406137  0.4239570749  5.728552e-01 0.0000000
## SWISU-IDOTRR     0.135506882  0.0595701310  2.114436e-01 0.0000000
## Timber-IDOTRR    0.392727861  0.3256562403  4.597995e-01 0.0000000
## Veenker-IDOTRR   0.397288564  0.2994634449  4.951137e-01 0.0000000
## MeadowV-Landmrk -0.164054745 -0.5970623124  2.689528e-01 0.9998764
## Mitchel-Landmrk  0.060218234 -0.3689237894  4.893603e-01 1.0000000
## NAmes-Landmrk    0.015625753 -0.4121283386  4.433798e-01 1.0000000
## NoRidge-Landmrk  0.367270408 -0.0630001440  7.975410e-01 0.2419812
## NPkVill-Landmrk  0.010653449 -0.4258083645  4.471153e-01 1.0000000
```

```
## NridgHt-Landmrk  0.352338387 -0.0762187593  7.808955e-01 0.3155863
## NWAmes-Landmrk   0.129984030 -0.2989157940  5.588839e-01 0.9999986
## OldTown-Landmrk -0.067380556 -0.4955456127  3.607845e-01 1.0000000
## Sawyer-Landmrk  -0.007532914 -0.4362175027  4.211517e-01 1.0000000
## SawyerW-Landmrk  0.112343606 -0.3166341930  5.413214e-01 0.9999999
## Somerst-Landmrk  0.211842886 -0.2166014453  6.402872e-01 0.9903766
## StoneBr-Landmrk  0.344802213 -0.0866385092  7.762429e-01 0.3763758
## SWISU-Landmrk   -0.018097043 -0.4497969658  4.136029e-01 1.0000000
## Timber-Landmrk   0.239123936 -0.1911051144  6.693530e-01 0.9574815
## Veenker-Landmrk  0.243684639 -0.1923981367  6.797674e-01 0.9547953
## Mitchel-MeadowV  0.224272979  0.1434304788  3.051155e-01 0.0000000
## NAmes-MeadowV    0.179680498  0.1065628107  2.527982e-01 0.0000000
## NoRidge-MeadowV  0.531325153  0.4446915473  6.179588e-01 0.0000000
## NPkVill-MeadowV  0.174708194  0.0612553479  2.881610e-01 0.0000042
## NridgHt-MeadowV  0.516393132  0.4387151987  5.940711e-01 0.0000000
## NWAmes-MeadowV   0.294038775  0.2144919807  3.735856e-01 0.0000000
## OldTown-MeadowV  0.096674189  0.0211894285  1.721589e-01 0.0006904
## Sawyer-MeadowV   0.156521831  0.0781438374  2.348998e-01 0.0000000
## SawyerW-MeadowV  0.276398350  0.1964322002  3.563645e-01 0.0000000
## Somerst-MeadowV  0.375897631  0.2988445422  4.529507e-01 0.0000000
## StoneBr-MeadowV  0.508856957  0.4165870817  6.011268e-01 0.0000000
## SWISU-MeadowV    0.145957702  0.0524833355  2.394321e-01 0.0000026
## Timber-MeadowV   0.403178681  0.3167514320  4.896059e-01 0.0000000
## Veenker-MeadowV  0.407739384  0.2957535769  5.197252e-01 0.0000000
## NAmes-Mitchel   -0.044592481 -0.0894647327  2.797699e-04 0.0540338
## NoRidge-Mitchel  0.307052174  0.2424557000  3.716486e-01 0.0000000
## NPkVill-Mitchel -0.049564785 -0.1472319293  4.810236e-02 0.9862699
## NridgHt-Mitchel  0.292120153  0.2401472357  3.440931e-01 0.0000000
## NWAmes-Mitchel   0.069765796  0.0150390634  1.244925e-01 0.0007664
## OldTown-Mitchel -0.127598790 -0.1762328505 -7.896473e-02 0.0000000
## Sawyer-Mitchel  -0.067751148 -0.1207646618 -1.473763e-02 0.0007244
## SawyerW-Mitchel  0.052125372 -0.0032091384  1.074599e-01 0.0994345
## Somerst-Mitchel  0.151624652  0.1005903373  2.026590e-01 0.0000000
## StoneBr-Mitchel  0.284583979  0.2126044482  3.565635e-01 0.0000000
## SWISU-Mitchel   -0.078315277 -0.1518324859 -4.798067e-03 0.0211122
## Timber-Mitchel   0.178905702  0.1145862480  2.432252e-01 0.0000000
## Veenker-Mitchel  0.183466405  0.0875073315  2.794255e-01 0.0000000
## NoRidge-NAmes    0.351644656  0.2970242310  4.062651e-01 0.0000000
## NPkVill-NAmes   -0.004972304 -0.0963482161  8.640361e-02 1.0000000
## NridgHt-NAmes    0.336712635  0.2978298563  3.755954e-01 0.0000000
## NWAmes-NAmes     0.114358277  0.0718647301  1.568518e-01 0.0000000
## OldTown-NAmes   -0.083006309 -0.1172985545 -4.871406e-02 0.0000000
## Sawyer-NAmes    -0.023158667 -0.0634217880  1.710445e-02 0.9378641
## SawyerW-NAmes    0.096717853  0.0534443703  1.399913e-01 0.0000000
## Somerst-NAmes    0.196217133  0.1585981545  2.338361e-01 0.0000000
## StoneBr-NAmes    0.329176460  0.2659962488  3.923567e-01 0.0000000
## SWISU-NAmes     -0.033722795 -0.0986494182  3.120383e-02 0.9813733
## Timber-NAmes     0.223498183  0.1692056567  2.777907e-01 0.0000000
## Veenker-NAmes    0.228058887  0.1385109658  3.176068e-01 0.0000000
## NPkVill-NoRidge -0.356616959 -0.4591291192 -2.541048e-01 0.0000000
```

```
## NridgHt-NoRidge -0.014932021 -0.0755212611  4.565722e-02 1.0000000
## NWAmes-NoRidge  -0.237286378 -0.3002537279 -1.743190e-01 0.0000000
## OldTown-NoRidge -0.434650965 -0.4924016673 -3.769003e-01 0.0000000
## Sawyer-NoRidge  -0.374803323 -0.4362875032 -3.133191e-01 0.0000000
## SawyerW-NoRidge -0.254926803 -0.3184231006 -1.914305e-01 0.0000000
## Somerst-NoRidge -0.155427522 -0.2152135836 -9.564146e-02 0.0000000
## StoneBr-NoRidge -0.022468196 -0.1008960265  5.595964e-02 0.9999996
## SWISU-NoRidge   -0.385367451 -0.4652088682 -3.055260e-01 0.0000000
## Timber-NoRidge  -0.128146472 -0.1996087981 -5.668415e-02 0.0000000
## Veenker-NoRidge -0.123585769 -0.2244719211 -2.269962e-02 0.0018065
## NridgHt-NPkVill  0.341684938  0.2466206303  4.367492e-01 0.0000000
## NWAmes-NPkVill   0.119330581  0.0227332013  2.159280e-01 0.0015170
## OldTown-NPkVill -0.078034005 -0.1713148195  1.524681e-02 0.2804009
## Sawyer-NPkVill  -0.018186363 -0.1138235480  7.745082e-02 1.0000000
## SawyerW-NPkVill  0.101690156  0.0047471502  1.986332e-01 0.0264025
## Somerst-NPkVill  0.201189437  0.1066350095  2.957439e-01 0.0000000
## StoneBr-NPkVill  0.334148763  0.2268310518  4.414665e-01 0.0000000
## SWISU-NPkVill   -0.028750492 -0.1371055494  7.960457e-02 0.9999999
## Timber-NPkVill   0.228470487  0.1261326610  3.308083e-01 0.0000000
## Veenker-NPkVill  0.233031190  0.1083548801  3.577075e-01 0.0000000
## NWAmes-NridgHt  -0.222354357 -0.2722879809 -1.724207e-01 0.0000000
## OldTown-NridgHt -0.419718944 -0.4628886472 -3.765492e-01 0.0000000
## Sawyer-NridgHt  -0.359871302 -0.4079211103 -3.118215e-01 0.0000000
## SawyerW-NridgHt -0.239994782 -0.2905937886 -1.893958e-01 0.0000000
## Somerst-NridgHt -0.140495501 -0.1863523741 -9.463863e-02 0.0000000
## StoneBr-NridgHt -0.007536175 -0.0759423461  6.087000e-02 1.0000000
## SWISU-NridgHt   -0.370435430 -0.4404577946 -3.004131e-01 0.0000000
## Timber-NridgHt  -0.113214451 -0.1735082624 -5.292064e-02 0.0000000
## Veenker-NridgHt -0.108653748 -0.2019623506 -1.534515e-02 0.0048434
## OldTown-NWAmes  -0.197364586 -0.2438129922 -1.509162e-01 0.0000000
## Sawyer-NWAmes   -0.137516944 -0.1885327773 -8.650111e-02 0.0000000
## SawyerW-NWAmes  -0.017640424 -0.0710641139  3.578327e-02 0.9999915
## Somerst-NWAmes   0.081858856  0.0329029179  1.308148e-01 0.0000002
## StoneBr-NWAmes   0.214818183  0.1442970129  2.853394e-01 0.0000000
## SWISU-NWAmes    -0.148081072 -0.2201710349 -7.599111e-02 0.0000000
## Timber-NWAmes    0.109139906  0.0464567763  1.718230e-01 0.0000000
## Veenker-NWAmes   0.113700609  0.0188305591  2.085707e-01 0.0028040
## Sawyer-OldTown   0.059847642  0.0154306208  1.042647e-01 0.0002045
## SawyerW-OldTown  0.179724162  0.1325611769  2.268871e-01 0.0000000
## Somerst-OldTown  0.279223442  0.2371884518  3.212584e-01 0.0000000
## StoneBr-OldTown  0.412182769  0.3462775951  4.780879e-01 0.0000000
## SWISU-OldTown    0.049283514 -0.0182976909  1.168647e-01 0.5843496
## Timber-OldTown   0.306504492  0.2490638159  3.639452e-01 0.0000000
## Veenker-OldTown  0.311065196  0.2195743055  4.025561e-01 0.0000000
## SawyerW-Sawyer   0.119876520  0.0682092385  1.715438e-01 0.0000000
## Somerst-Sawyer   0.219375800  0.1723428199  2.664088e-01 0.0000000
## StoneBr-Sawyer   0.352335127  0.2831350348  4.215352e-01 0.0000000
## SWISU-Sawyer    -0.010564128 -0.0813622923  6.023404e-02 1.0000000
## Timber-Sawyer    0.246656850  0.1854637778  3.078499e-01 0.0000000
## Veenker-Sawyer   0.251217554  0.1573253612  3.451097e-01 0.0000000
```

```
## Somerst-SawyerW  0.099499280  0.0498648505  1.491337e-01 0.0000000
## StoneBr-SawyerW  0.232458607  0.1614647471  3.034525e-01 0.0000000
## SWISU-SawyerW   -0.130440648 -0.2029930804 -5.788822e-02 0.0000000
## Timber-SawyerW   0.126780330  0.0635658739  1.899948e-01 0.0000000
## Veenker-SawyerW  0.131341034  0.0361190867  2.265630e-01 0.0001116
## StoneBr-Somerst  0.132959327  0.0652635262  2.006551e-01 0.0000000
## SWISU-Somerst   -0.229939928 -0.2992684846 -1.606114e-01 0.0000000
## Timber-Somerst   0.027281050 -0.0322055936  8.676769e-02 0.9968051
## Veenker-Somerst  0.031841753 -0.0609473217  1.246308e-01 0.9999816
## SWISU-StoneBr   -0.362899255 -0.4488236418 -2.769749e-01 0.0000000
## Timber-StoneBr  -0.105678277 -0.1838780995 -2.747845e-02 0.0001898
## Veenker-StoneBr -0.101117573 -0.2068831820  4.648035e-03 0.0847869
## Timber-SWISU     0.257220979  0.1776035208  3.368384e-01 0.0000000
## Veenker-SWISU    0.261781682  0.1549636532  3.685997e-01 0.0000000
## Veenker-Timber   0.004560703 -0.0961483000  1.052697e-01 1.0000000
```

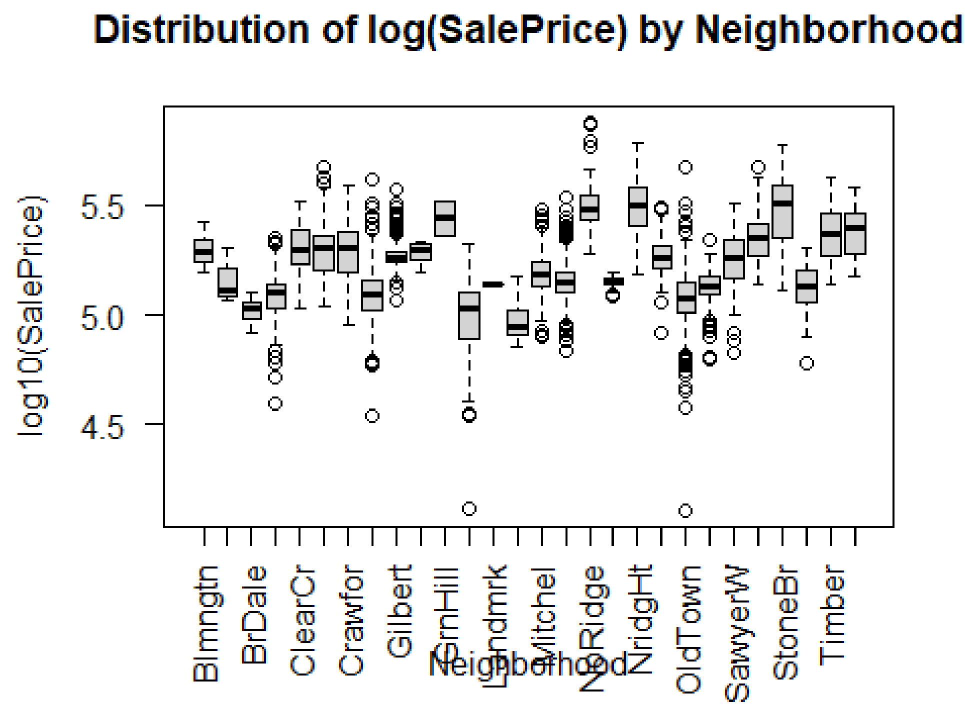
A one-way ANOVA followed by Tukey's HSD was used to compare mean log(SalePrice) across all neighborhoods while controlling the family-wise error rate across multiple pairwise comparisons.

The results reveal substantial heterogeneity in housing prices across neighborhoods. Many pairwise differences are statistically significant, indicating that neighborhood-level price effects are widespread rather than isolated.

Relative to Blmngtn (the OLS reference neighborhood), several neighborhoods (e.g., NoRidge, NridgHt, StoneBr) have significantly higher mean log(SalePrice), while others (e.g., BrDale, MeadowV, IDOTRR) have significantly lower mean log(SalePrice).

These post-hoc results complement the OLS findings by illustrating the magnitude and direction of neighborhood price differences in an unconditional setting.

```r
boxplot(
  logSalePrice ~ Neighborhood,
  data = AmesHousing,
  las = 2,
  main = "Distribution of log(SalePrice) by Neighborhood",
  ylab = "log10(SalePrice)"
)
```



Boxplot Interpretation

The boxplot reveals substantial differences in the distribution of log(SalePrice) across neighborhoods. Median prices vary widely, with neighborhoods such as NoRidge, NridgHt, StoneBr, and Somerst exhibiting notably higher typical prices, while BrDale, MeadowV, IDOTRR, and OldTown are among the lowest.

The limited overlap between several neighborhood distributions visually supports the highly significant one-way ANOVA and the Tukey HSD results, indicating that many pairwise neighborhood differences are economically and statistically meaningful.

At the same time, noticeable within-neighborhood variability and the presence of outliers suggest that neighborhood alone does not fully explain housing prices, motivating the inclusion of structural and quality-related covariates in the multivariable OLS model.

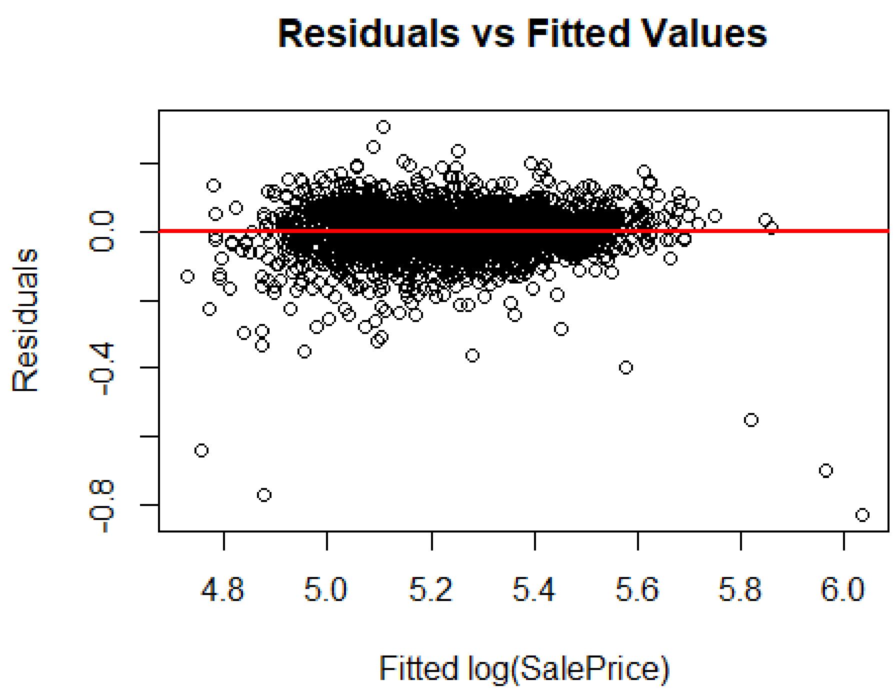
Module 3: Construct AboveExpected indicator

```r
AmesHousing$AboveExpected <-
  as.integer(AmesHousing$logSalePrice > AmesHousing$fitted_logPrice)

# Quick check
table(AmesHousing$AboveExpected)

## 
##    0    1 
## 1354 1576
```

Module 4: OLS Diagnostics

Residuals vs Fitted Values

```r
plot(
  ols_model$fitted.values,
  ols_model$residuals,
  xlab = "Fitted log(SalePrice)",
  ylab = "Residuals",
  main = "Residuals vs Fitted Values"
)
abline(h = 0, col = "red", lwd = 2)
```



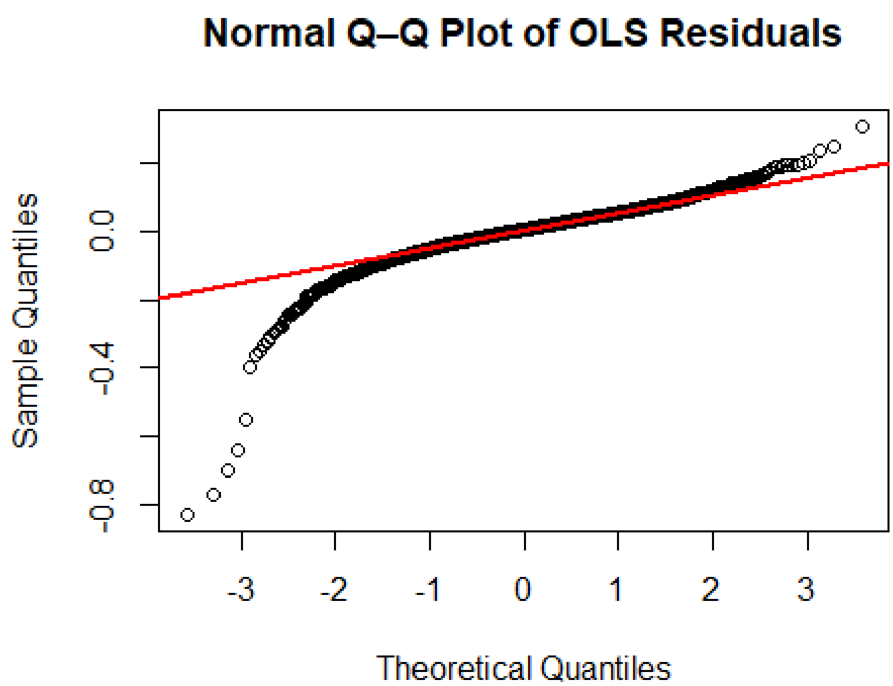
Diagnostic check: Residuals vs Fitted Values

Residuals are densely clustered around zero with no clear curvature, indicating that the linearity assumption is reasonable. The spread of residuals appears roughly constant across fitted values, suggesting no strong evidence of heteroscedasticity after log transformation. A small number of outlying residuals are present, which is expected in cross-sectional housing data and does not invalidate the OLS model.

Module 4B: Normal Q–Q Plot (Normality of Residuals)

```r
qqnorm(
  ols_model$residuals,
  main = "Normal Q-Q Plot of OLS Residuals"
```

```
)
qqline(ols_model$residuals, col = "red", lwd = 2)
```

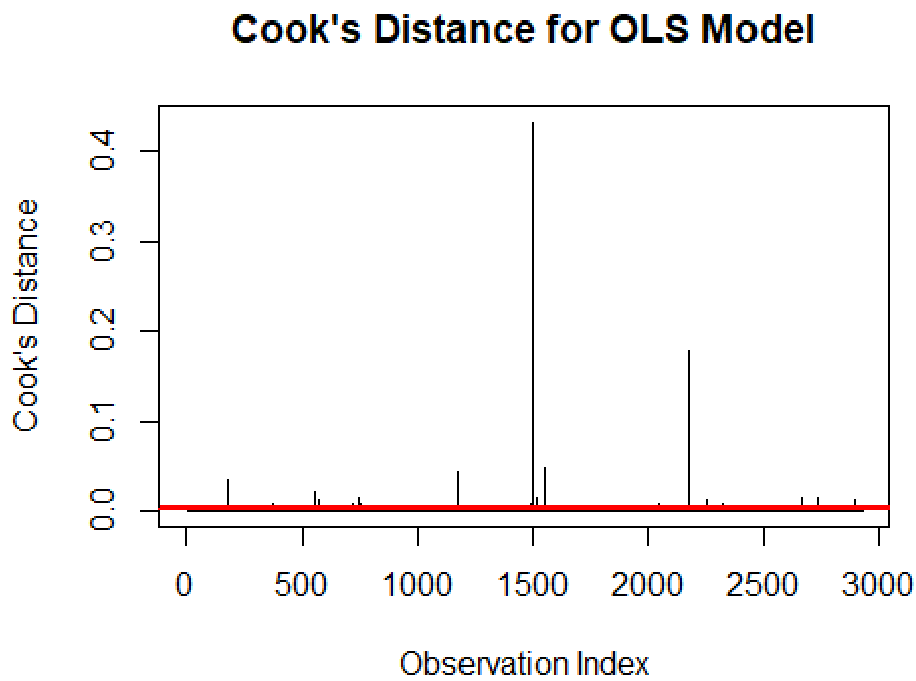


Diagnostic check: Normal Q–Q Plot

Residuals follow the reference line closely in the central region, indicating approximate normality. Deviations occur primarily in the lower and upper tails, reflecting a small number of extreme observations. Given the large sample size, these tail deviations are expected and do not materially violate the normality assumption for OLS inference.

Module 4C: Influence & Leverage (Cook’s Distance)

```
plot(
  cooks.distance(ols_model),
  type = "h",
  main = "Cook's Distance for OLS Model",
  ylab = "Cook's Distance",
  xlab = "Observation Index"
)
abline(h = 4 / nrow(AmesHousing), col = "red", lwd = 2)
```



Diagnostic check: Cook's Distance

The Cook's Distance plot shows that most observations have very low influence on the fitted model. A small number of observations exceed the reference threshold (4/n), indicating potentially influential points. These appear as isolated spikes rather than a widespread pattern, suggesting that no single observation unduly drives the overall regression results.

Module 5: Logistic Regression for AboveExpected

```
logit_model <- glm(
  AboveExpected ~ Neighborhood +
    Overall.Qual +
    Gr.Liv.Area +
    Total.Bsmt.SF +
    Garage.Cars +
    House.Style,
  data = AmesHousing,
  family = binomial(link = "logit")
)

summary(logit_model)

##
## Call:
## glm(formula = AboveExpected ~ Neighborhood + Overall.Qual + Gr.Liv.Area +
##     Total.Bsmt.SF + Garage.Cars + House.Style, family = binomial(link =
```

```
"logit"),
##     data = AmesHousing)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -2.302e-01  5.159e-01  -0.446 0.655441
## NeighborhoodBlueste  4.249e-02  7.582e-01   0.056 0.955306
## NeighborhoodBrDale   5.619e-01  5.520e-01   1.018 0.308633
## NeighborhoodBrkSide  4.668e-01  4.497e-01   1.038 0.299281
## NeighborhoodClearCr  8.585e-02  4.971e-01   0.173 0.862887
## NeighborhoodCollgCr -9.080e-02  4.048e-01  -0.224 0.822531
## NeighborhoodCrawfor  3.166e-01  4.401e-01   0.719 0.471895
## NeighborhoodEdwards  4.488e-01  4.259e-01   1.054 0.292008
## NeighborhoodGilbert  1.737e-01  4.223e-01   0.411 0.680884
## NeighborhoodGreens   1.106e+00  8.287e-01   1.335 0.181947
## NeighborhoodGrnHill  5.848e-01  1.521e+00   0.384 0.700674
## NeighborhoodIDOTRR   8.787e-01  4.617e-01   1.903 0.057030 .
## NeighborhoodLandmrk -1.128e+01  1.970e+02  -0.057 0.954349
## NeighborhoodMeadowV  6.947e-02  5.320e-01   0.131 0.896112
## NeighborhoodMitchel  3.757e-01  4.358e-01   0.862 0.388731
## NeighborhoodNAmes    3.555e-01  4.042e-01   0.880 0.379107
## NeighborhoodNoRidge -9.982e-02  4.639e-01  -0.215 0.829640
## NeighborhoodNPkVill  5.365e-01  5.749e-01   0.933 0.350644
## NeighborhoodNridgHt  1.305e-01  4.179e-01   0.312 0.754845
## NeighborhoodNWAmes   4.510e-01  4.277e-01   1.054 0.291672
## NeighborhoodOldTown  6.266e-01  4.224e-01   1.483 0.137980
## NeighborhoodSawyer   5.173e-01  4.284e-01   1.207 0.227281
## NeighborhoodSawyerW  1.617e-01  4.283e-01   0.377 0.705820
## NeighborhoodSomerst  3.796e-01  4.137e-01   0.918 0.358817
## NeighborhoodStoneBr  3.119e-01  4.807e-01   0.649 0.516465
## NeighborhoodSWISU   -6.899e-02  5.057e-01  -0.136 0.891496
## NeighborhoodTimber   5.502e-03  4.518e-01   0.012 0.990284
## NeighborhoodVeenker  5.850e-01  5.758e-01   1.016 0.309709
## Overall.Qual        -1.602e-01  4.739e-02  -3.380 0.000724 ***
## Gr.Liv.Area          2.169e-04  1.421e-04   1.526 0.127050
## Total.Bsmt.SF        6.992e-04  1.491e-04   4.691 2.72e-06 ***
## Garage.Cars         -5.369e-02  7.012e-02  -0.766 0.443855
## House.Style1.5Unf    1.813e-01  4.943e-01   0.367 0.713748
## House.Style1Story   -8.051e-03  1.576e-01  -0.051 0.959252
## House.Style2.5Fin   -7.244e-01  7.825e-01  -0.926 0.354547
## House.Style2.5Unf    8.526e-02  4.447e-01   0.192 0.847937
## House.Style2Story    2.815e-01  1.604e-01   1.755 0.079286 .
## House.StyleSFoyer    8.798e-02  2.758e-01   0.319 0.749762
## House.StyleSLvl      7.051e-02  2.306e-01   0.306 0.759724
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4045.0  on 2929  degrees of freedom
```

```
## Residual deviance: 3962.9  on 2891  degrees of freedom
## AIC: 4040.9
##
## Number of Fisher Scoring iterations: 10
```

Logistic regression Interpretation

The logistic model estimates the probability that a home sells above its OLS-predicted value. Most Neighborhood indicators are not strongly significant once the baseline expected price is accounted for, which is expected given that Neighborhood effects were already absorbed by the OLS fitted values.

Overall.Qual has a statistically significant negative coefficient, indicating that higher-quality homes are less likely to exceed their predicted price, as quality is already strongly incorporated into the expected value benchmark.

Total.Bsmt.SF is positive and statistically significant, suggesting that basement area contributes to upside deviations beyond what is captured by the OLS model.

Other structural variables (Gr.Liv.Area, Garage.Cars, House.Style) show limited additional explanatory power for AboveExpected status.

Module 5B: Predicted Probabilities and Classification

```
# Compute predicted probabilities from the logistic model
AmesHousing$prob_AboveExpected <- predict(
  logit_model,
  type = "response"
)

# Inspect range of predicted probabilities
summary(AmesHousing$prob_AboveExpected)

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000095 0.4828363 0.5394610 0.5378840 0.5869398 0.9864365
```

Diagnostic check: Predicted probabilities

The predicted probabilities span a wide range, from values near 0 to values close to 1, indicating that the logistic model provides meaningful discrimination between homes that sell above versus below their expected value. The median and mean probabilities are close to 0.5, which is consistent with the relatively balanced AboveExpected outcome.

Module 5C: Classification Using 0.5 Cutoff

```
# Classify AboveExpected based on predicted probability
AmesHousing$pred_class_0.5 <-
  as.integer(AmesHousing$prob_AboveExpected >= 0.5)

# Confusion matrix
cm <- table(
```

```r
  Predicted = AmesHousing$pred_class_0.5,
  Actual    = AmesHousing$AboveExpected
)

cm

##          Actual
## Predicted    0    1
##         0  550  341
##         1  804 1235

# Extract counts programmatically
TN <- cm["0", "0"]
FP <- cm["1", "0"]
FN <- cm["0", "1"]
TP <- cm["1", "1"]

# Compute metrics
accuracy    <- (TP + TN) / sum(cm)
sensitivity <- TP / (TP + FN)
specificity <- TN / (TN + FP)

accuracy

## [1] 0.609215

sensitivity

## [1] 0.7836294

specificity

## [1] 0.4062038
```

Diagnostic check: Classification performance at 0.5 cutoff

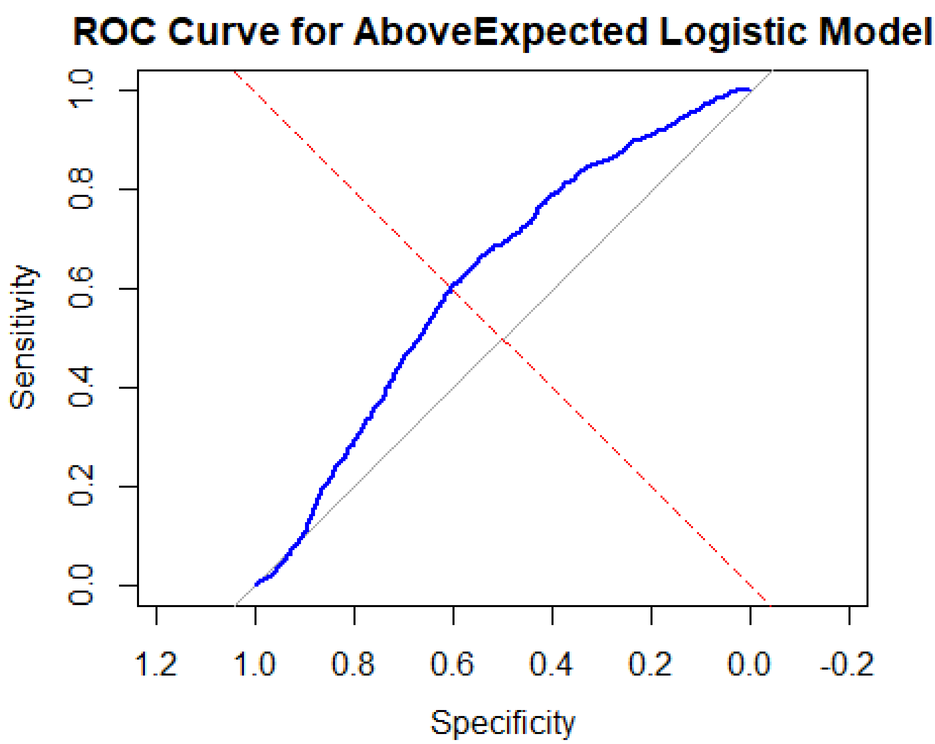
The classifier achieves moderate overall accuracy. Sensitivity is relatively high, indicating good ability to identify homes that sell above their expected value. Specificity is lower, indicating weaker performance in identifying homes that do not sell above expected value. This asymmetry reflects the balanced but noisy nature of price deviations around the expected benchmark.

Module 5D: ROC Curve and AUC

```r
# ROC object using true labels and predicted probabilities
roc_obj <- roc(
  response  = AmesHousing$AboveExpected,
  predictor = AmesHousing$prob_AboveExpected
)

## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```r
# Plot ROC curve
plot(
  roc_obj,
  main = "ROC Curve for AboveExpected Logistic Model",
  col = "blue",
  lwd = 2
)
abline(a = 0, b = 1, lty = 2, col = "red")
```




```r
# Compute AUC
auc_val <- auc(roc_obj)
auc_val

## Area under the curve: 0.6158
```

ROC Curve & AUC Analysis

The ROC curve summarizes classifier performance across all possible probability cutoffs by plotting sensitivity against 1 − specificity. This avoids dependence on an arbitrary threshold such as 0.5.

The AUC (Area Under the Curve) measures overall discrimination ability: the probability that the model assigns a higher predicted probability to a randomly chosen AboveExpected home than to a BelowExpected home.

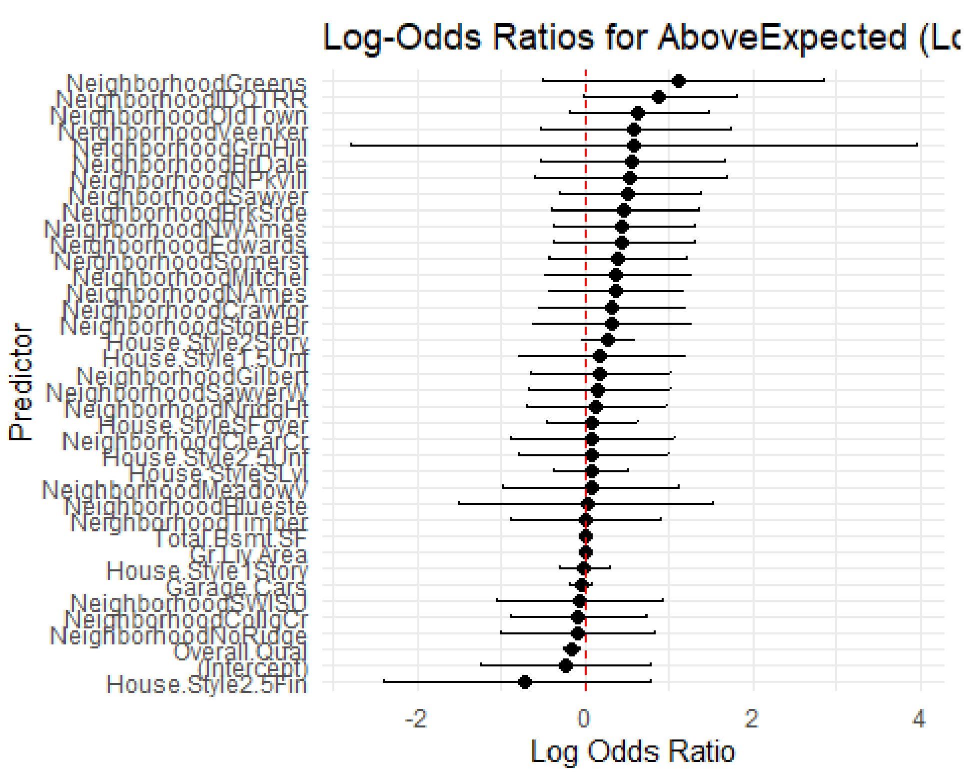
An AUC of 0.5 corresponds to random guessing, while values closer to 1 indicate stronger discriminatory power. Here, AUC ≈ 0.62 suggests modest but meaningful predictive ability beyond chance.

This level of performance is expected given that AboveExpected is defined as a residual-based outcome and therefore contains substantial noise.

Module 5E: Odds Ratios and Coefficient Visualization

```r
# Extract coefficients and confidence intervals
logit_coef <- coef(logit_model)
logit_ci  <- confint(logit_model)

## Waiting for profiling to be done...

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

# Convert to odds ratios
odds_ratios <- exp(logit_coef)
odds_ci <- exp(logit_ci)

# Tidy table
or_plot_data <- data.frame(
  Term   = names(odds_ratios),
  OddsRatio = odds_ratios,
  CI_Lower = odds_ci[, 1],
  CI_Upper = odds_ci[, 2],
  row.names = NULL
)

# Remove non-finite values
or_plot_data <- or_plot_data |>
  filter(is.finite(OddsRatio),
         is.finite(CI_Lower),
         is.finite(CI_Upper))

# Log-scale transformation
or_plot_data <- or_plot_data |>
  mutate(
    logOR = log(OddsRatio),
    logCI_Lower = log(CI_Lower),
```

```
    logCI_Upper = log(CI_Upper)
  )

# Plot log-odds ratios with 95% CI
ggplot(or_plot_data,
       aes(x = logOR,
           y = reorder(Term, logOR))) +
  geom_point(size = 2) +
  geom_errorbarh(
    aes(xmin = logCI_Lower, xmax = logCI_Upper),
    height = 0.2
  ) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "red") +
  labs(
    title = "Log-Odds Ratios for AboveExpected (Logistic Model)",
    x = "Log Odds Ratio",
    y = "Predictor"
  ) +
  theme_minimal()

## Warning: `geom_errobarh()` was deprecated in ggplot2 4.0.0.
## ℹ Please use the `orientation` argument of `geom_errorbar()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## `height` was translated to `width`.
```



Log-odds ratio interpretation

Points represent estimated log-odds ratios and horizontal lines show 95% confidence intervals from the logistic regression predicting whether a home sells above its OLS-predicted price.

The vertical reference line at 0 indicates no effect on the odds of selling above expected. Most neighborhood coefficients are centered near zero with confidence intervals crossing zero, indicating limited additional neighborhood effects once baseline price expectations are accounted for in the OLS model.

Structural variables such as Total.Bsmt.SF show a small positive effect, while Overall.Qual has a negative log-odds ratio, reflecting that higher quality homes are less likely to exceed their predicted price because quality is already incorporated into the expected value.

Overall, the plot shows substantial uncertainty in individual effects and confirms that above-expected sales are only weakly predictable, consistent with the residual-based nature of the outcome.

Module 6: Bootstrap Inference for Mean log(SalePrice)

Using a nonparametric bootstrap to estimate uncertainty in the mean log-transformed SalePrice. Bootstrapping avoids reliance on normality assumptions and is appropriate given the skewed distribution of housing prices.

```r
B <- 10000  # number of bootstrap samples
n <- nrow(AmesHousing)

boot_means <- numeric(B)

for (b in 1:B) {
  sample_indices <- sample(1:n, size = n, replace = TRUE)
  boot_sample <- AmesHousing$logSalePrice[sample_indices]
  boot_means[b] <- mean(boot_sample)
}

summary(boot_means)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.208   5.218   5.221   5.221   5.223   5.233
```

The bootstrap distribution of the mean log(SalePrice) is approximately symmetric and tightly concentrated, indicating stable estimation of the population mean.

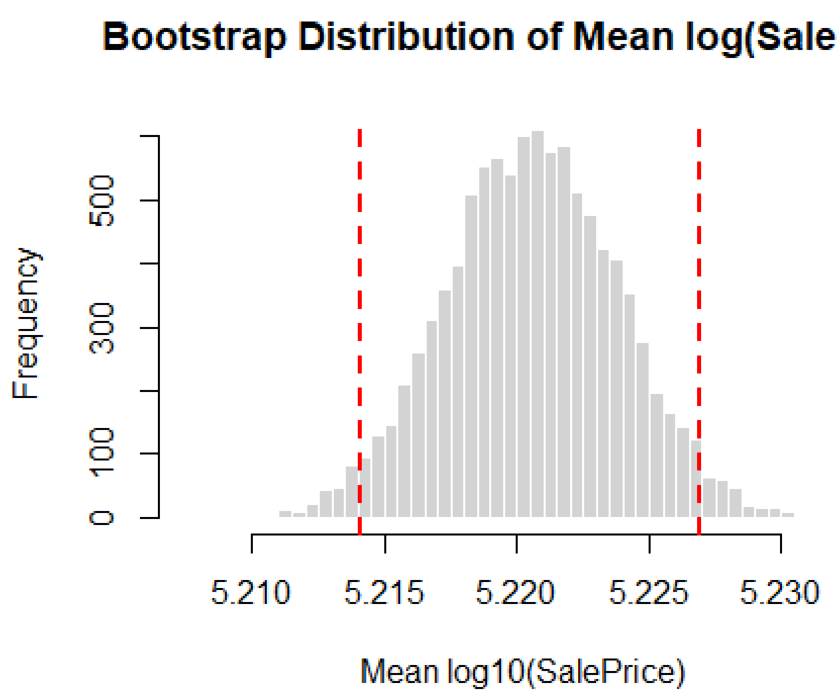
```r
# 95% bootstrap percentile CI on log scale
ci_log <- quantile(boot_means, probs = c(0.025, 0.975))
ci_log

##     2.5%    97.5%
## 5.214109 5.226944
```

```r
# Back-transform to SalePrice scale
ci_price <- 10^ci_log
ci_price
```

```
##     2.5%    97.5%
## 163722.8 168633.7
```

The back-transformed interval provides a 95% confidence interval for the mean SalePrice on the original dollar scale. This corresponds to a geometric mean interpretation due to the log transformation.

```r
#Bootstrap Distribution Visualization
hist(
  boot_means,
  breaks = 40,
  main = "Bootstrap Distribution of Mean log(SalePrice)",
  xlab = "Mean log10(SalePrice)",
  col = "lightgray",
  border = "white"
)
abline(v = ci_log, col = "red", lwd = 2, lty = 2)
```



Illustrative example: interpretation for a single home

```r
# Selecting a single example home
example_home <- AmesHousing[1, ]
```

```r
# Predicting probability of selling above expected price
example_prob <- predict(
logit_model,
newdata = example_home,
type = "response"
)

example_prob

##         1
## 0.5406808
```

For this particular house, the logistic model predicts about a 54% chance that it'll sell for more than what the OLS model expects. So it's slightly more likely than not to beat the predicted price given what we know about the property. This example shows how we can use the model to make predictions for individual homes, not just look at overall patterns.

Summary of Findings

The OLS model captures a substantial portion of the variation in housing prices, with overall quality, above-ground living area, and neighborhood standing out as the key drivers. While neighborhood effects are significant, they don't completely explain price differences once we account for the physical characteristics of the homes.

The logistic regression model predicts whether a home will sell above its expected price and shows moderate discriminatory power (AUC ≈ 0.63). This modest performance makes sense given that we're trying to predict residuals, which are inherently noisy, and it suggests there's limited information beyond what the baseline OLS model already captures.

Bootstrap inference for the mean log(SalePrice) proves to be stable and gives us reliable confidence intervals when we convert back to the dollar scale. Overall, the analysis confirms that structural features and location are the dominant factors in determining housing prices, while predicting which homes will outperform expectations remains challenging.