Modeling housing prices

Chase Mathis

Import Data/libraries

```
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.1 --

v ggplot2 3.3.5     v purrr   0.3.4
v tibble  3.1.6     v dplyr   1.0.7
v tidyr   1.1.4     v stringr 1.4.0
v readr   2.1.1     v forcats 0.5.1

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(tidymodels)
```

```
Registered S3 method overwritten by 'tune':
  method                   from
  required_pkgs.model_spec parsnip

-- Attaching packages --------------------------------------- tidymodels 0.1.4 --

v broom        0.7.11     v rsample      0.1.1
v dials        0.1.0      v tune         0.1.6
v infer        1.0.0      v workflows    0.2.4
v modeldata    0.1.1      v workflowsets 0.1.0
v parsnip      0.1.7      v yardstick    0.0.9
v recipes      0.2.0
```

```
-- Conflicts --------------------------------------- tidymodels_conflicts() --
x scales::discard() masks purrr::discard()
x dplyr::filter()   masks stats::filter()
x recipes::fixed()  masks stringr::fixed()
x dplyr::lag()      masks stats::lag()
x yardstick::spec() masks readr::spec()
x recipes::step()   masks stats::step()
* Search for functions across packages at https://www.tidymodels.org/find/
```

```
library(knitr)
library(GGally)
```

```
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
```

```
test_house <- read_csv("data/test.csv")
```

```
Rows: 1459 Columns: 80

-- Column specification --------------------------------------------------------
Delimiter: ","
chr (43): MSZoning, Street, Alley, LotShape, LandContour, Utilities, LotConf...
dbl (37): Id, MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond, Ye...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
train_house <- read_csv("data/train.csv")
```

```
Rows: 1460 Columns: 81

-- Column specification --------------------------------------------------------
Delimiter: ","
chr (43): MSZoning, Street, Alley, LotShape, LandContour, Utilities, LotConf...
dbl (38): Id, MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond, Ye...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Goal

Create a model predicting sales price.

Selecting variables

```
glimpse(train_house)
```

```
Rows: 1,460
Columns: 81
$ Id            <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1~
$ MSSubClass    <dbl> 60, 20, 60, 70, 60, 50, 20, 60, 50, 190, 20, 60, 20, 20,~
$ MSZoning      <chr> "RL", "RL", "RL", "RL", "RL", "RL", "RL", "RL", "RM", "R~
$ LotFrontage   <dbl> 65, 80, 68, 60, 84, 85, 75, NA, 51, 50, 70, 85, NA, 91, ~
$ LotArea       <dbl> 8450, 9600, 11250, 9550, 14260, 14115, 10084, 10382, 612~
$ Street        <chr> "Pave", "Pave", "Pave", "Pave", "Pave", "Pave", "Pave", ~
$ Alley         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ LotShape      <chr> "Reg", "Reg", "IR1", "IR1", "IR1", "IR1", "Reg", "IR1", ~
$ LandContour   <chr> "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", ~
$ Utilities     <chr> "AllPub", "AllPub", "AllPub", "AllPub", "AllPub", "AllPu~
$ LotConfig     <chr> "Inside", "FR2", "Inside", "Corner", "FR2", "Inside", "I~
$ LandSlope     <chr> "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", ~
$ Neighborhood  <chr> "CollgCr", "Veenker", "CollgCr", "Crawfor", "NoRidge", "~
$ Condition1    <chr> "Norm", "Feedr", "Norm", "Norm", "Norm", "Norm", "Norm",~
$ Condition2    <chr> "Norm", "Norm", "Norm", "Norm", "Norm", "Norm", "Norm", ~
$ BldgType      <chr> "1Fam", "1Fam", "1Fam", "1Fam", "1Fam", "1Fam", "1Fam", ~
$ HouseStyle    <chr> "2Story", "1Story", "2Story", "2Story", "2Story", "1.5Fi~
$ OverallQual   <dbl> 7, 6, 7, 7, 8, 5, 8, 7, 7, 5, 5, 9, 5, 7, 6, 7, 6, 4, 5,~
$ OverallCond   <dbl> 5, 8, 5, 5, 5, 5, 5, 6, 5, 6, 5, 5, 6, 5, 5, 8, 7, 5, 5,~
$ YearBuilt     <dbl> 2003, 1976, 2001, 1915, 2000, 1993, 2004, 1973, 1931, 19~
$ YearRemodAdd  <dbl> 2003, 1976, 2002, 1970, 2000, 1995, 2005, 1973, 1950, 19~
$ RoofStyle     <chr> "Gable", "Gable", "Gable", "Gable", "Gable", "Gable", "G~
$ RoofMatl      <chr> "CompShg", "CompShg", "CompShg", "CompShg", "CompShg", "~
$ Exterior1st   <chr> "VinylSd", "MetalSd", "VinylSd", "Wd Sdng", "VinylSd", "~
$ Exterior2nd   <chr> "VinylSd", "MetalSd", "VinylSd", "Wd Shng", "VinylSd", "~
$ MasVnrType    <chr> "BrkFace", "None", "BrkFace", "None", "BrkFace", "None",~
$ MasVnrArea    <dbl> 196, 0, 162, 0, 350, 0, 186, 240, 0, 0, 0, 286, 0, 306, ~
$ ExterQual     <chr> "Gd", "TA", "Gd", "TA", "Gd", "TA", "Gd", "TA", "TA", "T~
$ ExterCond     <chr> "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "T~
$ Foundation    <chr> "PConc", "CBlock", "PConc", "BrkTil", "PConc", "Wood", "~
```

```
$ BsmtQual      <chr> "Gd", "Gd", "Gd", "TA", "Gd", "Gd", "Ex", "Gd", "TA", "T~
$ BsmtCond      <chr> "TA", "TA", "TA", "Gd", "TA", "TA", "TA", "TA", "TA", "T~
$ BsmtExposure  <chr> "No", "Gd", "Mn", "No", "Av", "No", "Av", "Mn", "No", "N~
$ BsmtFinType1  <chr> "GLQ", "ALQ", "GLQ", "ALQ", "GLQ", "GLQ", "GLQ", "ALQ", ~
$ BsmtFinSF1    <dbl> 706, 978, 486, 216, 655, 732, 1369, 859, 0, 851, 906, 99~
$ BsmtFinType2  <chr> "Unf", "Unf", "Unf", "Unf", "Unf", "Unf", "Unf", "BLQ", ~
$ BsmtFinSF2    <dbl> 0, 0, 0, 0, 0, 0, 0, 32, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ BsmtUnfSF     <dbl> 150, 284, 434, 540, 490, 64, 317, 216, 952, 140, 134, 17~
$ TotalBsmtSF   <dbl> 856, 1262, 920, 756, 1145, 796, 1686, 1107, 952, 991, 10~
$ Heating       <chr> "GasA", "GasA", "GasA", "GasA", "GasA", "GasA", "GasA", ~
$ HeatingQC     <chr> "Ex", "Ex", "Ex", "Gd", "Ex", "Ex", "Ex", "Ex", "Gd", "E~
$ CentralAir    <chr> "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "~
$ Electrical    <chr> "SBrkr", "SBrkr", "SBrkr", "SBrkr", "SBrkr", "SBrkr", "S~
$ `1stFlrSF`    <dbl> 856, 1262, 920, 961, 1145, 796, 1694, 1107, 1022, 1077, ~
$ `2ndFlrSF`    <dbl> 854, 0, 866, 756, 1053, 566, 0, 983, 752, 0, 0, 1142, 0,~
$ LowQualFinSF  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ GrLivArea     <dbl> 1710, 1262, 1786, 1717, 2198, 1362, 1694, 2090, 1774, 10~
$ BsmtFullBath  <dbl> 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1,~
$ BsmtHalfBath  <dbl> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ FullBath      <dbl> 2, 2, 2, 1, 2, 1, 2, 2, 2, 1, 1, 3, 1, 2, 1, 1, 1, 2, 1,~
$ HalfBath      <dbl> 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1,~
$ BedroomAbvGr  <dbl> 3, 3, 3, 3, 4, 1, 3, 3, 2, 2, 3, 4, 2, 3, 2, 2, 2, 2, 3,~
$ KitchenAbvGr  <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1,~
$ KitchenQual   <chr> "Gd", "TA", "Gd", "Gd", "Gd", "TA", "Gd", "TA", "TA", "T~
$ TotRmsAbvGrd  <dbl> 8, 6, 6, 7, 9, 5, 7, 7, 8, 5, 5, 11, 4, 7, 5, 5, 5, 6, 6~
$ Functional    <chr> "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", ~
$ Fireplaces    <dbl> 0, 1, 1, 1, 1, 0, 1, 2, 2, 2, 0, 2, 0, 1, 1, 0, 1, 0, 0,~
$ FireplaceQu   <chr> NA, "TA", "TA", "Gd", "TA", NA, "Gd", "TA", "TA", "TA", ~
$ GarageType    <chr> "Attchd", "Attchd", "Attchd", "Detchd", "Attchd", "Attch~
$ GarageYrBlt   <dbl> 2003, 1976, 2001, 1998, 2000, 1993, 2004, 1973, 1931, 19~
$ GarageFinish  <chr> "RFn", "RFn", "RFn", "Unf", "RFn", "Unf", "RFn", "RFn", ~
$ GarageCars    <dbl> 2, 2, 2, 3, 3, 2, 2, 2, 2, 1, 1, 3, 1, 3, 1, 2, 2, 2, 2,~
$ GarageArea    <dbl> 548, 460, 608, 642, 836, 480, 636, 484, 468, 205, 384, 7~
$ GarageQual    <chr> "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "Fa", "G~
$ GarageCond    <chr> "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "T~
$ PavedDrive    <chr> "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "~
$ WoodDeckSF    <dbl> 0, 298, 0, 0, 192, 40, 255, 235, 90, 0, 0, 147, 140, 160~
$ OpenPorchSF   <dbl> 61, 0, 42, 35, 84, 30, 57, 204, 0, 4, 0, 21, 0, 33, 213,~
$ EnclosedPorch <dbl> 0, 0, 0, 272, 0, 0, 0, 228, 205, 0, 0, 0, 0, 0, 176, 0, ~
$ `3SsnPorch`   <dbl> 0, 0, 0, 0, 0, 320, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
$ ScreenPorch   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 176, 0, 0, 0, 0, 0, ~
$ PoolArea      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ PoolQC        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

```
$ Fence         <chr> NA, NA, NA, NA, NA, "MnPrv", NA, NA, NA, NA, NA, NA, NA,~
$ MiscFeature   <chr> NA, NA, NA, NA, NA, "Shed", NA, "Shed", NA, NA, NA, NA, ~
$ MiscVal       <dbl> 0, 0, 0, 0, 0, 700, 0, 350, 0, 0, 0, 0, 0, 0, 0, 0, 700,~
$ MoSold        <dbl> 2, 5, 9, 2, 12, 10, 8, 11, 4, 1, 2, 7, 9, 8, 5, 7, 3, 10~
$ YrSold        <dbl> 2008, 2007, 2008, 2006, 2008, 2009, 2007, 2009, 2008, 20~
$ SaleType      <chr> "WD", "WD", "WD", "WD", "WD", "WD", "WD", "WD", "WD", "W~
$ SaleCondition <chr> "Normal", "Normal", "Normal", "Abnorml", "Normal", "Norm~
$ SalePrice     <dbl> 208500, 181500, 223500, 140000, 250000, 143000, 307000, ~
```

```
# from: https://sebastiansauer.github.io/NAs-with-dplyr/
train_house %>%
  select_if(function(x) any(is.na(x))) %>%
  summarise_each(funs(sum(is.na(.)))) -> extra_NA
```

```
Warning: `summarise_each_()` was deprecated in dplyr 0.7.0.
Please use `across()` instead.
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
Warning: `funs()` was deprecated in dplyr 0.8.0.
Please use a list of either functions or lambdas:

  # Simple named list:
  list(mean = mean, median = median)

  # Auto named with `tibble::lst()`:
  tibble::lst(mean, median)

  # Using lambdas
  list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
extra_NA <- pivot_longer(cols = everything(), extra_NA, values_to = 'count')
extra_NA %>%
  mutate(prop = count/nrow(train_house)) %>%
  arrange(desc(prop)) %>%
  select(name, prop) %>%
  kable(digits = 3)
```

name	prop
PoolQC	0.995
MiscFeature	0.963
Alley	0.938
Fence	0.808
FireplaceQu	0.473
LotFrontage	0.177
GarageType	0.055
GarageYrBlt	0.055
GarageFinish	0.055
GarageQual	0.055
GarageCond	0.055
BsmtExposure	0.026
BsmtFinType2	0.026
BsmtQual	0.025
BsmtCond	0.025
BsmtFinType1	0.025
MasVnrType	0.005
MasVnrArea	0.005
Electrical	0.001

Looking at the columns and how many values are n/a, I am going to take out the top 5 in this table for these predictors have so many n/a values.

```
colum <- extra_NA %>%
  mutate(prop = count/nrow(train_house)) %>%
  arrange(desc(prop)) %>%
  slice_head(n = 5) %>%
  pull(name)

train_house <- train_house %>%
  select(!colum)
```

```
Note: Using an external vector in selections is ambiguous.
i Use `all_of(colum)` instead of `colum` to silence this message.
i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
This message is displayed once per session.
```

One character has only one unique factor, which causes an error in the regression. Also n/a values should be ommited.

```r
lst <- train_house %>%
  select(where(is.character)) %>%
  sapply(unique)

lst
```

```
$MSZoning
[1] "RL"      "RM"      "C (all)" "FV"      "RH"

$Street
[1] "Pave" "Grvl"

$LotShape
[1] "Reg" "IR1" "IR2" "IR3"

$LandContour
[1] "Lvl" "Bnk" "Low" "HLS"

$Utilities
[1] "AllPub" "NoSeWa"

$LotConfig
[1] "Inside"  "FR2"     "Corner"  "CulDSac" "FR3"

$LandSlope
[1] "Gtl" "Mod" "Sev"

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

$Condition1
[1] "Norm"   "Feedr"  "PosN"   "Artery" "RRAe"   "RRNn"   "RRAn"   "PosA"
[9] "RRNe"

$Condition2
[1] "Norm"   "Artery" "RRNn"   "Feedr"  "PosN"   "PosA"   "RRAn"   "RRAe"

$BldgType
[1] "1Fam"   "2fmCon" "Duplex" "TwnhsE" "Twnhs"
```

```
$HouseStyle
[1] "2Story" "1Story" "1.5Fin" "1.5Unf" "SFoyer" "SLvl"   "2.5Unf" "2.5Fin"

$RoofStyle
[1] "Gable"   "Hip"     "Gambrel" "Mansard" "Flat"    "Shed"

$RoofMatl
[1] "CompShg" "WdShngl" "Metal"   "WdShake" "Membran" "Tar&Grv" "Roll"
[8] "ClyTile"

$Exterior1st
 [1] "VinylSd" "MetalSd" "Wd Sdng" "HdBoard" "BrkFace" "WdShing" "CemntBd"
 [8] "Plywood" "AsbShng" "Stucco"  "BrkComm" "AsphShn" "Stone"   "ImStucc"
[15] "CBlock"

$Exterior2nd
 [1] "VinylSd" "MetalSd" "Wd Shng" "HdBoard" "Plywood" "Wd Sdng" "CmentBd"
 [8] "BrkFace" "Stucco"  "AsbShng" "Brk Cmn" "ImStucc" "AsphShn" "Stone"
[15] "Other"   "CBlock"

$MasVnrType
[1] "BrkFace" "None"    "Stone"   "BrkCmn"  NA

$ExterQual
[1] "Gd" "TA" "Ex" "Fa"

$ExterCond
[1] "TA" "Gd" "Fa" "Po" "Ex"

$Foundation
[1] "PConc"  "CBlock" "BrkTil" "Wood"   "Slab"   "Stone"

$BsmtQual
[1] "Gd" "TA" "Ex" NA   "Fa"

$BsmtCond
[1] "TA" "Gd" NA   "Fa" "Po"

$BsmtExposure
[1] "No" "Gd" "Mn" "Av" NA

$BsmtFinType1
```

```
[1] "GLQ" "ALQ" "Unf" "Rec" "BLQ" NA    "LwQ"

$BsmtFinType2
[1] "Unf" "BLQ" NA    "ALQ" "Rec" "LwQ" "GLQ"

$Heating
[1] "GasA"  "GasW"  "Grav"  "Wall"  "OthW"  "Floor"

$HeatingQC
[1] "Ex" "Gd" "TA" "Fa" "Po"

$CentralAir
[1] "Y" "N"

$Electrical
[1] "SBrkr" "FuseF" "FuseA" "FuseP" "Mix"   NA

$KitchenQual
[1] "Gd" "TA" "Ex" "Fa"

$Functional
[1] "Typ"  "Min1" "Maj1" "Min2" "Mod"  "Maj2" "Sev"

$GarageType
[1] "Attchd"  "Detchd"  "BuiltIn" "CarPort" NA        "Basment" "2Types"

$GarageFinish
[1] "RFn" "Unf" "Fin" NA

$GarageQual
[1] "TA" "Fa" "Gd" NA   "Ex" "Po"

$GarageCond
[1] "TA" "Fa" NA   "Gd" "Po" "Ex"

$PavedDrive
[1] "Y" "N" "P"

$SaleType
[1] "WD"    "New"   "COD"   "ConLD" "ConLI" "CWD"   "ConLw" "Con"   "Oth"

$SaleCondition
[1] "Normal"  "Abnorml" "Partial" "AdjLand" "Alloca"  "Family"
```

```r
train_house <- train_house %>%
  select(!Utilities)
```

Must narrow down number of predictors. Backwards selection

```r
# must use regular stats::lm on for aic
house_fit_0 <- lm(SalePrice ~ 1, data = train_house)
house_fit_1 <- lm(SalePrice ~ ., data = train_house)

#step_backwards <- stats::step(house_fit_1, direction = "backward")

formula_back <- SalePrice ~ MSSubClass + MSZoning + LotArea + Street + LandContour +
    LotConfig + LandSlope + Neighborhood + Condition1 + Condition2 +
    BldgType + OverallQual + OverallCond + YearBuilt + RoofMatl +
    MasVnrType + MasVnrArea + ExterQual + BsmtQual + BsmtExposure +
    BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF + CentralAir + `1stFlrSF` +
    `2ndFlrSF` + FullBath + BedroomAbvGr + KitchenAbvGr + KitchenQual +
    Functional + GarageCars + GarageArea + GarageQual + GarageCond +
    PoolArea + SaleCondition
```

Forwards and backwards selection rendered same formula.

```r
good_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(formula_back, data = train_house)

good_fit %>%
  tidy() %>%
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-1482592.252	144816.673	-10.238	0.000
MSSubClass	-121.452	48.283	-2.515	0.012
MSZoningFV	36858.724	12477.522	2.954	0.003
MSZoningRH	23696.032	13071.608	1.813	0.070
MSZoningRL	27792.954	10839.598	2.564	0.010
MSZoningRM	23666.725	10144.963	2.333	0.020
LotArea	0.740	0.099	7.488	0.000
StreetPave	43817.181	14004.597	3.129	0.002
LandContourHLS	10420.363	5312.144	1.962	0.050

term	estimate	std.error	statistic	p.value
LandContourLow	-9108.767	6663.128	-1.367	0.172
LandContourLvl	6217.801	3909.715	1.590	0.112
LotConfigCulDSac	6758.555	3146.712	2.148	0.032
LotConfigFR2	-9126.919	4086.984	-2.233	0.026
LotConfigFR3	-12234.333	12753.867	-0.959	0.338
LotConfigInside	-1323.253	1769.888	-0.748	0.455
LandSlopeMod	5356.722	3982.009	1.345	0.179
LandSlopeSev	-37136.635	10538.559	-3.524	0.000
NeighborhoodBlueste	-4526.601	18479.349	-0.245	0.807
NeighborhoodBrDale	-3049.678	10685.152	-0.285	0.775
NeighborhoodBrkSide	-203.412	9232.263	-0.022	0.982
NeighborhoodClearCr	-9560.805	8909.902	-1.073	0.283
NeighborhoodCollgCr	-10510.658	6987.355	-1.504	0.133
NeighborhoodCrawfor	12791.502	8274.234	1.546	0.122
NeighborhoodEdwards	-22392.734	7791.235	-2.874	0.004
NeighborhoodGilbert	-11914.913	7421.236	-1.606	0.109
NeighborhoodIDOTRR	-440.771	10701.406	-0.041	0.967
NeighborhoodMeadowV	-7200.564	10537.463	-0.683	0.495
NeighborhoodMitchel	-22953.663	7989.925	-2.873	0.004
NeighborhoodNAmes	-17305.654	7467.295	-2.318	0.021
NeighborhoodNoRidge	25169.562	8027.672	3.135	0.002
NeighborhoodNPkVill	-1328.359	10382.951	-0.128	0.898
NeighborhoodNridgHt	14202.011	7254.259	1.958	0.050
NeighborhoodNWAmes	-23174.149	7604.438	-3.047	0.002
NeighborhoodOldTown	-11509.184	9426.234	-1.221	0.222
NeighborhoodSawyer	-12185.742	7882.137	-1.546	0.122
NeighborhoodSawyerW	-7597.016	7505.205	-1.012	0.312
NeighborhoodSomerst	-3566.611	8769.058	-0.407	0.684
NeighborhoodStoneBr	29851.943	7962.514	3.749	0.000
NeighborhoodSWISU	-7469.864	9591.584	-0.779	0.436
NeighborhoodTimber	-17072.875	7878.745	-2.167	0.030
NeighborhoodVeenker	890.041	9896.853	0.090	0.928
Condition1Feedr	3919.264	5325.899	0.736	0.462
Condition1Norm	12987.047	4233.013	3.068	0.002
Condition1PosA	7423.741	9798.120	0.758	0.449
Condition1PosN	13252.336	7293.420	1.817	0.069
Condition1RRAe	-11037.788	9008.455	-1.225	0.221
Condition1RRAn	10233.241	6807.547	1.503	0.133
Condition1RRNe	-3162.664	17516.499	-0.181	0.857
Condition1RRNn	12057.948	12597.470	0.957	0.339
Condition2Feedr	-20380.369	23185.007	-0.879	0.380

term	estimate	std.error	statistic	p.value
Condition2Norm	-7700.850	19725.309	-0.390	0.696
Condition2PosA	30062.637	31663.487	0.949	0.343
Condition2PosN	-247442.347	27031.739	-9.154	0.000
Condition2RRAe	-21803.583	31288.179	-0.697	0.486
Condition2RRAn	2093.760	31126.857	0.067	0.946
Condition2RRNn	-1345.953	26755.383	-0.050	0.960
BldgType2fmCon	8513.263	9307.234	0.915	0.361
BldgTypeDuplex	-9980.772	7793.713	-1.281	0.201
BldgTypeTwnhs	-12952.255	7274.440	-1.781	0.075
BldgTypeTwnhsE	-7935.286	5804.736	-1.367	0.172
OverallQual	7184.678	1010.772	7.108	0.000
OverallCond	6051.171	779.738	7.761	0.000
YearBuilt	387.873	65.964	5.880	0.000
RoofMatlCompShg	678165.770	30437.111	22.281	0.000
RoofMatlMembran	748415.521	42060.851	17.794	0.000
RoofMatlMetal	720454.243	40822.621	17.648	0.000
RoofMatlRoll	681632.812	39300.596	17.344	0.000
RoofMatlTar&Grv	665833.556	31911.372	20.865	0.000
RoofMatlWdShake	685721.532	32699.898	20.970	0.000
RoofMatlWdShngl	722778.401	31580.038	22.887	0.000
MasVnrTypeBrkFace	10025.819	6735.796	1.488	0.137
MasVnrTypeNone	14560.294	6799.392	2.141	0.032
MasVnrTypeStone	15117.738	7125.356	2.122	0.034
MasVnrArea	18.336	5.684	3.226	0.001
ExterQualFa	-9561.408	12558.026	-0.761	0.447
ExterQualGd	-21362.882	4689.787	-4.555	0.000
ExterQualTA	-23318.052	5197.306	-4.487	0.000
BsmtQualFa	-15429.132	6253.032	-2.467	0.014
BsmtQualGd	-21033.244	3237.635	-6.496	0.000
BsmtQualTA	-18737.100	3991.278	-4.695	0.000
BsmtExposureGd	16168.102	3030.470	5.335	0.000
BsmtExposureMn	-4160.680	3008.057	-1.383	0.167
BsmtExposureNo	-6390.801	2131.948	-2.998	0.003
BsmtFinSF1	37.815	4.652	8.129	0.000
BsmtFinSF2	26.126	5.902	4.427	0.000
BsmtUnfSF	18.831	4.595	4.099	0.000
CentralAirY	4360.455	3920.833	1.112	0.266
`1stFlrSF`	57.951	4.838	11.978	0.000
`2ndFlrSF`	64.161	3.130	20.497	0.000
FullBath	2157.120	2071.933	1.041	0.298
BedroomAbvGr	-4295.958	1291.501	-3.326	0.001

term	estimate	std.error	statistic	p.value
KitchenAbvGr	-17929.756	6610.656	-2.712	0.007
KitchenQualFa	-24531.098	6556.319	-3.742	0.000
KitchenQualGd	-25441.474	3459.180	-7.355	0.000
KitchenQualTA	-25032.152	3888.797	-6.437	0.000
FunctionalMaj2	-8796.976	15020.561	-0.586	0.558
FunctionalMin1	-2509.398	9177.460	-0.273	0.785
FunctionalMin2	-1724.791	9145.001	-0.189	0.850
FunctionalMod	-7120.673	11035.893	-0.645	0.519
FunctionalSev	-52618.288	27457.533	-1.916	0.056
FunctionalTyp	9563.732	8049.176	1.188	0.235
GarageCars	5438.394	2257.094	2.409	0.016
GarageArea	13.793	7.470	1.846	0.065
GarageQualFa	-125024.498	27138.290	-4.607	0.000
GarageQualGd	-117485.100	27811.992	-4.224	0.000
GarageQualPo	-132336.920	33433.362	-3.958	0.000
GarageQualTA	-123573.264	26832.297	-4.605	0.000
GarageCondFa	117433.852	32160.075	3.652	0.000
GarageCondGd	111518.001	32892.560	3.390	0.001
GarageCondPo	123138.681	34550.330	3.564	0.000
GarageCondTA	120652.600	31721.013	3.804	0.000
PoolArea	88.697	17.441	5.086	0.000
SaleConditionAdjLand	30320.103	23992.531	1.264	0.207
SaleConditionAlloca	14738.357	10550.543	1.397	0.163
SaleConditionFamily	-408.372	6028.654	-0.068	0.946
SaleConditionNormal	5413.079	2791.561	1.939	0.053
SaleConditionPartial	17751.503	3868.309	4.589	0.000

Building a recipe

```
house_rec_1 <- recipe(formula_back, data = train_house) %>%
step_center(all_numeric_predictors()) %>% # mean center
step_dummy(all_nominal_predictors()) %>% # dummy coding
step_zv(all_predictors())# remove zero variance variables
```

```
house_spec <- linear_reg() %>%
  set_engine("lm")

house_wflow_1 <- workflow() %>%
```

```
add_model(house_spec) %>%
add_recipe(house_rec_1)
```