

Week5 HW5-Ames Housing Data Analysis

ChengHan Li (Alex)

June 23, 2024

Contents

```
#  
knitr::opts_chunk$set(echo = TRUE)  
#  
library(AmesHousing)  
library(ggplot2)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
# Ames  
data <- make_ames()  
  
#  
head(data)
```

```
## # A tibble: 6 x 81  
##   MS_SubClass      MS_Zoning Lot_Frontage Lot_Area Street Alley Lot_Shape  
##   <fct>          <fct>         <dbl>   <int> <fct>  <fct> <fct>  
## 1 One_Story_1946_and_New~ Resident~      141   31770 Pave   No_A~ Slightly~  
## 2 One_Story_1946_and_New~ Resident~       80   11622 Pave   No_A~ Regular  
## 3 One_Story_1946_and_New~ Resident~       81   14267 Pave   No_A~ Slightly~  
## 4 One_Story_1946_and_New~ Resident~       93   11160 Pave   No_A~ Regular  
## 5 Two_Story_1946_and_New~ Resident~       74   13830 Pave   No_A~ Slightly~
```

```
## 6 Two_Story_1946_and_New~ Resident~      78      9978 Pave  No_A~ Slightly~
## # i 74 more variables: Land_Contour <fct>, Utilities <fct>, Lot_Config <fct>,
## #   Land_Slope <fct>, Neighborhood <fct>, Condition_1 <fct>, Condition_2 <fct>,
## #   Bldg_Type <fct>, House_Style <fct>, Overall_Qual <fct>, Overall_Cond <fct>,
## #   Year_Built <int>, Year_Remod_Add <int>, Roof_Style <fct>, Roof_Matl <fct>,
## #   Exterior_1st <fct>, Exterior_2nd <fct>, Mas_Vnr_Type <fct>,
## #   Mas_Vnr_Area <dbl>, Exter_Qual <fct>, Exter_Cond <fct>, Foundation <fct>,
## #   Bsmt_Qual <fct>, Bsmt_Cond <fct>, Bsmt_Exposure <fct>, ...
```

```
#
summary(data)
```

```
##
##              MS_SubClass
## One_Story_1946_and_Newer_All_Styles :1079
## Two_Story_1946_and_Newer           : 575
## One_and_Half_Story_Finished_All_Ages: 287
## One_Story_PUD_1946_and_Newer       : 192
## One_Story_1945_and_Older           : 139
## Two_Story_PUD_1946_and_Newer       : 129
## (Other)                            : 529
##
##              MS_Zoning      Lot_Frontage      Lot_Area
## Floating_Village_Residential: 139  Min.    : 0.00  Min.    : 1300
## Residential_High_Density    : 27   1st Qu.: 43.00 1st Qu.: 7440
## Residential_Low_Density     :2273  Median : 63.00 Median : 9436
## Residential_Medium_Density  : 462  Mean   : 57.65 Mean   : 10148
## A_agr                      : 2     3rd Qu.: 78.00 3rd Qu.: 11555
## C_all                      : 25    Max.    :313.00 Max.    :215245
## I_all                      : 2
##
##      Street      Alley      Lot_Shape      Land_Contour
## Grvl: 12  Gravel      : 120  Regular      :1859  Bnk: 117
## Pave:2918  No_Alley_Access:2732 Slightly_Irregular : 979  HLS: 120
##          Paved      : 78   Moderately_Irregular: 76  Low: 60
##                               Irregular      : 16  Lvl:2633
##
##
##
##      Utilities      Lot_Config      Land_Slope      Neighborhood
## AllPub:2927  Corner : 511  Gtl:2789  North_Ames      : 443
## NoSeWa: 1    CulDSac: 180  Mod: 125  College_Creek   : 267
## NoSewr: 2    FR2      : 85  Sev: 16   Old_Town        : 239
##          FR3      : 14          Edwards            : 194
##          Inside :2140          Somerset            : 182
##                               Northridge_Heights: 166
##                               (Other)           :1439
##
##      Condition_1      Condition_2      Bldg_Type      House_Style
## Norm :2522  Norm :2900  OneFam :2425  One_Story      :1481
## Feedr : 164  Feedr : 13  TwoFmCon: 62  Two_Story      : 873
## Artery : 92  Artery : 5  Duplex : 109  One_and_Half_Fin: 314
## RRAn : 50   PosA : 4   Twnhs : 101  SLvl           : 128
## PosN : 39   PosN : 4   TwnhsE : 233  SFoyer         : 83
## RRAe : 28   RRNn : 2          Two_and_Half_Unf: 24
## (Other): 35  (Other): 2          (Other)         : 27
##
##      Overall_Qual      Overall_Cond      Year_Built      Year_Remod_Add
## Average :825  Average :1654  Min. :1872  Min. :1950
```

```

## Above_Average:732 Above_Average: 533 1st Qu.:1954 1st Qu.:1965
## Good :602 Good : 390 Median :1973 Median :1993
## Very_Good :350 Very_Good : 144 Mean :1971 Mean :1984
## Below_Average:226 Below_Average: 101 3rd Qu.:2001 3rd Qu.:2004
## Excellent :107 Fair : 50 Max. :2010 Max. :2010
## (Other) : 88 (Other) : 58
## Roof_Style Roof_Mat1 Exterior_1st Exterior_2nd Mas_Vnr_Type
## Flat : 20 CompShg:2887 VinylSd:1026 VinylSd:1015 BrkCmn : 25
## Gable :2321 Tar&Grv: 23 MetalSd: 450 MetalSd: 447 BrkFace: 880
## Gambrel: 22 WdShake: 9 HdBoard: 442 HdBoard: 406 CBlock : 1
## Hip : 551 WdShngl: 7 Wd Sdng: 420 Wd Sdng: 397 None :1775
## Mansard: 11 ClyTile: 1 Plywood: 221 Plywood: 274 Stone : 249
## Shed : 5 Membran: 1 CemntBd: 126 CmentBd: 126
## (Other): 2 (Other): 245 (Other): 265
## Mas_Vnr_Area Exter_Qual Exter_Cond Foundation
## Min. : 0.0 Excellent: 107 Excellent: 12 BrkTil: 311
## 1st Qu.: 0.0 Fair : 35 Fair : 67 CBlock:1244
## Median : 0.0 Good : 989 Good : 299 PConc :1310
## Mean : 101.1 Typical :1799 Poor : 3 Slab : 49
## 3rd Qu.: 162.8 Typical :2549 Stone : 11
## Max. :1600.0 Wood : 5
##
## Bsmt_Qual Bsmt_Cond Bsmt_Exposure BsmtFin_Type_1
## Excellent : 258 Excellent : 3 Av : 418 ALQ :429
## Fair : 88 Fair : 104 Gd : 284 BLQ :269
## Good :1219 Good : 122 Mn : 239 GLQ :859
## No_Basement: 80 No_Basement: 80 No :1906 LwQ :154
## Poor : 2 Poor : 5 No_Basement: 83 No_Basement: 80
## Typical :1283 Typical :2616 Rec :288
## Unf :851
## BsmtFin_SF_1 BsmtFin_Type_2 BsmtFin_SF_2 Bsmt_Unf_SF
## Min. :0.000 ALQ : 53 Min. : 0.00 Min. : 0.0
## 1st Qu.:3.000 BLQ : 68 1st Qu.: 0.00 1st Qu.: 219.0
## Median :3.000 GLQ : 34 Median : 0.00 Median : 465.5
## Mean :4.177 LwQ : 89 Mean : 49.71 Mean : 559.1
## 3rd Qu.:7.000 No_Basement: 81 3rd Qu.: 0.00 3rd Qu.: 801.8
## Max. :7.000 Rec : 106 Max. :1526.00 Max. :2336.0
## Unf :2499
## Total_Bsmt_SF Heating Heating_QC Central_Air Electrical
## Min. : 0 Floor: 1 Excellent:1495 N: 196 FuseA : 188
## 1st Qu.: 793 GasA :2885 Fair : 92 Y:2734 FuseF : 50
## Median : 990 GasW : 27 Good : 476 FuseP : 8
## Mean :1051 Grav : 9 Poor : 3 Mix : 1
## 3rd Qu.:1302 OthW : 2 Typical : 864 SBrkr :2682
## Max. :6110 Wall : 6 Unknown: 1
##
## First_Flr_SF Second_Flr_SF Low_Qual_Fin_SF Gr_Liv_Area
## Min. : 334.0 Min. : 0.0 Min. : 0.000 Min. : 334
## 1st Qu.: 876.2 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.:1126
## Median :1084.0 Median : 0.0 Median : 0.000 Median :1442
## Mean :1159.6 Mean : 335.5 Mean : 4.677 Mean :1500
## 3rd Qu.:1384.0 3rd Qu.: 703.8 3rd Qu.: 0.000 3rd Qu.:1743
## Max. :5095.0 Max. :2065.0 Max. :1064.000 Max. :5642
##

```

```

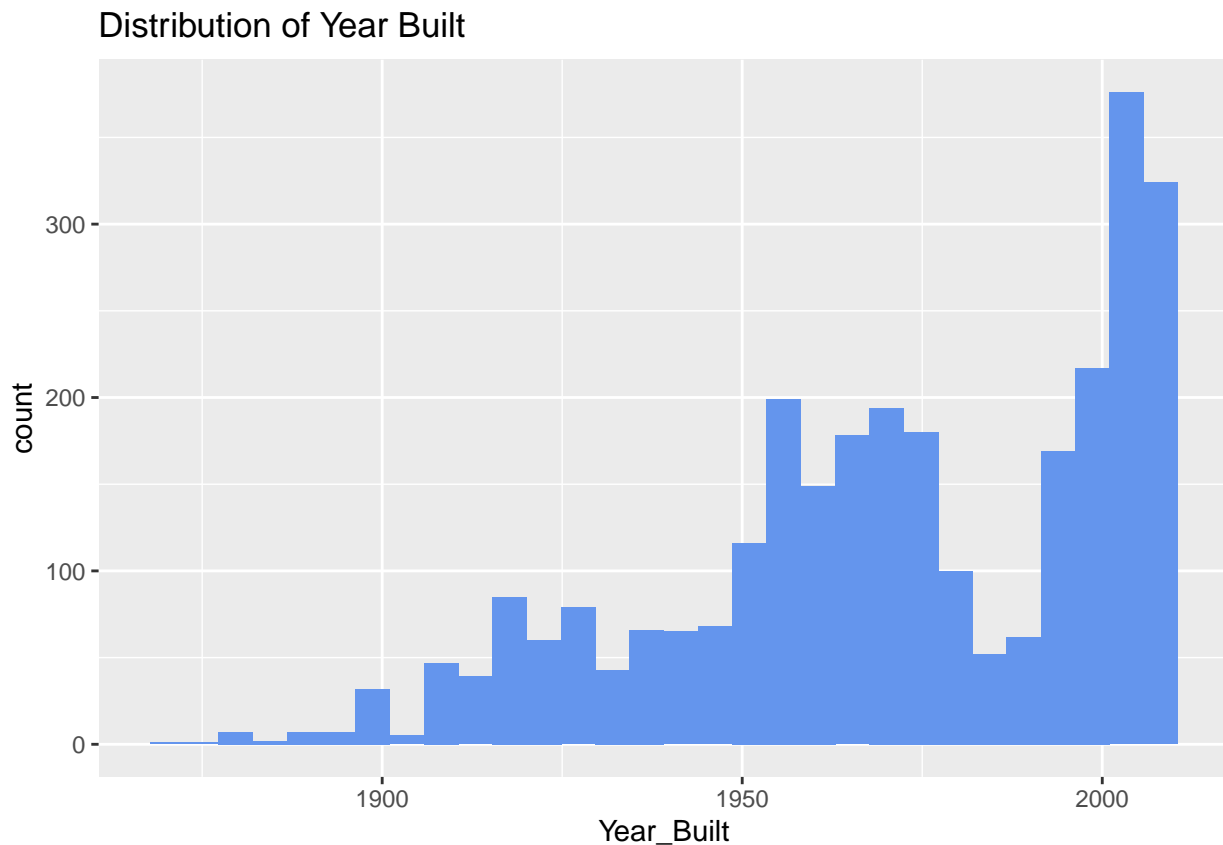
## Bsmt_Full_Bath    Bsmt_Half_Bath    Full_Bath    Half_Bath
## Min.    :0.0000    Min.    :0.00000    Min.    :0.000    Min.    :0.0000
## 1st Qu.:0.0000    1st Qu.:0.00000    1st Qu.:1.000    1st Qu.:0.0000
## Median :0.0000    Median :0.00000    Median :2.000    Median :0.0000
## Mean    :0.4311    Mean    :0.06109    Mean    :1.567    Mean    :0.3795
## 3rd Qu.:1.0000    3rd Qu.:0.00000    3rd Qu.:2.000    3rd Qu.:1.0000
## Max.    :3.0000    Max.    :2.00000    Max.    :4.000    Max.    :2.0000
##
## Bedroom_AbvGr    Kitchen_AbvGr    Kitchen_Qual    TotRms_AbvGrd
## Min.    :0.000    Min.    :0.000    Excellent: 205    Min.    : 2.000
## 1st Qu.:2.000    1st Qu.:1.000    Fair      : 70    1st Qu.: 5.000
## Median :3.000    Median :1.000    Good     :1160    Median : 6.000
## Mean    :2.854    Mean    :1.044    Poor     : 1    Mean    : 6.443
## 3rd Qu.:3.000    3rd Qu.:1.000    Typical  :1494    3rd Qu.: 7.000
## Max.    :8.000    Max.    :3.000                Max.    :15.000
##
## Functional    Fireplaces    Fireplace_Qu    Garage_Type
## Typ      :2728    Min.    :0.0000    Excellent   : 43    Attchd      :1731
## Min2     : 70    1st Qu.:0.0000    Fair       : 75    Basement    : 36
## Min1     : 65    Median :1.0000    Good       : 744    BuiltIn     : 186
## Mod      : 35    Mean    :0.5993    No_Fireplace:1422    CarPort     : 15
## Maj1     : 19    3rd Qu.:1.0000    Poor       : 46    Detchd      : 782
## Maj2     : 9    Max.    :4.0000    Typical    : 600    More_Than_Two_Types: 23
## (Other): 4                No_Garage      : 157
##
## Garage_Finish    Garage_Cars    Garage_Area    Garage_Qual
## Fin      : 728    Min.    :0.000    Min.    : 0.0    Excellent: 3
## No_Garage: 159    1st Qu.:1.000    1st Qu.: 320.0    Fair      : 124
## RFn      : 812    Median :2.000    Median : 480.0    Good      : 24
## Unf      :1231    Mean    :1.766    Mean    : 472.7    No_Garage: 159
##           3rd Qu.:2.000    3rd Qu.: 576.0    Poor      : 5
##           Max.    :5.000    Max.    :1488.0    Typical   :2615
##
##
## Garage_Cond    Paved_Drive    Wood_Deck_SF    Open_Porch_SF
## Excellent: 3    Dirt_Gravel   : 216    Min.    : 0.00    Min.    : 0.00
## Fair      : 74    Partial_Pavement: 62    1st Qu.: 0.00    1st Qu.: 0.00
## Good      : 15    Paved         :2652    Median   : 0.00    Median   : 27.00
## No_Garage: 159                Mean    : 93.75    Mean     : 47.53
## Poor      : 14                3rd Qu.: 168.00    3rd Qu.: 70.00
## Typical   :2665                Max.    :1424.00    Max.     :742.00
##
##
## Enclosed_Porch    Three_season_porch    Screen_Porch    Pool_Area
## Min.    : 0.00    Min.    : 0.000    Min.    : 0    Min.    : 0.000
## 1st Qu.: 0.00    1st Qu.: 0.000    1st Qu.: 0    1st Qu.: 0.000
## Median : 0.00    Median : 0.000    Median : 0    Median : 0.000
## Mean    : 23.01    Mean    : 2.592    Mean    : 16    Mean    : 2.243
## 3rd Qu.: 0.00    3rd Qu.: 0.000    3rd Qu.: 0    3rd Qu.: 0.000
## Max.    :1012.00    Max.    :508.000    Max.    :576    Max.    :800.000
##
##
## Pool_QC    Fence    Misc_Feature    Misc_Val
## Excellent: 4    Good_Privacy   : 118    Elev: 1    Min.    : 0.00
## Fair      : 2    Good_Wood      : 112    Gar2: 5    1st Qu.: 0.00
## Good      : 4    Minimum_Privacy : 330    None:2824    Median   : 0.00
## No_Pool   :2917    Minimum_Wood_Wire: 12    Othr: 4    Mean    : 50.63
## Typical   : 3    No_Fence       :2358    Shed: 95    3rd Qu.: 0.00

```

```
##                               TenC:    1    Max.    :17000.00
##
##      Mo_Sold      Year_Sold      Sale_Type      Sale_Condition      Sale_Price
##  Min.    : 1.000    Min.    :2006    WD       :2536    Abnorml: 190    Min.    : 12789
##  1st Qu.: 4.000    1st Qu.:2007    New      : 239    AdjLand: 12    1st Qu.:129500
##  Median : 6.000    Median :2008    COD      :  87    Alloca  : 24    Median :160000
##  Mean   : 6.216    Mean   :2008    ConLD    :  26    Family  : 46    Mean   :180796
##  3rd Qu.: 8.000    3rd Qu.:2009    CWD      :  12    Normal  :2413   3rd Qu.:213500
##  Max.   :12.000    Max.   :2010    ConLI    :   9    Partial: 245    Max.   :755000
##
##      Longitude      Latitude
##  Min.    : -93.69    Min.    :41.99
##  1st Qu.: -93.66    1st Qu.:42.02
##  Median : -93.64    Median :42.03
##  Mean   : -93.64    Mean   :42.03
##  3rd Qu.: -93.62    3rd Qu.:42.05
##  Max.   : -93.58    Max.   :42.06
##
```

```
#
options(max.print = 1000)

# 'Year_Built'
ggplot(data, aes(x = Year_Built)) +
  geom_histogram(bins = 30, fill = "cornflowerblue") +
  ggtitle("Distribution of Year Built")
```



```

# 'SalePrice'
if ("SalePrice" %in% names(data)) {
  # 'Gr_Liv_Area' 'SalePrice'
  ggplot(data, aes(x = Gr_Liv_Area, y = SalePrice)) +
    geom_point(alpha = 0.6) +
    geom_smooth(method = lm, color = "red", se = FALSE) +
    ggtitle("Relationship between Ground Living Area and Sale Price")
} else {
  cat("The SalePrice column is missing from the dataset.\n")
}

```

The SalePrice column is missing from the dataset.

```

# 'SalePrice'
if ("SalePrice" %in% names(data)) {
  X <- data$Gr_Liv_Area
  y <- data$SalePrice
  trainIndex <- createDataPartition(y, p = .8, list = FALSE, times = 1)
  X_train <- X[trainIndex]
  X_test <- X[-trainIndex]
  y_train <- y[trainIndex]
  y_test <- y[-trainIndex]
  model <- lm(y_train ~ X_train)
  y_pred <- predict(model, newdata = data.frame(X_train = X_test))
  mse <- mean((y_test - y_pred)^2)
  r2 <- summary(model)$r.squared
  cat("Mean Squared Error: ", mse, "\n")
  cat("R^2 Score: ", r2, "\n")
} else {
  cat("The SalePrice column is missing from the dataset.\n")
}

```

The SalePrice column is missing from the dataset.