Residential home prices

Loading the packages

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

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

```
library(janitor)
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(ggridges)
library(rmarkdown)
```

Hypothesis

What are all factors could influence the residential home prices.

Condition - 1

Type of neighbourhood preferred. # sub-conditions Type of zone preferred Type of house style Type of building type Type of facilities Shape of the property flatness level Foundation type Type of heating Electrical system Kitchen quality Garage location Interior finish of the garage paved driveway Type of sale

Reading the data, cleaning the column names.

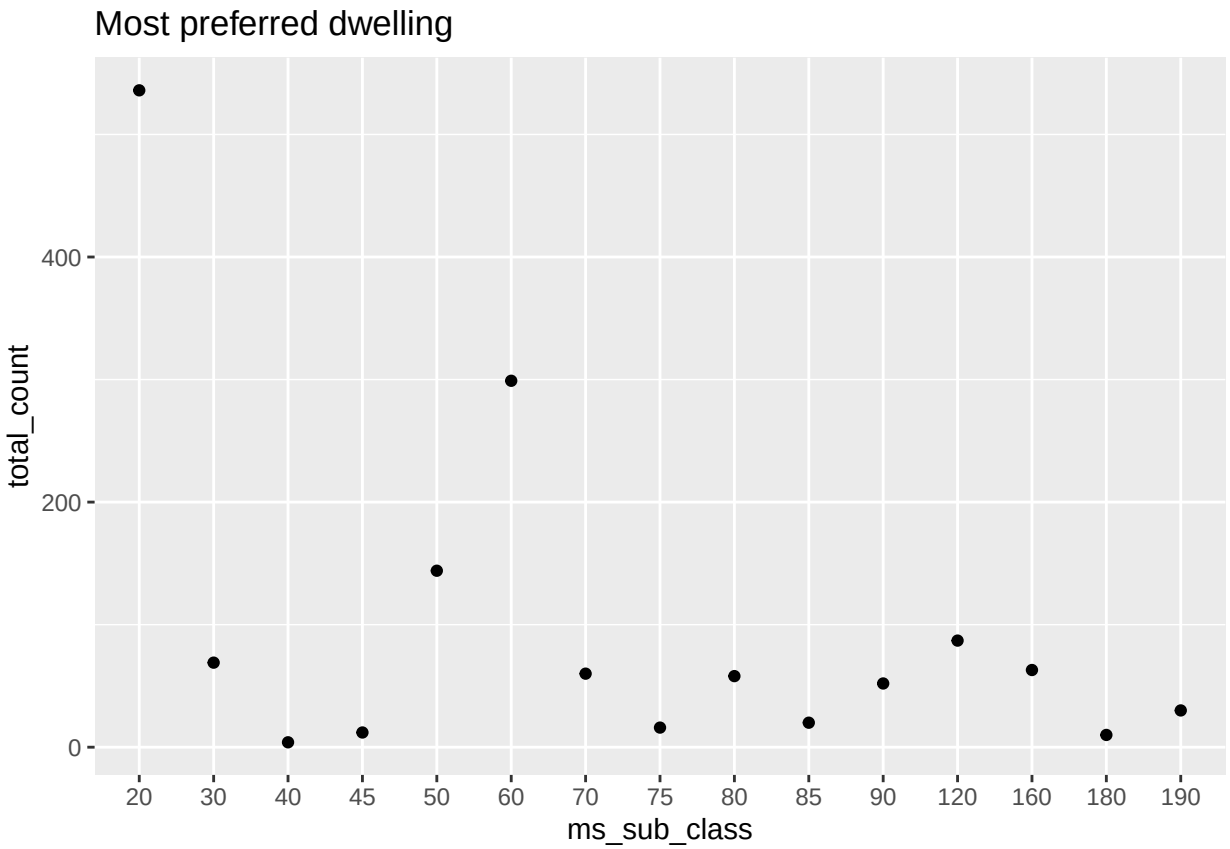
```
data <- read_csv("train.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   Id = col_double(),
##   MSSubClass = col_double(),
##   LotFrontage = col_double(),
##   LotArea = col_double(),
##   OverallQual = col_double(),
##   OverallCond = col_double(),
##   YearBuilt = col_double(),
##   YearRemodAdd = col_double(),
##   MasVnrArea = col_double(),
##   BsmtFinSF1 = col_double(),
##   BsmtFinSF2 = col_double(),
##   BsmtUnfSF = col_double(),
##   TotalBsmtSF = col_double(),
##   `1stFlrSF` = col_double(),
##   `2ndFlrSF` = col_double(),
##   LowQualFinSF = col_double(),
##   GrLivArea = col_double(),
##   BsmtFullBath = col_double(),
##   BsmtHalfBath = col_double(),
##   FullBath = col_double()
##   # ... with 18 more columns
## )
## i Use `spec()` for the full column specifications.
```

```
data <- clean_names(data)
data$ms_sub_class <- as.factor(data$ms_sub_class)
```

A general look at the type of dwelling and the most preferred dwelling.

```
data %>%
  select(ms_sub_class) %>%
  group_by(ms_sub_class) %>%
  summarise(total_count = n()) %>%
ggplot(aes(x = (ms_sub_class),y = total_count)) +
  geom_point(fill = "tomato3") +
  labs(title = "Most preferred dwelling") +
  xlab("ms_sub_class")
```



Looking for missing values

```
colSums(is.na(data))
```

```
##              id     ms_sub_class         ms_zoning    lot_frontage        lot_area
##               0                0                 0             259               0
##          street            alley       lot_shape    land_contour       utilities
##               0             1369                 0               0               0
##      lot_config       land_slope    neighborhood      condition1      condition2
##               0                0                 0               0               0
##       bldg_type      house_style    overall_qual    overall_cond      year_built
##               0                0                 0               0               0
##  year_remod_add       roof_style       roof_matl     exterior1st     exterior2nd
##               0                0                 0               0               0
##    mas_vnr_type     mas_vnr_area      exter_qual      exter_cond      foundation
##               8                8                 0               0               0
##       bsmt_qual       bsmt_cond   bsmt_exposure   bsmt_fin_type1    bsmt_fin_sf1
##              37               37                38              37               0
##  bsmt_fin_type2    bsmt_fin_sf2     bsmt_unf_sf   total_bsmt_sf         heating
##              38                0                 0               0               0
##      heating_qc      central_air      electrical     x1st_flr_sf     x2nd_flr_sf
##               0                0                 1               0               0
## low_qual_fin_sf     gr_liv_area   bsmt_full_bath   bsmt_half_bath       full_bath
##               0                0                 0               0               0
```

```
##       half_bath  bedroom_abv_gr  kitchen_abv_gr    kitchen_qual tot_rms_abv_grd
##               0               0               0               0               0
##      functional      fireplaces    fireplace_qu     garage_type   garage_yr_blt
##               0               0             690              81              81
##   garage_finish     garage_cars     garage_area     garage_qual     garage_cond
##              81               0               0              81              81
##     paved_drive    wood_deck_sf   open_porch_sf  enclosed_porch     x3ssn_porch
##               0               0               0               0               0
##    screen_porch       pool_area         pool_qc           fence    misc_feature
##               0               0            1453            1179            1406
##        misc_val         mo_sold         yr_sold       sale_type  sale_condition
##               0               0               0               0               0
##      sale_price
##               0
```

```
str(data)
```

```
## spec_tbl_df[,81] [1,460 x 81] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ id             : num [1:1460] 1 2 3 4 5 6 7 8 9 10 ...
##  $ ms_sub_class   : Factor w/ 15 levels "20","30","40",..: 6 1 6 7 6 5 1 6 5 15 ...
##  $ ms_zoning      : chr [1:1460] "RL" "RL" "RL" "RL" ...
##  $ lot_frontage   : num [1:1460] 65 80 68 60 84 85 75 NA 51 50 ...
##  $ lot_area       : num [1:1460] 8450 9600 11250 9550 14260 ...
##  $ street         : chr [1:1460] "Pave" "Pave" "Pave" "Pave" ...
##  $ alley          : chr [1:1460] NA NA NA NA ...
##  $ lot_shape      : chr [1:1460] "Reg" "Reg" "IR1" "IR1" ...
##  $ land_contour   : chr [1:1460] "Lvl" "Lvl" "Lvl" "Lvl" ...
##  $ utilities      : chr [1:1460] "AllPub" "AllPub" "AllPub" "AllPub" ...
##  $ lot_config     : chr [1:1460] "Inside" "FR2" "Inside" "Corner" ...
##  $ land_slope     : chr [1:1460] "Gtl" "Gtl" "Gtl" "Gtl" ...
##  $ neighborhood   : chr [1:1460] "CollgCr" "Veenker" "CollgCr" "Crawfor" ...
##  $ condition1     : chr [1:1460] "Norm" "Feedr" "Norm" "Norm" ...
##  $ condition2     : chr [1:1460] "Norm" "Norm" "Norm" "Norm" ...
##  $ bldg_type      : chr [1:1460] "1Fam" "1Fam" "1Fam" "1Fam" ...
##  $ house_style    : chr [1:1460] "2Story" "1Story" "2Story" "2Story" ...
##  $ overall_qual   : num [1:1460] 7 6 7 7 8 5 8 7 7 5 ...
##  $ overall_cond   : num [1:1460] 5 8 5 5 5 5 5 6 5 6 ...
##  $ year_built     : num [1:1460] 2003 1976 2001 1915 2000 ...
##  $ year_remod_add : num [1:1460] 2003 1976 2002 1970 2000 ...
##  $ roof_style     : chr [1:1460] "Gable" "Gable" "Gable" "Gable" ...
##  $ roof_matl      : chr [1:1460] "CompShg" "CompShg" "CompShg" "CompShg" ...
##  $ exterior1st    : chr [1:1460] "VinylSd" "MetalSd" "VinylSd" "Wd Sdng" ...
##  $ exterior2nd    : chr [1:1460] "VinylSd" "MetalSd" "VinylSd" "Wd Shng" ...
##  $ mas_vnr_type   : chr [1:1460] "BrkFace" "None" "BrkFace" "None" ...
##  $ mas_vnr_area   : num [1:1460] 196 0 162 0 350 0 186 240 0 0 ...
##  $ exter_qual     : chr [1:1460] "Gd" "TA" "Gd" "TA" ...
##  $ exter_cond     : chr [1:1460] "TA" "TA" "TA" "TA" ...
##  $ foundation     : chr [1:1460] "PConc" "CBlock" "PConc" "BrkTil" ...
##  $ bsmt_qual      : chr [1:1460] "Gd" "Gd" "Gd" "TA" ...
##  $ bsmt_cond      : chr [1:1460] "TA" "TA" "TA" "Gd" ...
##  $ bsmt_exposure  : chr [1:1460] "No" "Gd" "Mn" "No" ...
##  $ bsmt_fin_type1 : chr [1:1460] "GLQ" "ALQ" "GLQ" "ALQ" ...
##  $ bsmt_fin_sf1   : num [1:1460] 706 978 486 216 655 ...
##  $ bsmt_fin_type2 : chr [1:1460] "Unf" "Unf" "Unf" "Unf" ...
##  $ bsmt_fin_sf2   : num [1:1460] 0 0 0 0 0 0 0 32 0 0 ...
```

```
##  $ bsmt_unf_sf    : num [1:1460] 150 284 434 540 490 64 317 216 952 140 ...
##  $ total_bsmt_sf  : num [1:1460] 856 1262 920 756 1145 ...
##  $ heating        : chr [1:1460] "GasA" "GasA" "GasA" "GasA" ...
##  $ heating_qc     : chr [1:1460] "Ex" "Ex" "Ex" "Gd" ...
##  $ central_air    : chr [1:1460] "Y" "Y" "Y" "Y" ...
##  $ electrical     : chr [1:1460] "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...
##  $ x1st_flr_sf    : num [1:1460] 856 1262 920 961 1145 ...
##  $ x2nd_flr_sf    : num [1:1460] 854 0 866 756 1053 ...
##  $ low_qual_fin_sf: num [1:1460] 0 0 0 0 0 0 0 0 0 0 ...
##  $ gr_liv_area    : num [1:1460] 1710 1262 1786 1717 2198 ...
##  $ bsmt_full_bath : num [1:1460] 1 0 1 1 1 1 1 1 0 1 ...
##  $ bsmt_half_bath : num [1:1460] 0 1 0 0 0 0 0 0 0 0 ...
##  $ full_bath      : num [1:1460] 2 2 2 1 2 1 2 2 2 1 ...
##  $ half_bath      : num [1:1460] 1 0 1 0 1 1 0 1 0 0 ...
##  $ bedroom_abv_gr : num [1:1460] 3 3 3 3 4 1 3 3 2 2 ...
##  $ kitchen_abv_gr : num [1:1460] 1 1 1 1 1 1 1 1 2 2 ...
##  $ kitchen_qual   : chr [1:1460] "Gd" "TA" "Gd" "Gd" ...
##  $ tot_rms_abv_grd: num [1:1460] 8 6 6 7 9 5 7 7 8 5 ...
##  $ functional     : chr [1:1460] "Typ" "Typ" "Typ" "Typ" ...
##  $ fireplaces     : num [1:1460] 0 1 1 1 1 0 1 2 2 2 ...
##  $ fireplace_qu   : chr [1:1460] NA "TA" "TA" "Gd" ...
##  $ garage_type    : chr [1:1460] "Attchd" "Attchd" "Attchd" "Detchd" ...
##  $ garage_yr_blt  : num [1:1460] 2003 1976 2001 1998 2000 ...
##  $ garage_finish  : chr [1:1460] "RFn" "RFn" "RFn" "Unf" ...
##  $ garage_cars    : num [1:1460] 2 2 2 3 3 2 2 2 2 1 ...
##  $ garage_area    : num [1:1460] 548 460 608 642 836 480 636 484 468 205 ...
##  $ garage_qual    : chr [1:1460] "TA" "TA" "TA" "TA" ...
##  $ garage_cond    : chr [1:1460] "TA" "TA" "TA" "TA" ...
##  $ paved_drive    : chr [1:1460] "Y" "Y" "Y" "Y" ...
##  $ wood_deck_sf   : num [1:1460] 0 298 0 0 192 40 255 235 90 0 ...
##  $ open_porch_sf  : num [1:1460] 61 0 42 35 84 30 57 204 0 4 ...
##  $ enclosed_porch : num [1:1460] 0 0 0 272 0 0 0 228 205 0 ...
##  $ x3ssn_porch    : num [1:1460] 0 0 0 0 0 320 0 0 0 0 ...
##  $ screen_porch   : num [1:1460] 0 0 0 0 0 0 0 0 0 0 ...
##  $ pool_area      : num [1:1460] 0 0 0 0 0 0 0 0 0 0 ...
##  $ pool_qc        : chr [1:1460] NA NA NA NA ...
##  $ fence          : chr [1:1460] NA NA NA NA ...
##  $ misc_feature   : chr [1:1460] NA NA NA NA ...
##  $ misc_val       : num [1:1460] 0 0 0 0 0 700 0 350 0 0 ...
##  $ mo_sold        : num [1:1460] 2 5 9 2 12 10 8 11 4 1 ...
##  $ yr_sold        : num [1:1460] 2008 2007 2008 2006 2008 ...
##  $ sale_type      : chr [1:1460] "WD" "WD" "WD" "WD" ...
##  $ sale_condition : chr [1:1460] "Normal" "Normal" "Normal" "Abnorml" ...
##  $ sale_price     : num [1:1460] 208500 181500 223500 140000 250000 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   MSSubClass = col_double(),
##   ..   MSZoning = col_character(),
##   ..   LotFrontage = col_double(),
##   ..   LotArea = col_double(),
##   ..   Street = col_character(),
##   ..   Alley = col_character(),
##   ..   LotShape = col_character(),
```

```
##   ..   LandContour = col_character(),
##   ..   Utilities = col_character(),
##   ..   LotConfig = col_character(),
##   ..   LandSlope = col_character(),
##   ..   Neighborhood = col_character(),
##   ..   Condition1 = col_character(),
##   ..   Condition2 = col_character(),
##   ..   BldgType = col_character(),
##   ..   HouseStyle = col_character(),
##   ..   OverallQual = col_double(),
##   ..   OverallCond = col_double(),
##   ..   YearBuilt = col_double(),
##   ..   YearRemodAdd = col_double(),
##   ..   RoofStyle = col_character(),
##   ..   RoofMatl = col_character(),
##   ..   Exterior1st = col_character(),
##   ..   Exterior2nd = col_character(),
##   ..   MasVnrType = col_character(),
##   ..   MasVnrArea = col_double(),
##   ..   ExterQual = col_character(),
##   ..   ExterCond = col_character(),
##   ..   Foundation = col_character(),
##   ..   BsmtQual = col_character(),
##   ..   BsmtCond = col_character(),
##   ..   BsmtExposure = col_character(),
##   ..   BsmtFinType1 = col_character(),
##   ..   BsmtFinSF1 = col_double(),
##   ..   BsmtFinType2 = col_character(),
##   ..   BsmtFinSF2 = col_double(),
##   ..   BsmtUnfSF = col_double(),
##   ..   TotalBsmtSF = col_double(),
##   ..   Heating = col_character(),
##   ..   HeatingQC = col_character(),
##   ..   CentralAir = col_character(),
##   ..   Electrical = col_character(),
##   ..   `1stFlrSF` = col_double(),
##   ..   `2ndFlrSF` = col_double(),
##   ..   LowQualFinSF = col_double(),
##   ..   GrLivArea = col_double(),
##   ..   BsmtFullBath = col_double(),
##   ..   BsmtHalfBath = col_double(),
##   ..   FullBath = col_double(),
##   ..   HalfBath = col_double(),
##   ..   BedroomAbvGr = col_double(),
##   ..   KitchenAbvGr = col_double(),
##   ..   KitchenQual = col_character(),
##   ..   TotRmsAbvGrd = col_double(),
##   ..   Functional = col_character(),
##   ..   Fireplaces = col_double(),
##   ..   FireplaceQu = col_character(),
##   ..   GarageType = col_character(),
##   ..   GarageYrBlt = col_double(),
##   ..   GarageFinish = col_character(),
##   ..   GarageCars = col_double(),
```

```
##   ..   GarageArea = col_double(),
##   ..   GarageQual = col_character(),
##   ..   GarageCond = col_character(),
##   ..   PavedDrive = col_character(),
##   ..   WoodDeckSF = col_double(),
##   ..   OpenPorchSF = col_double(),
##   ..   EnclosedPorch = col_double(),
##   ..   `3SsnPorch` = col_double(),
##   ..   ScreenPorch = col_double(),
##   ..   PoolArea = col_double(),
##   ..   PoolQC = col_character(),
##   ..   Fence = col_character(),
##   ..   MiscFeature = col_character(),
##   ..   MiscVal = col_double(),
##   ..   MoSold = col_double(),
##   ..   YrSold = col_double(),
##   ..   SaleType = col_character(),
##   ..   SaleCondition = col_character(),
##   ..   SalePrice = col_double()
##   .. )
```

Removing un-necessary columns.

```
data$id <- NULL
data$pool_qc <- NULL
data$fence <- NULL
data$misc_feature <- NULL
data$alley <- NULL
```

```
#cols <- c("ms_zoning", "street", "lot_shape", "land_contour", "utilities", "lot_config",
#          "land_slope", "neighbourhood", "condition1", "condition2", "bldg_type", "house_style",
#          "overall_qual", "overall_cond", "roof_style", "roof_matl", "exterior1st", "exterior2nd",
#          "mas_vnr_type", "exter_qual", "exter_cond", "foundation", "bsmt_qual", "bsmt_cond",
#         "bsmt_exposure", "bsmt_fin_type1", "bsmt_fin_type2", "heating", "heating_qc", "central_air",
#         "electrical", "kitchen_qual", "functional", "fireplace_qu", "garage_type", "garage_finish",
#          "garage_qual", "garage_cond", "paved_drive", "sale_type", "sale_condition")
```

Converting columns to factors.

```
x <- data %>%
  select_if(is.character)

length(colnames(x))
```

```
## [1] 39
```

```
y <- data%>%
  select_if(is.numeric)

colnames(y)
```

```
##  [1] "lot_frontage"    "lot_area"        "overall_qual"    "overall_cond"
##  [5] "year_built"      "year_remod_add"  "mas_vnr_area"    "bsmt_fin_sf1"
##  [9] "bsmt_fin_sf2"    "bsmt_unf_sf"     "total_bsmt_sf"   "x1st_flr_sf"
## [13] "x2nd_flr_sf"     "low_qual_fin_sf" "gr_liv_area"     "bsmt_full_bath"
## [17] "bsmt_half_bath"  "full_bath"       "half_bath"       "bedroom_abv_gr"
## [21] "kitchen_abv_gr"  "tot_rms_abv_grd" "fireplaces"      "garage_yr_blt"
## [25] "garage_cars"     "garage_area"     "wood_deck_sf"    "open_porch_sf"
## [29] "enclosed_porch"  "x3ssn_porch"     "screen_porch"    "pool_area"
## [33] "misc_val"        "mo_sold"         "yr_sold"         "sale_price"
```

```
col_as_factors <- c(colnames(x), "overall_qual", "overall_cond")

sum(length(col_as_factors))
```

```
## [1] 41
```

```
data[col_as_factors] <- lapply(data[col_as_factors], factor)
#str(data)
glimpse(data)
```

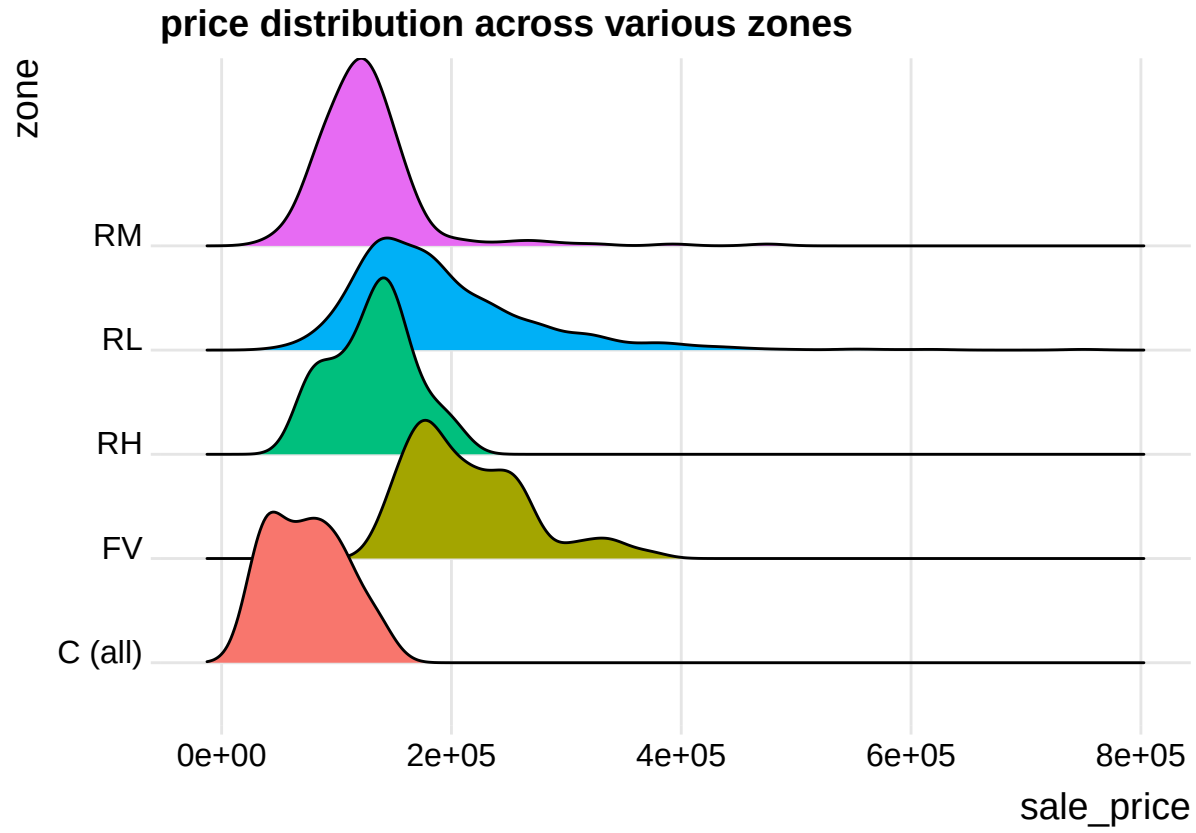
```
## Rows: 1,460
## Columns: 76
## $ ms_sub_class    <fct> 60, 20, 60, 70, 60, 50, 20, 60, 50, 190, 20, 60, 20, 2~
## $ ms_zoning       <fct> RL, RL, RL, RL, RL, RL, RL, RL, RM, RL, RL, RL, RL, RL~
## $ lot_frontage    <dbl> 65, 80, 68, 60, 84, 85, 75, NA, 51, 50, 70, 85, NA, 91~
## $ lot_area        <dbl> 8450, 9600, 11250, 9550, 14260, 14115, 10084, 10382, 6~
## $ street          <fct> Pave, Pave, Pave, Pave, Pave, Pave, Pave, Pave, Pave, ~
## $ lot_shape       <fct> Reg, Reg, IR1, IR1, IR1, IR1, Reg, IR1, Reg, Reg, Reg,~
## $ land_contour    <fct> Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl,~
## $ utilities       <fct> AllPub, AllPub, AllPub, AllPub, AllPub, AllPub, AllPub~
## $ lot_config      <fct> Inside, FR2, Inside, Corner, FR2, Inside, Inside, Corn~
## $ land_slope      <fct> Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl,~
## $ neighborhood    <fct> CollgCr, Veenker, CollgCr, Crawfor, NoRidge, Mitchel, ~
## $ condition1      <fct> Norm, Feedr, Norm, Norm, Norm, Norm, Norm, PosN, Arter~
## $ condition2      <fct> Norm, Norm, Norm, Norm, Norm, Norm, Norm, Norm, Norm, ~
## $ bldg_type       <fct> 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, ~
## $ house_style     <fct> 2Story, 1Story, 2Story, 2Story, 2Story, 1.5Fin, 1Story~
## $ overall_qual    <fct> 7, 6, 7, 7, 8, 5, 8, 7, 7, 5, 5, 9, 5, 7, 6, 7, 6, 4, ~
## $ overall_cond    <fct> 5, 8, 5, 5, 5, 5, 5, 6, 5, 6, 5, 5, 6, 5, 5, 8, 7, 5, ~
## $ year_built      <dbl> 2003, 1976, 2001, 1915, 2000, 1993, 2004, 1973, 1931, ~
## $ year_remod_add  <dbl> 2003, 1976, 2002, 1970, 2000, 1995, 2005, 1973, 1950, ~
## $ roof_style      <fct> Gable, Gable, Gable, Gable, Gable, Gable, Gable, Gable~
## $ roof_matl       <fct> CompShg, CompShg, CompShg, CompShg, CompShg, CompShg, ~
## $ exterior1st     <fct> VinylSd, MetalSd, VinylSd, Wd Sdng, VinylSd, VinylSd, ~
## $ exterior2nd     <fct> VinylSd, MetalSd, VinylSd, Wd Shng, VinylSd, VinylSd, ~
## $ mas_vnr_type    <fct> BrkFace, None, BrkFace, None, BrkFace, None, Stone, St~
## $ mas_vnr_area    <dbl> 196, 0, 162, 0, 350, 0, 186, 240, 0, 0, 0, 286, 0, 306~
## $ exter_qual      <fct> Gd, TA, Gd, TA, Gd, TA, Gd, TA, TA, TA, TA, Ex, TA, Gd~
## $ exter_cond      <fct> TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA~
## $ foundation      <fct> PConc, CBlock, PConc, BrkTil, PConc, Wood, PConc, CBlo~
## $ bsmt_qual       <fct> Gd, Gd, Gd, TA, Gd, Gd, Ex, Gd, TA, TA, TA, Ex, TA, Gd~
## $ bsmt_cond       <fct> TA, TA, TA, Gd, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA~
## $ bsmt_exposure   <fct> No, Gd, Mn, No, Av, No, Av, Mn, No, No, No, No, No, Av~
## $ bsmt_fin_type1  <fct> GLQ, ALQ, GLQ, ALQ, GLQ, GLQ, GLQ, ALQ, Unf, GLQ, Rec,~
## $ bsmt_fin_sf1    <dbl> 706, 978, 486, 216, 655, 732, 1369, 859, 0, 851, 906, ~
```

```
## $ bsmt_fin_type2  <fct> Unf, Unf, Unf, Unf, Unf, Unf, Unf, BLQ, Unf, Unf, Unf,~
## $ bsmt_fin_sf2    <dbl> 0, 0, 0, 0, 0, 0, 0, 32, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ bsmt_unf_sf     <dbl> 150, 284, 434, 540, 490, 64, 317, 216, 952, 140, 134, ~
## $ total_bsmt_sf   <dbl> 856, 1262, 920, 756, 1145, 796, 1686, 1107, 952, 991, ~
## $ heating         <fct> GasA, GasA, GasA, GasA, GasA, GasA, GasA, GasA, GasA, ~
## $ heating_qc      <fct> Ex, Ex, Ex, Gd, Ex, Ex, Ex, Ex, Gd, Ex, Ex, Ex, TA, Ex~
## $ central_air     <fct> Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, ~
## $ electrical      <fct> SBrkr, SBrkr, SBrkr, SBrkr, SBrkr, SBrkr, SBrkr, SBrkr~
## $ x1st_flr_sf     <dbl> 856, 1262, 920, 961, 1145, 796, 1694, 1107, 1022, 1077~
## $ x2nd_flr_sf     <dbl> 854, 0, 866, 756, 1053, 566, 0, 983, 752, 0, 0, 1142, ~
## $ low_qual_fin_sf <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ gr_liv_area     <dbl> 1710, 1262, 1786, 1717, 2198, 1362, 1694, 2090, 1774, ~
## $ bsmt_full_bath  <dbl> 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, ~
## $ bsmt_half_bath  <dbl> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ full_bath       <dbl> 2, 2, 2, 1, 2, 1, 2, 2, 2, 1, 1, 3, 1, 2, 1, 1, 1, 2, ~
## $ half_bath       <dbl> 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, ~
## $ bedroom_abv_gr  <dbl> 3, 3, 3, 3, 4, 1, 3, 3, 2, 2, 3, 4, 2, 3, 2, 2, 2, 2, ~
## $ kitchen_abv_gr  <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, ~
## $ kitchen_qual    <fct> Gd, TA, Gd, Gd, Gd, TA, Gd, TA, TA, TA, TA, Ex, TA, Gd~
## $ tot_rms_abv_grd <dbl> 8, 6, 6, 7, 9, 5, 7, 7, 8, 5, 5, 11, 4, 7, 5, 5, 5, 6,~
## $ functional      <fct> Typ, Typ, Typ, Typ, Typ, Typ, Typ, Typ, Min1, Typ, Typ~
## $ fireplaces      <dbl> 0, 1, 1, 1, 1, 0, 1, 2, 2, 2, 0, 2, 0, 1, 1, 0, 1, 0, ~
## $ fireplace_qu    <fct> NA, TA, TA, Gd, TA, NA, Gd, TA, TA, TA, NA, Gd, NA, Gd~
## $ garage_type     <fct> Attchd, Attchd, Attchd, Detchd, Attchd, Attchd, Attchd~
## $ garage_yr_blt   <dbl> 2003, 1976, 2001, 1998, 2000, 1993, 2004, 1973, 1931, ~
## $ garage_finish   <fct> RFn, RFn, RFn, Unf, RFn, Unf, RFn, RFn, Unf, RFn, Unf,~
## $ garage_cars     <dbl> 2, 2, 2, 3, 3, 2, 2, 2, 2, 1, 1, 3, 1, 3, 1, 2, 2, 2, ~
## $ garage_area     <dbl> 548, 460, 608, 642, 836, 480, 636, 484, 468, 205, 384,~
## $ garage_qual     <fct> TA, TA, TA, TA, TA, TA, TA, TA, Fa, Gd, TA, TA, TA, TA~
## $ garage_cond     <fct> TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA~
## $ paved_drive     <fct> Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, ~
## $ wood_deck_sf    <dbl> 0, 298, 0, 0, 192, 40, 255, 235, 90, 0, 0, 147, 140, 1~
## $ open_porch_sf   <dbl> 61, 0, 42, 35, 84, 30, 57, 204, 0, 4, 0, 21, 0, 33, 21~
## $ enclosed_porch  <dbl> 0, 0, 0, 272, 0, 0, 0, 228, 205, 0, 0, 0, 0, 0, 176, 0~
## $ x3ssn_porch     <dbl> 0, 0, 0, 0, 0, 320, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ screen_porch    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 176, 0, 0, 0, 0, 0~
## $ pool_area       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ misc_val        <dbl> 0, 0, 0, 0, 0, 700, 0, 350, 0, 0, 0, 0, 0, 0, 0, 0, 70~
## $ mo_sold         <dbl> 2, 5, 9, 2, 12, 10, 8, 11, 4, 1, 2, 7, 9, 8, 5, 7, 3, ~
## $ yr_sold         <dbl> 2008, 2007, 2008, 2006, 2008, 2009, 2007, 2009, 2008, ~
## $ sale_type       <fct> WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, New, WD, N~
## $ sale_condition  <fct> Normal, Normal, Normal, Abnorml, Normal, Normal, Norma~
## $ sale_price      <dbl> 208500, 181500, 223500, 140000, 250000, 143000, 307000~
```

price distribution for different zones

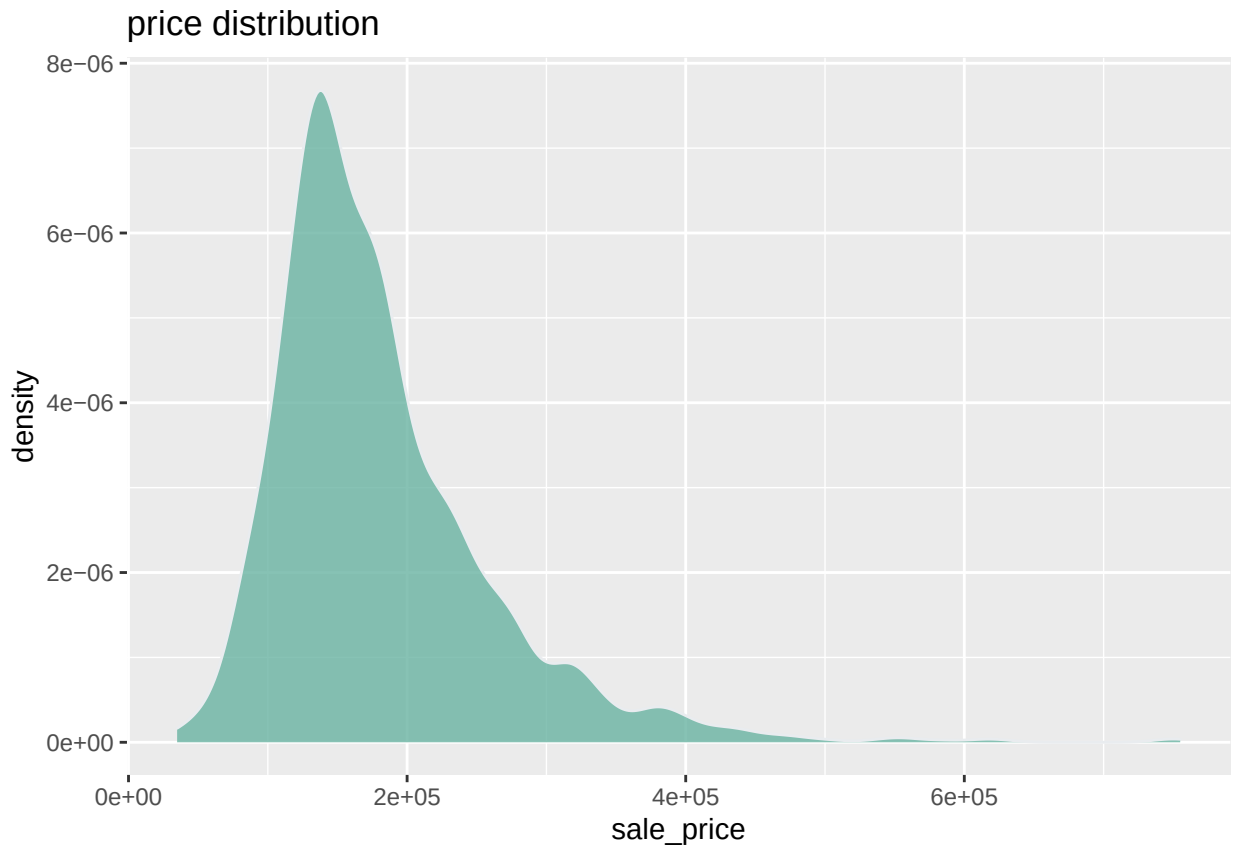
```
ggplot(data, aes(x = sale_price, y = ms_zoning, fill = ms_zoning)) +
  geom_density_ridges() +
  theme_ridges() +
  theme(legend.position = "none") +
  labs(title = "price distribution across various zones") +
  xlab("sale_price") + ylab("zone")
```

```
## Picking joint bandwidth of 15900
```



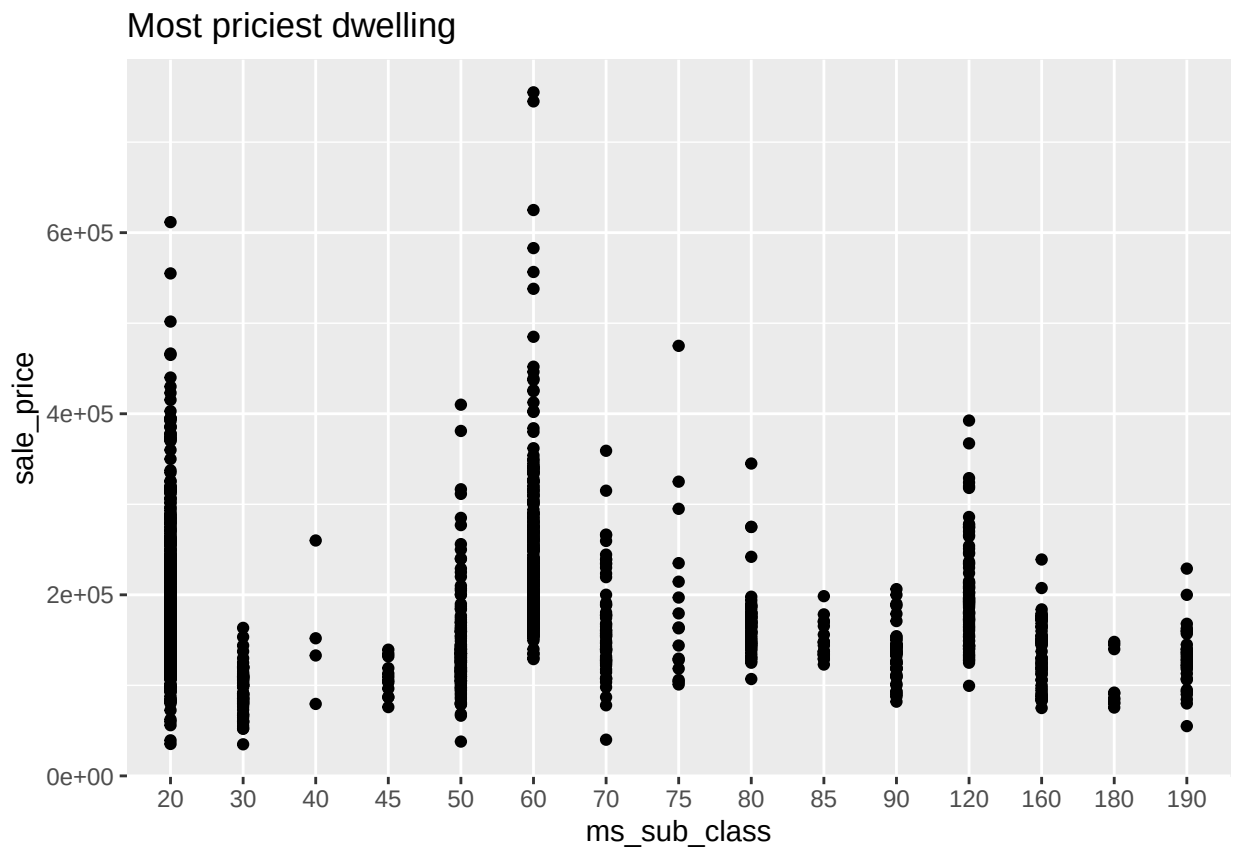
price distribution

```
data %>%
  ggplot( aes(x=sale_price)) +
  geom_density(fill="#69b3a2", color="#e9ecef", alpha=0.8) +
  labs(title = "price distribution")
```



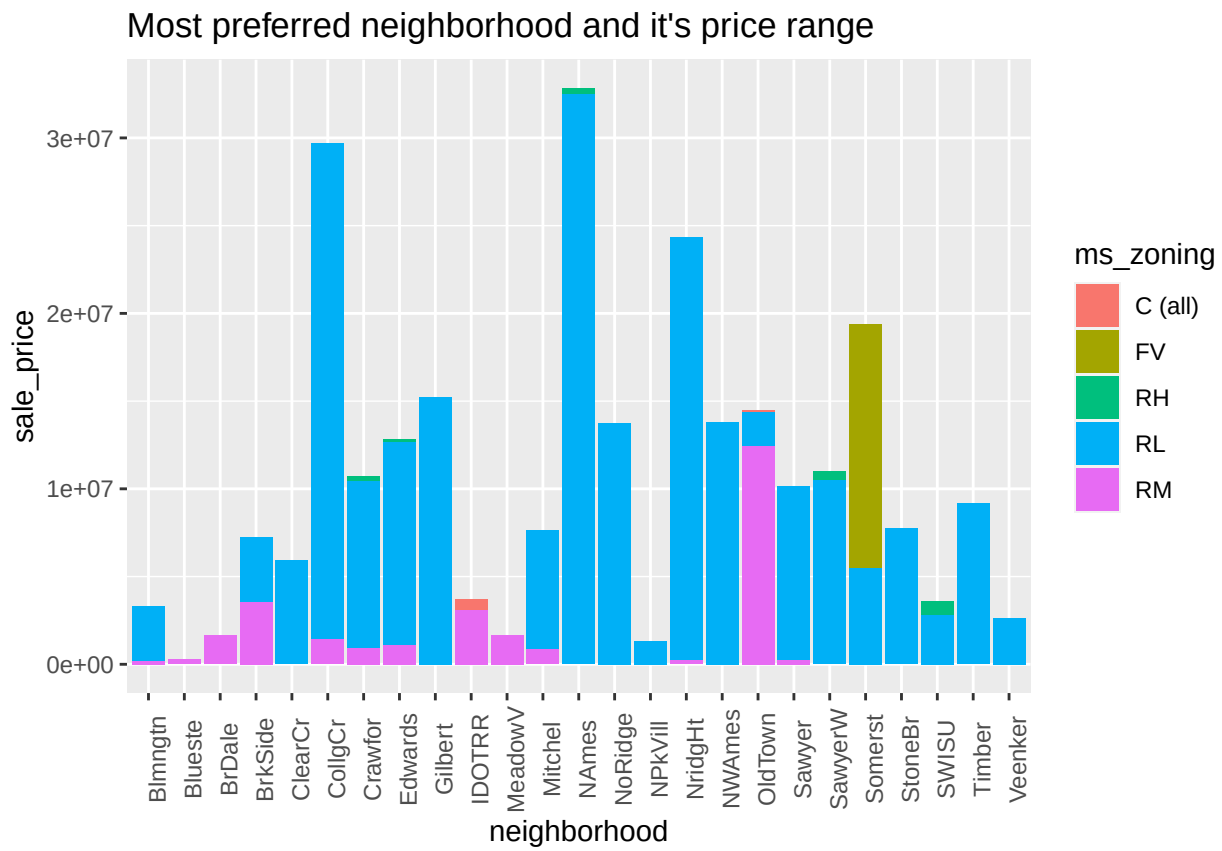
price of different dwellings

```
ggplot(data, aes(x=ms_sub_class, y=sale_price)) +
  geom_point() +
  theme(legend.position="none") +
  labs(title = "Most priciest dwelling")
```



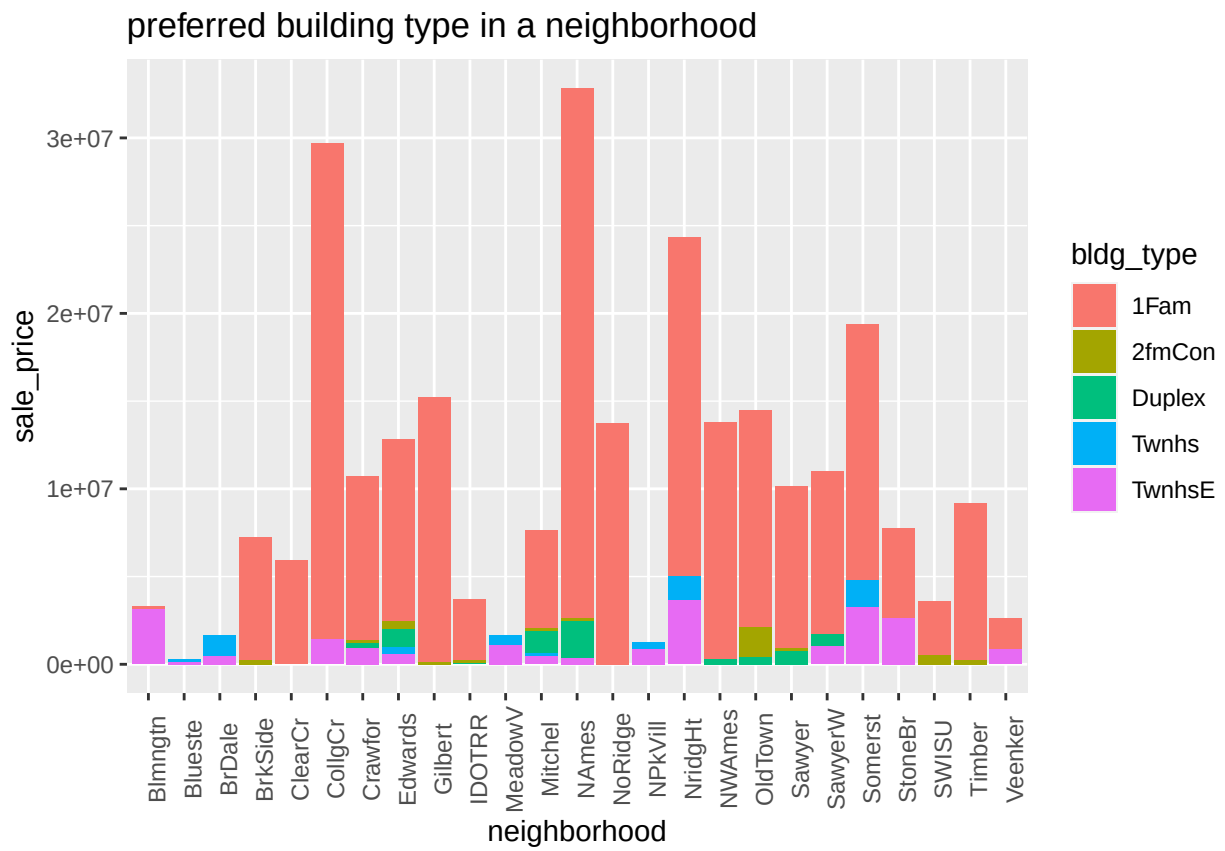
In a given neighborhood which zone is most preferred.

```
ggplot(data, aes(fill=ms_zoning, y=sale_price, x=neighborhood)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Most preferred neighborhood and it's price range")
```



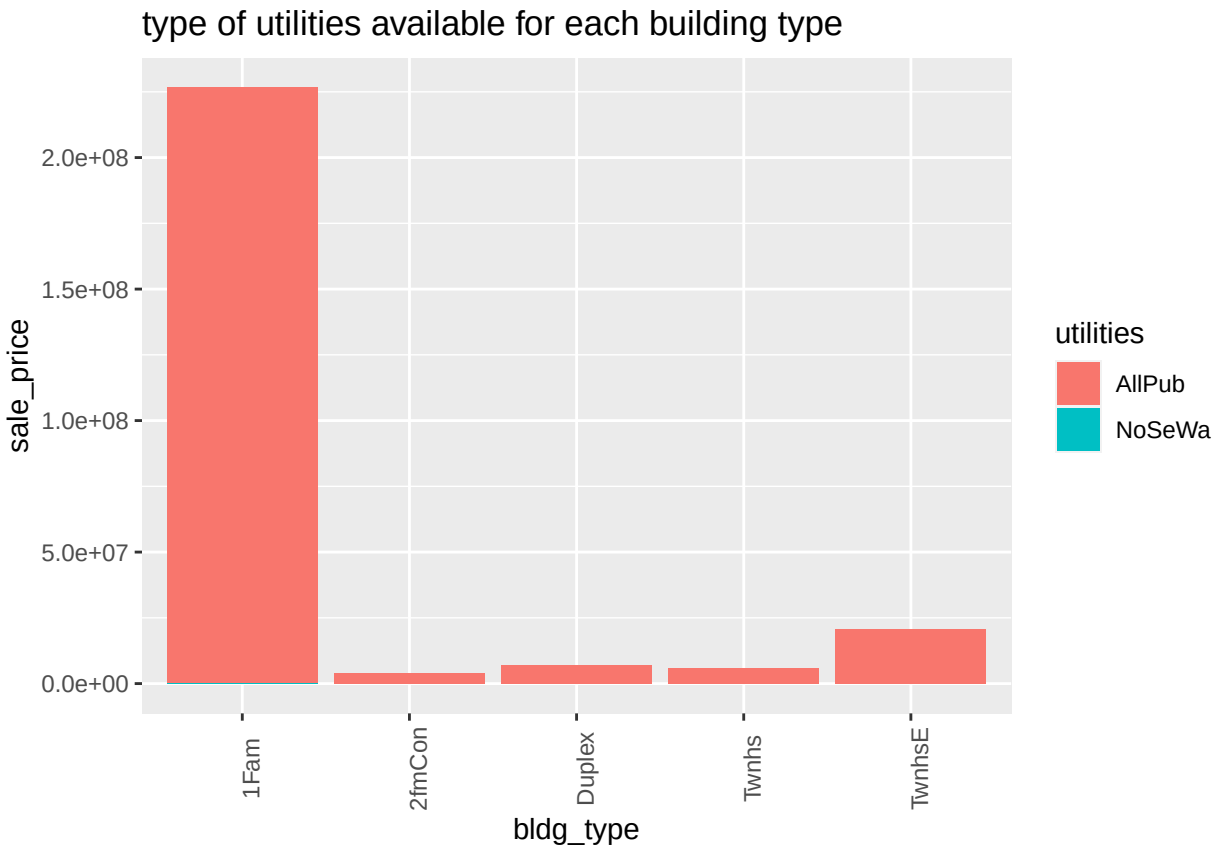
preferred building type in a neighborhood

```
ggplot(data, aes(fill=bldg_type, y=sale_price, x=neighborhood)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "preferred building type in a neighborhood")
```



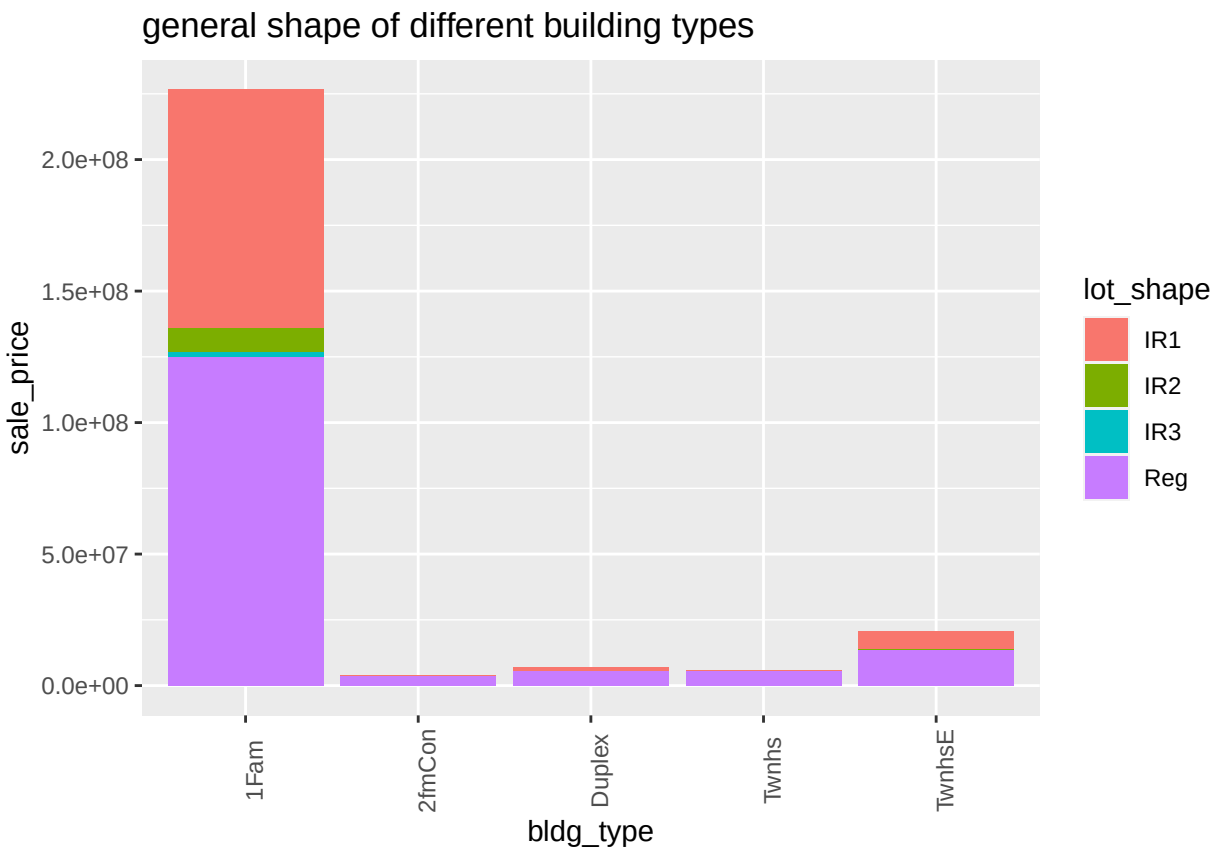
Type of utilities available for each building type

```
ggplot(data, aes(fill=utilities, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "type of utilities available for each building type")
```



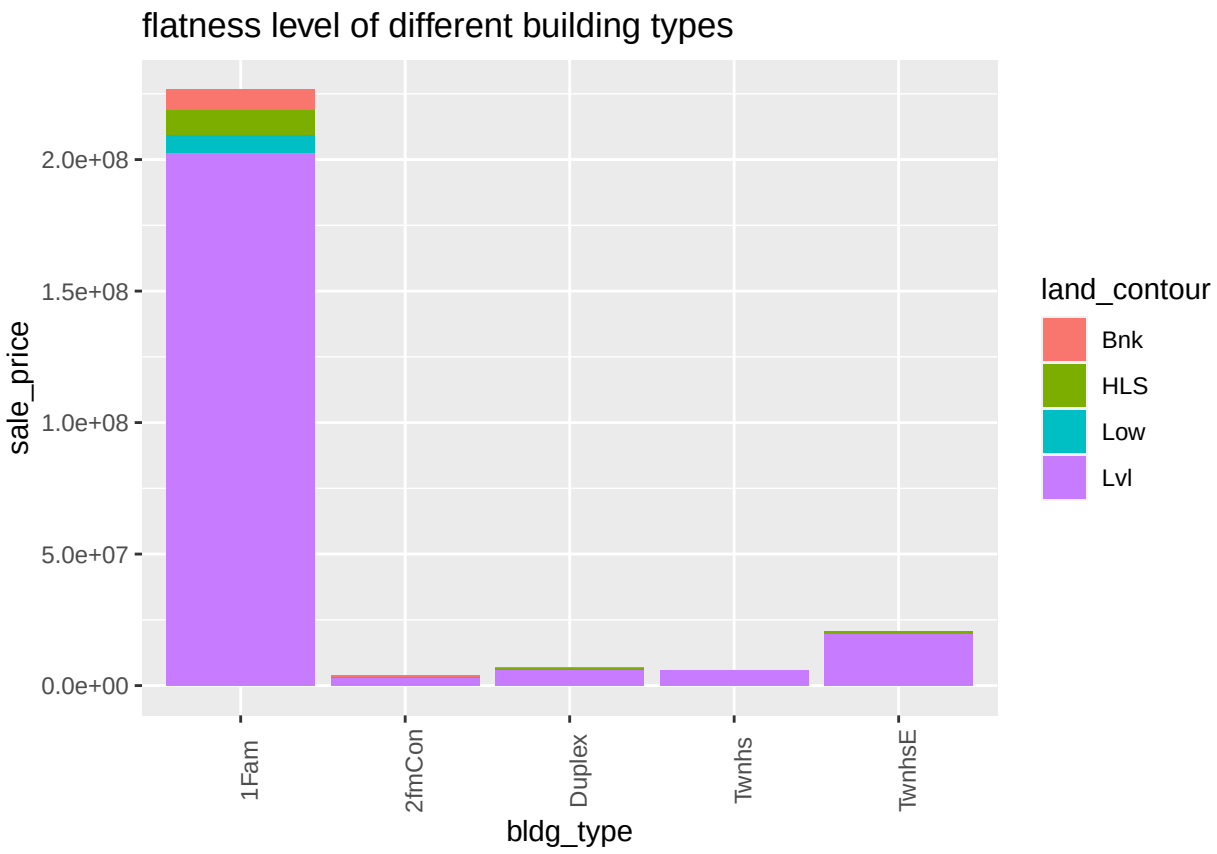
General shape of different building types

```
ggplot(data, aes(fill=lot_shape, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "general shape of different building types")
```



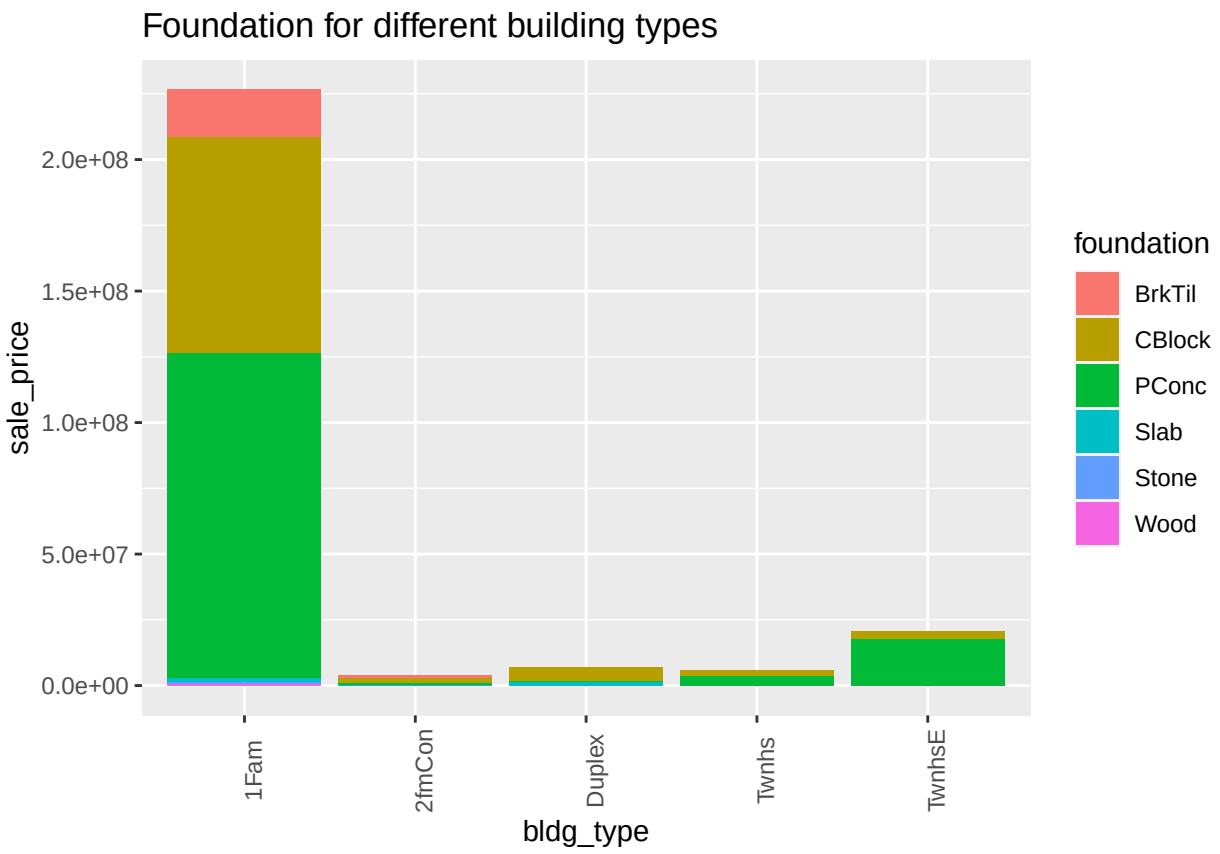
Flatness level of different building types

```
ggplot(data, aes(fill=land_contour, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "flatness level of different building types")
```



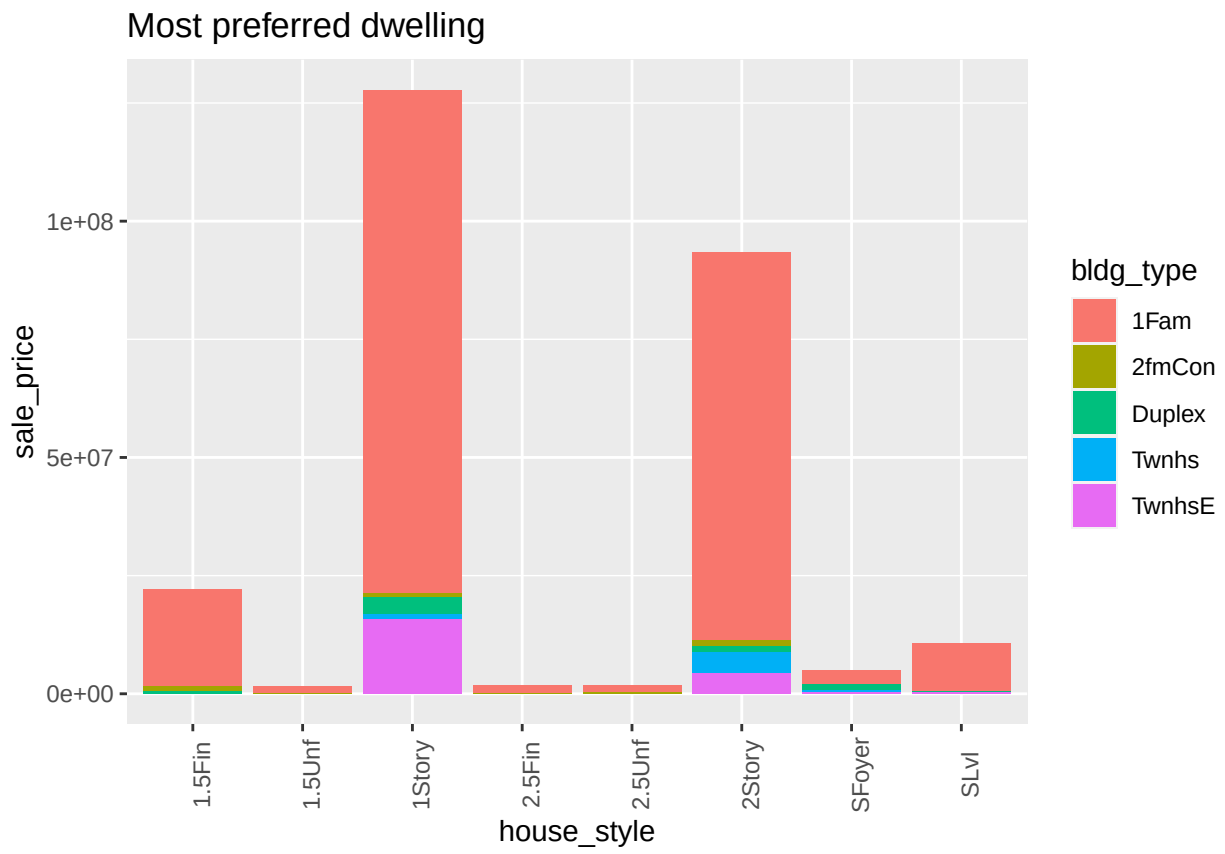
Type of foundation

```
ggplot(data, aes(fill=foundation, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Foundation for different building types")
```



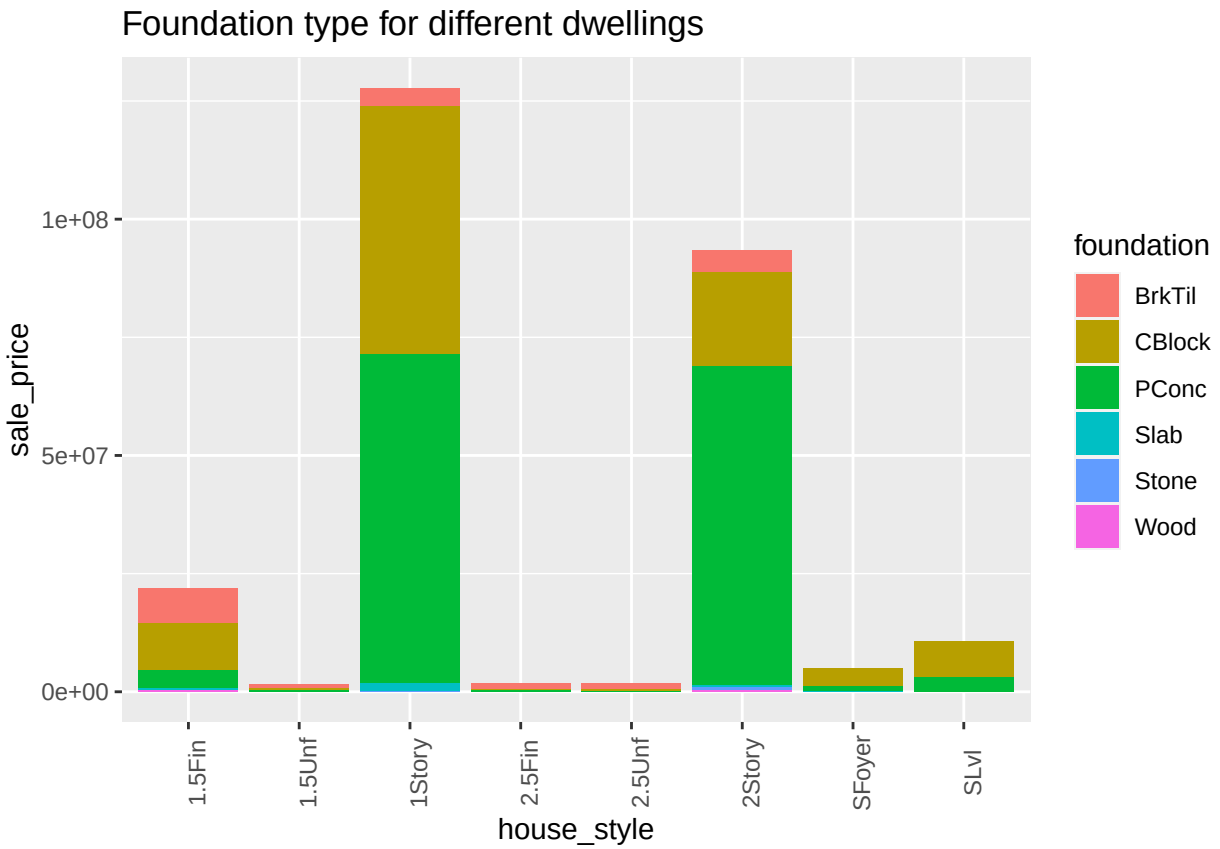
Different dwelling types

```
ggplot(data, aes(fill=bldg_type, y=sale_price, x=house_style)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Most preferred dwelling")
```



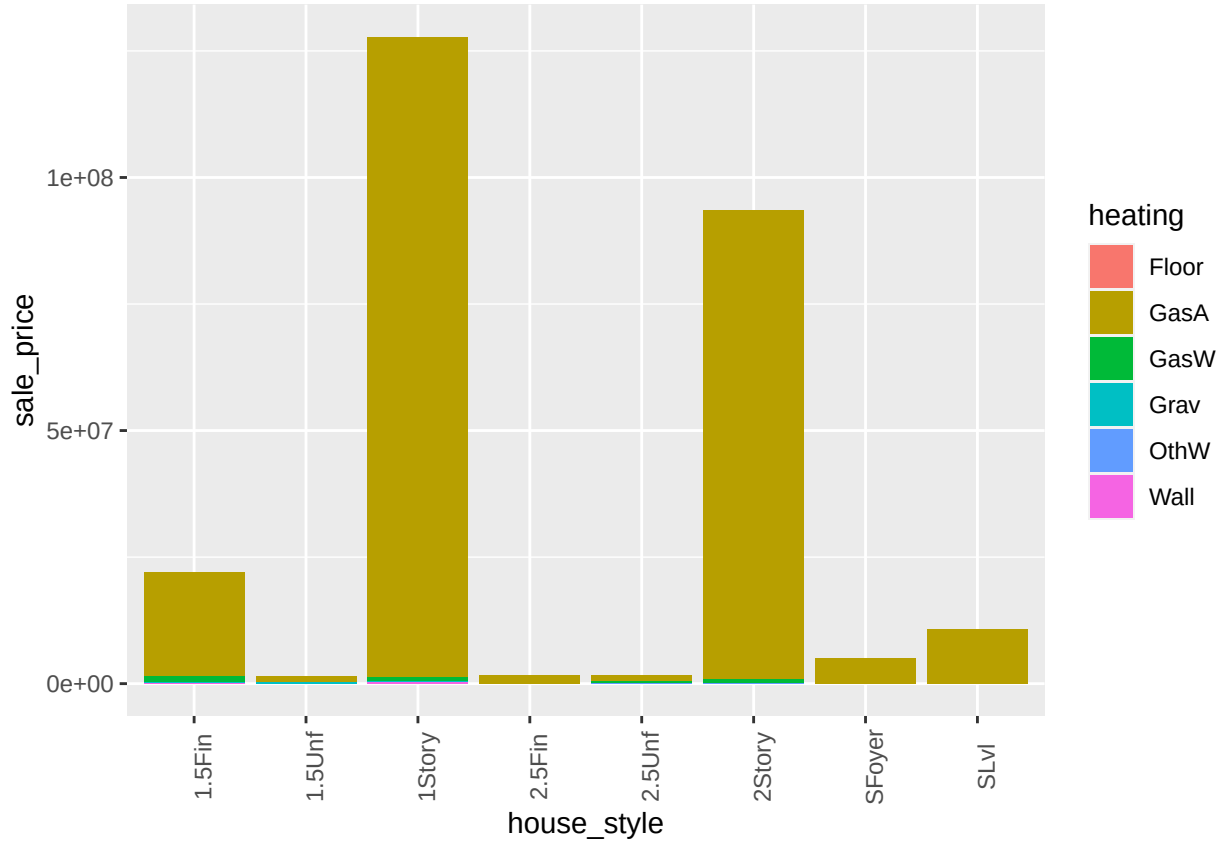
Type of foundation for different dwellings

```
ggplot(data, aes(fill=foundation, y=sale_price, x=house_style)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Foundation type for different dwellings")
```



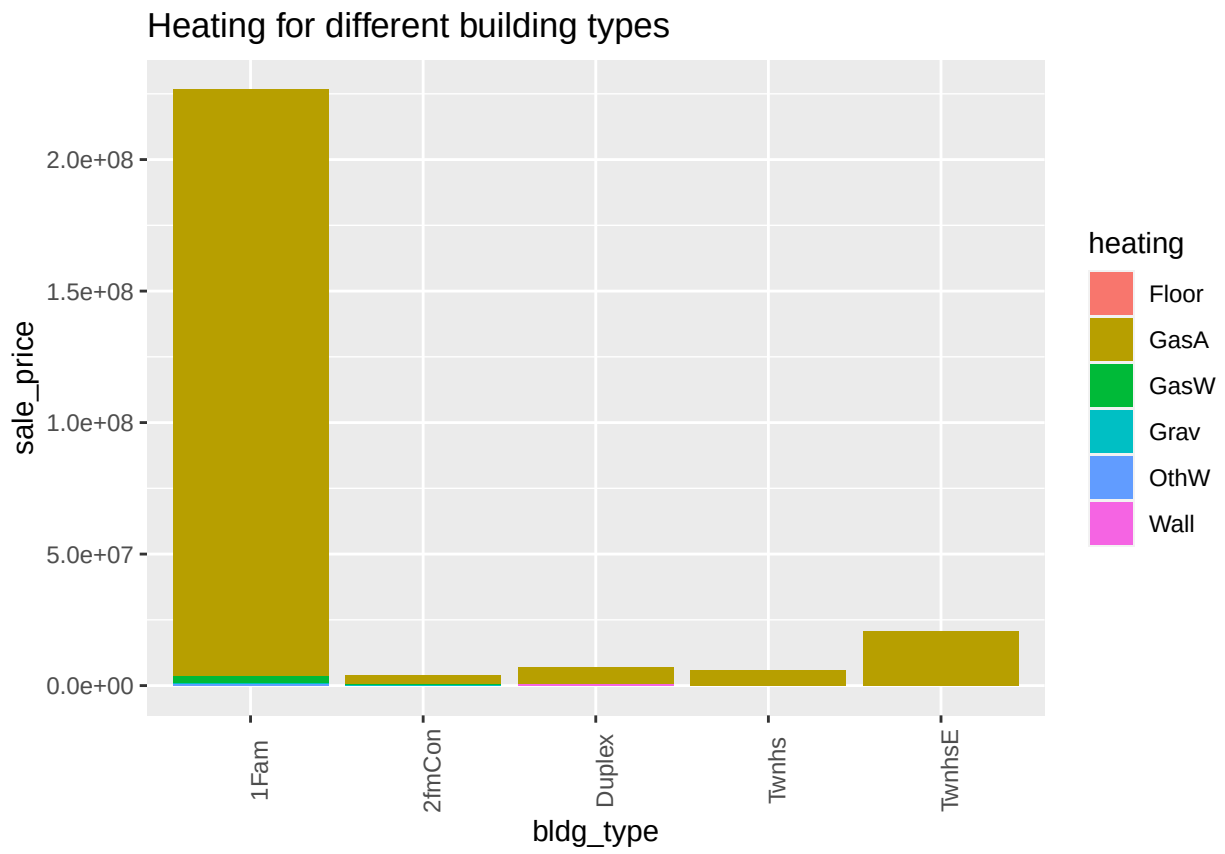
Kind of heating for different dwellings

```
ggplot(data, aes(fill=heating, y=sale_price, x=house_style)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs("heating for different house style")
```



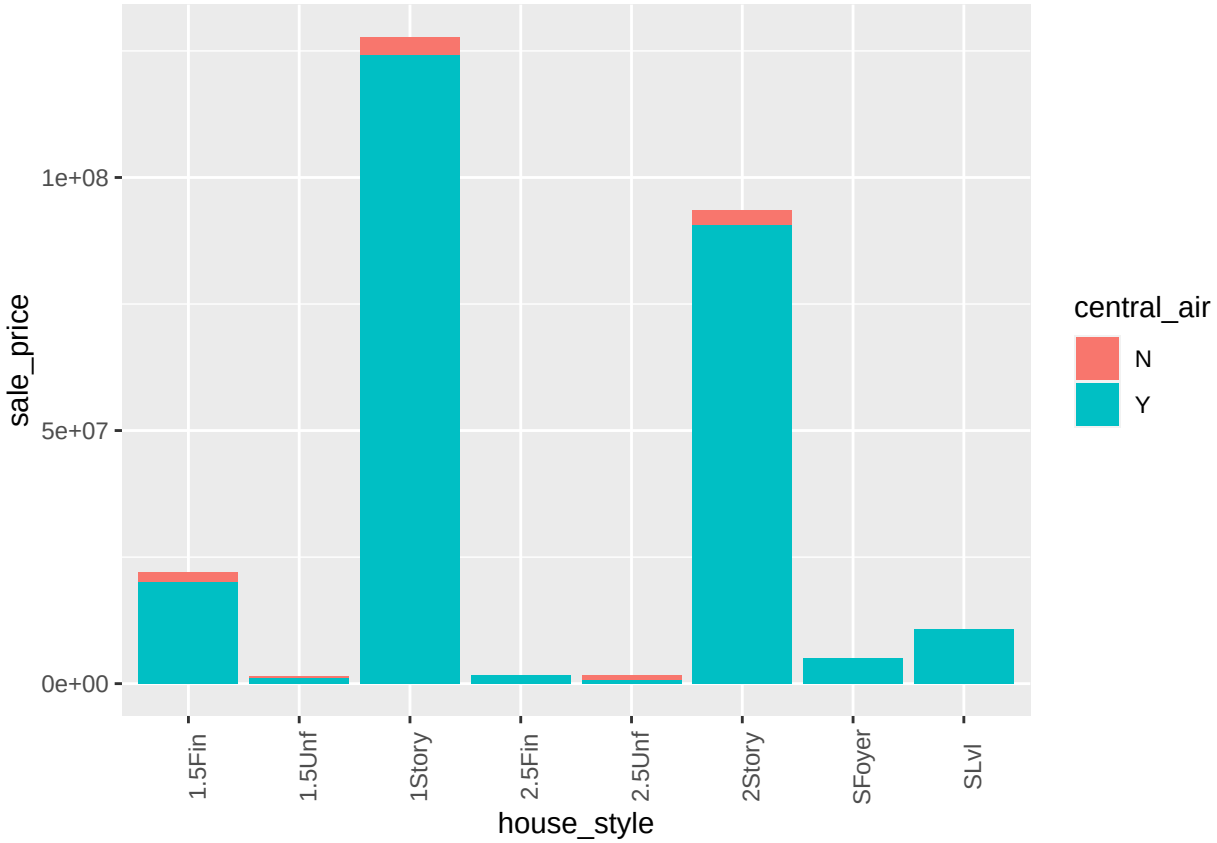
Kind of heating for different building types

```
ggplot(data, aes(fill=heating, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "Heating for different building types")
```



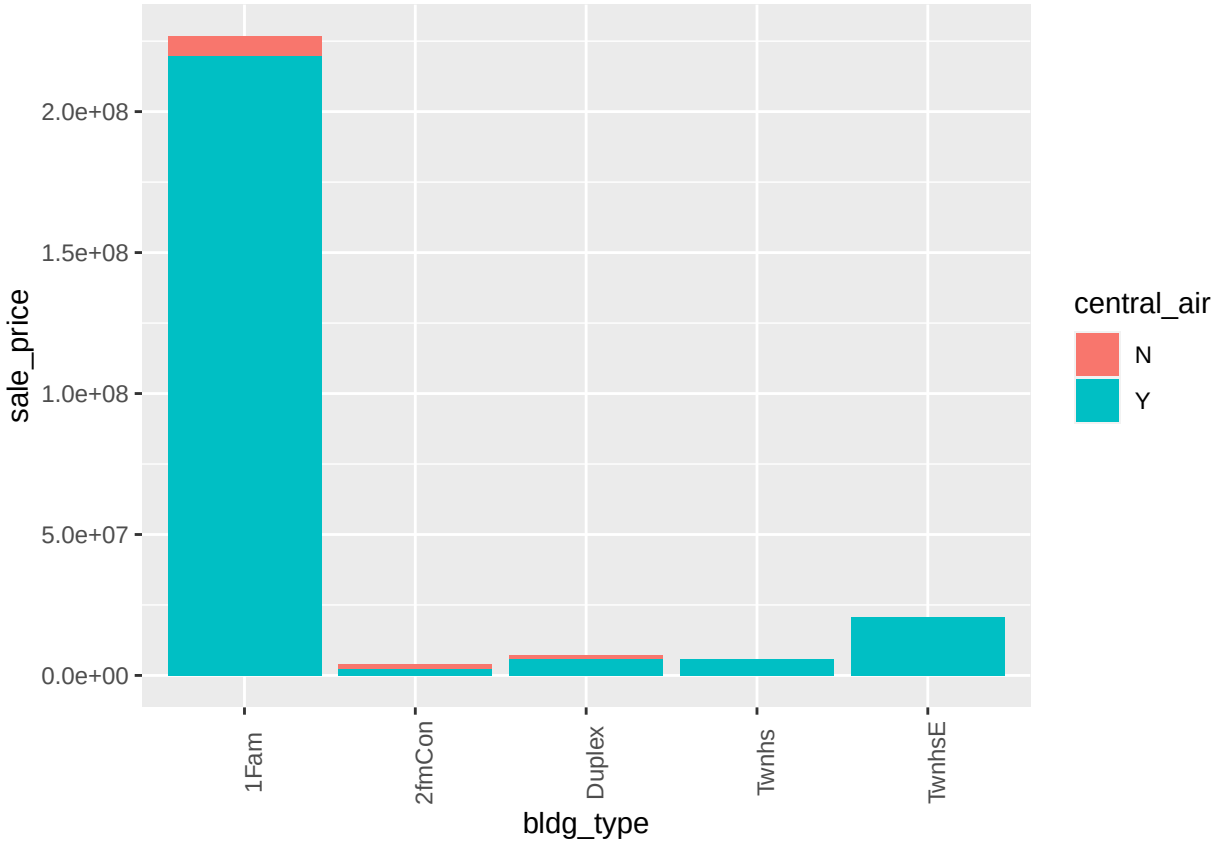
Air conditioning for different house styles

```
ggplot(data, aes(fill=central_air, y=sale_price, x=house_style)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs("air conditioning for different house style")
```



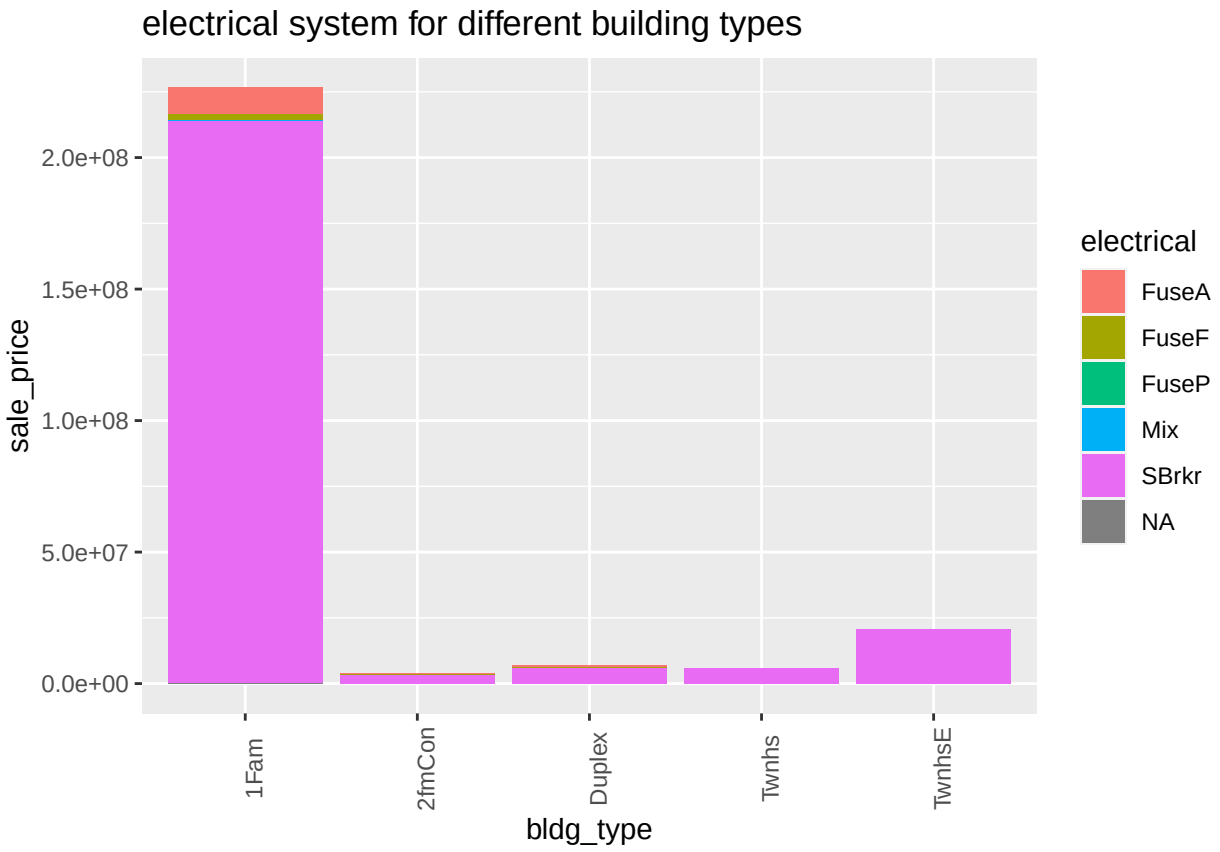
Air conditioning for different building types

```
ggplot(data, aes(fill=central_air, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs("air conditioning for different building types")
```



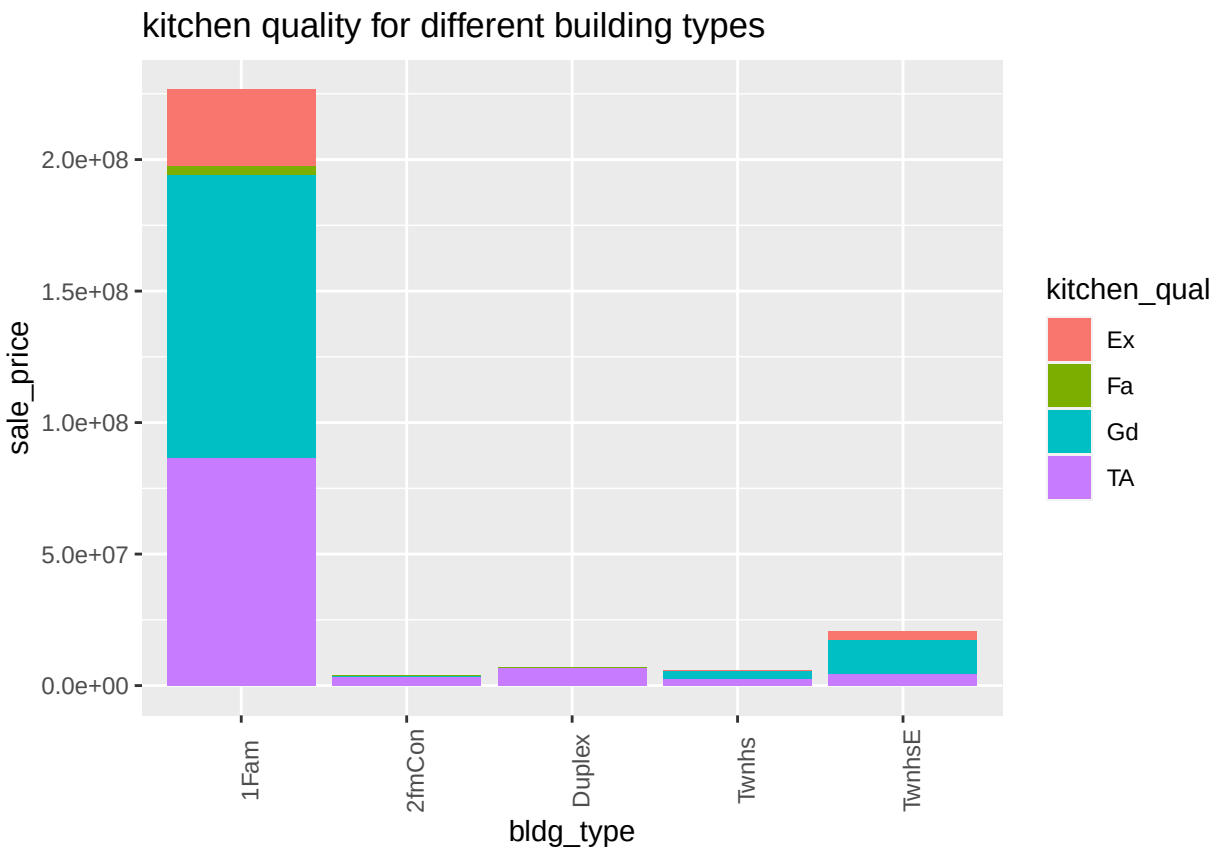
Electrical system for different building types

```
ggplot(data, aes(fill=electrical, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "electrical system for different building types")
```



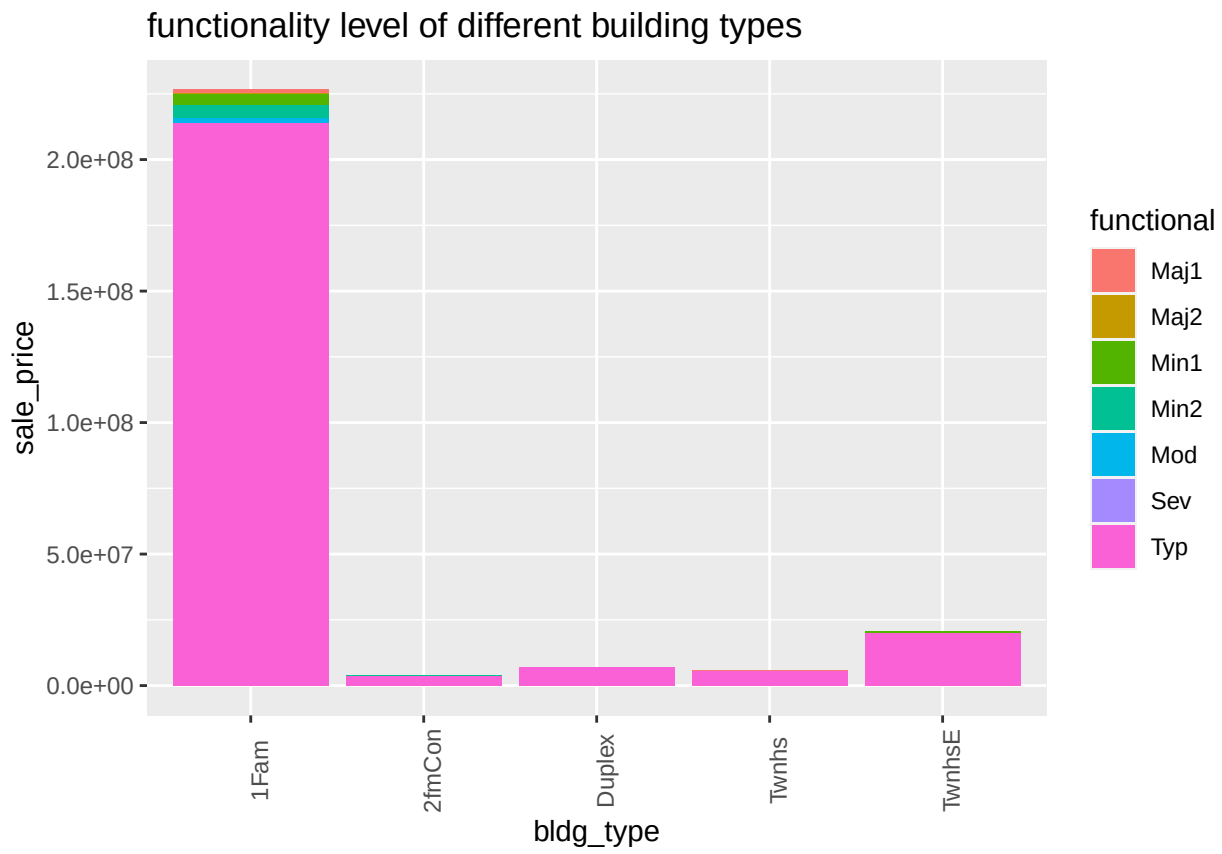
Kitchen quality for different building type

```
ggplot(data, aes(fill=kitchen_qual, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "kitchen quality for different building types")
```



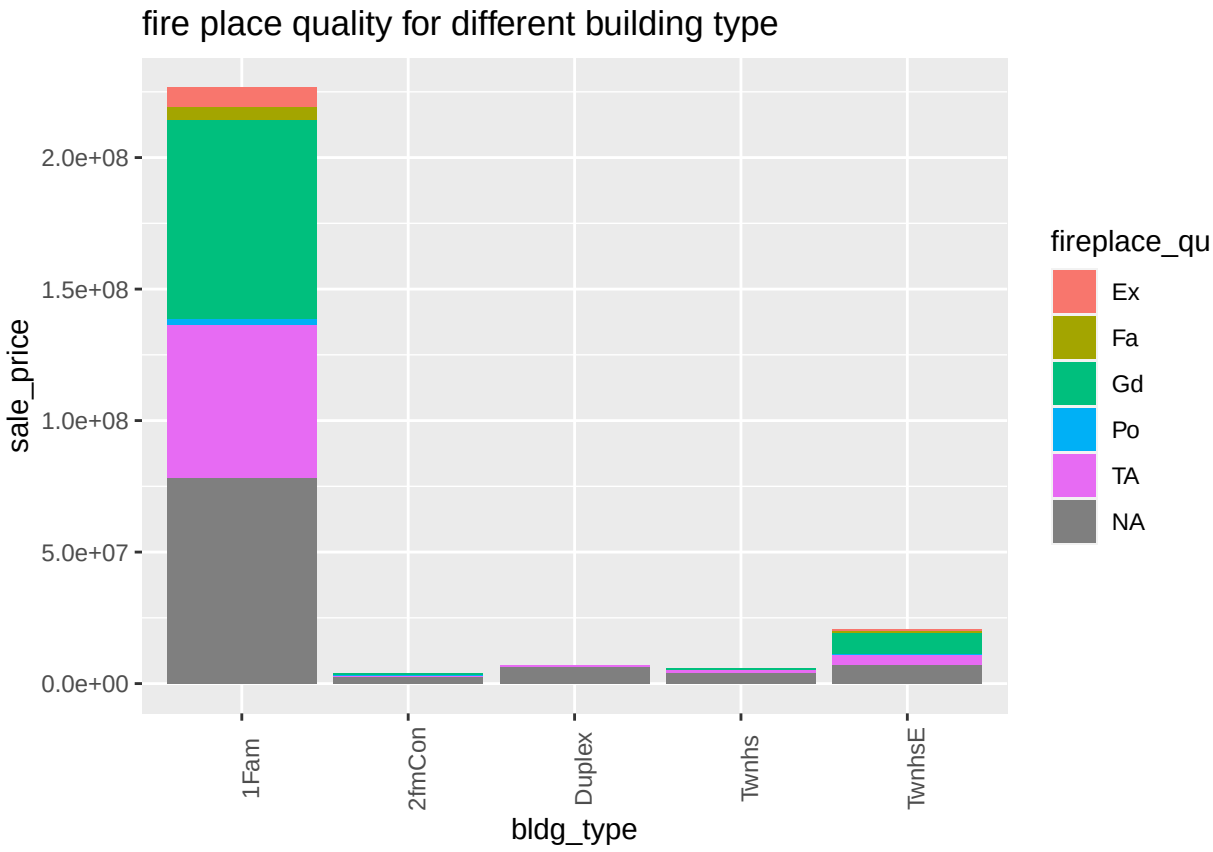
Functionality level of different building type

```
ggplot(data, aes(fill=functional, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "functionality level of different building types")
```



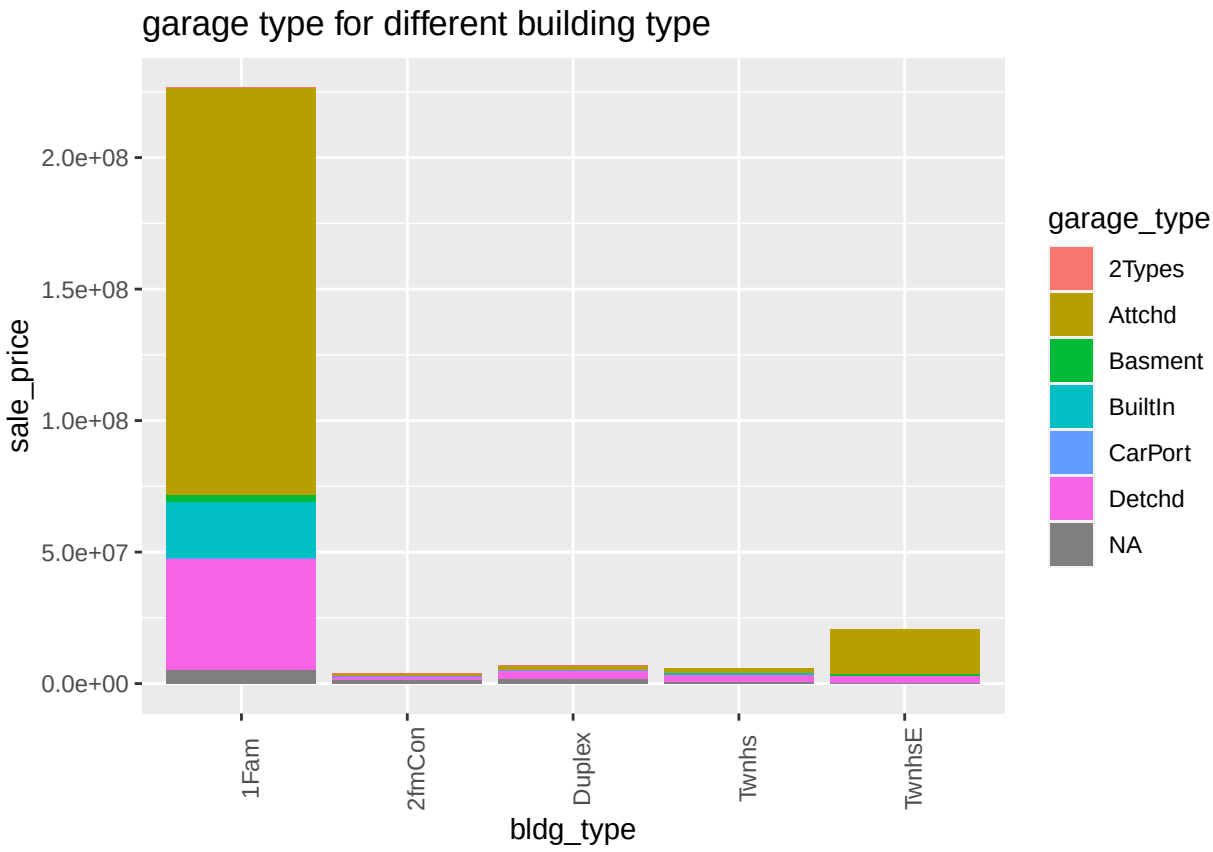
Fire place quality for different building type

```
ggplot(data, aes(fill=fireplace_qu, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "fire place quality for different building type")
```



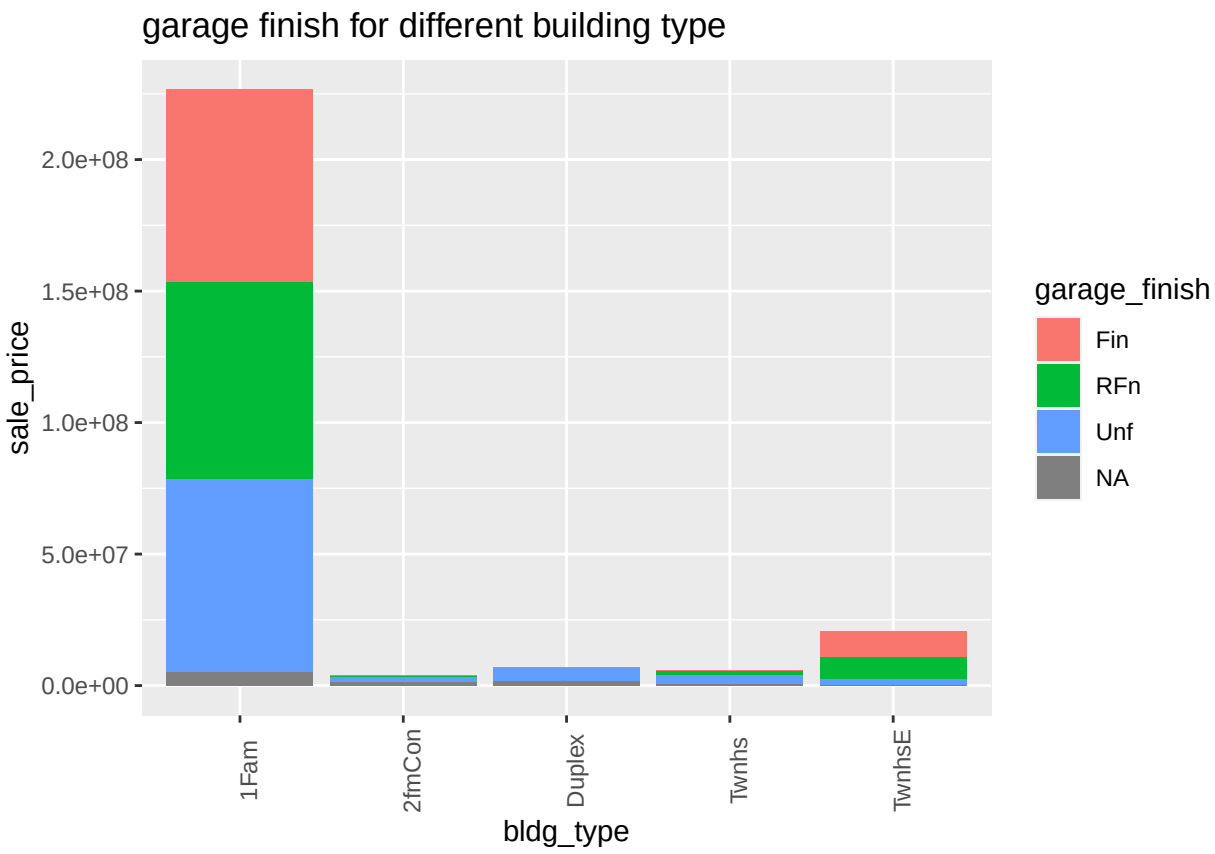
Garage type for different building type

```
ggplot(data, aes(fill=garage_type, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "garage type for different building type")
```



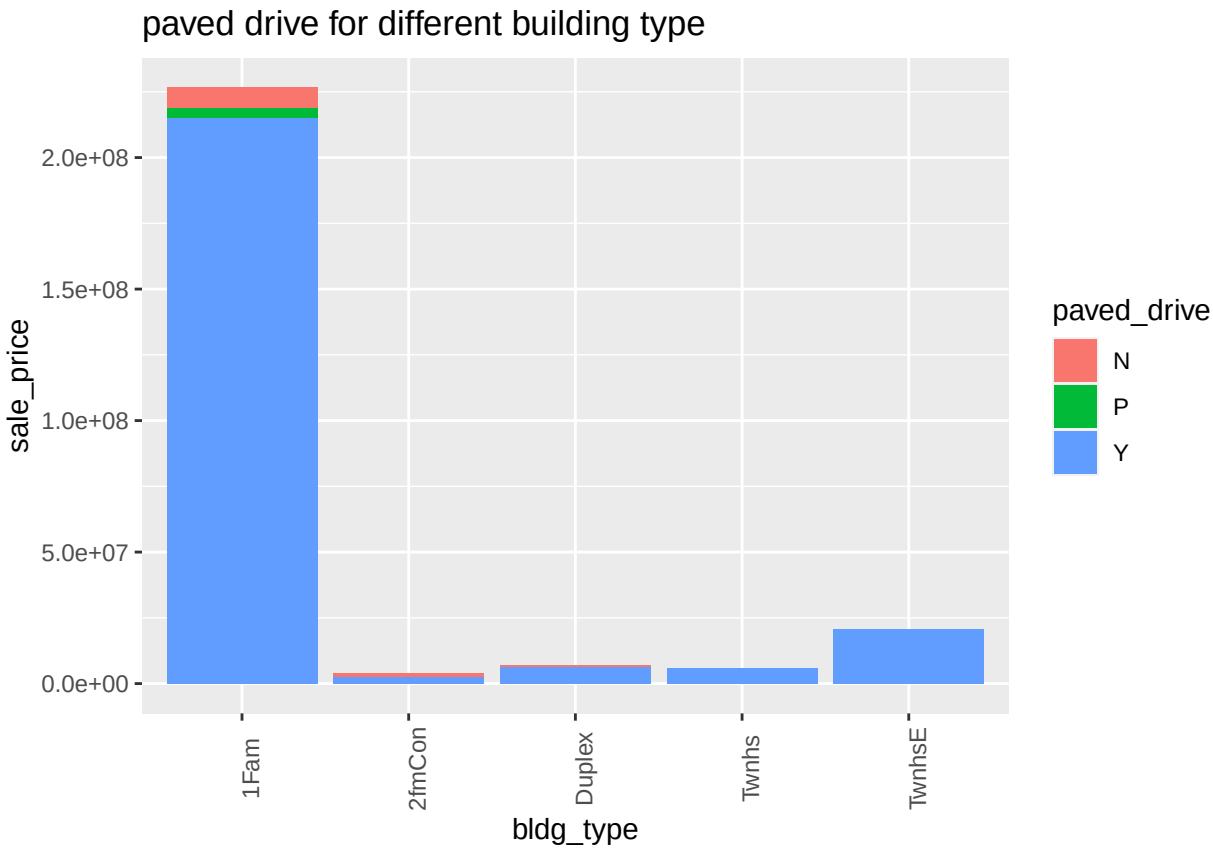
Garage finish for different building type

```
ggplot(data, aes(fill=garage_finish, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "garage finish for different building type")
```



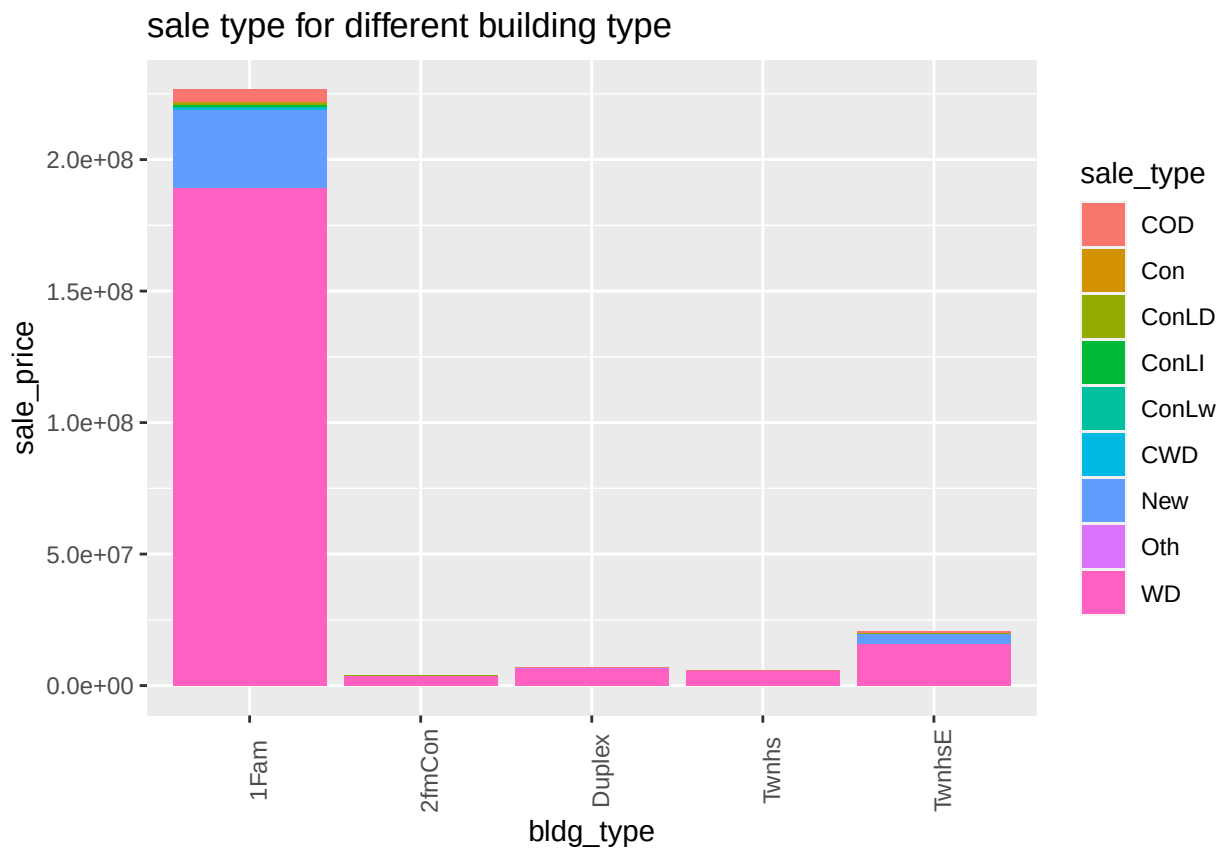
Paved drive for different building type

```
ggplot(data, aes(fill=paved_drive, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "paved drive for different building type")
```



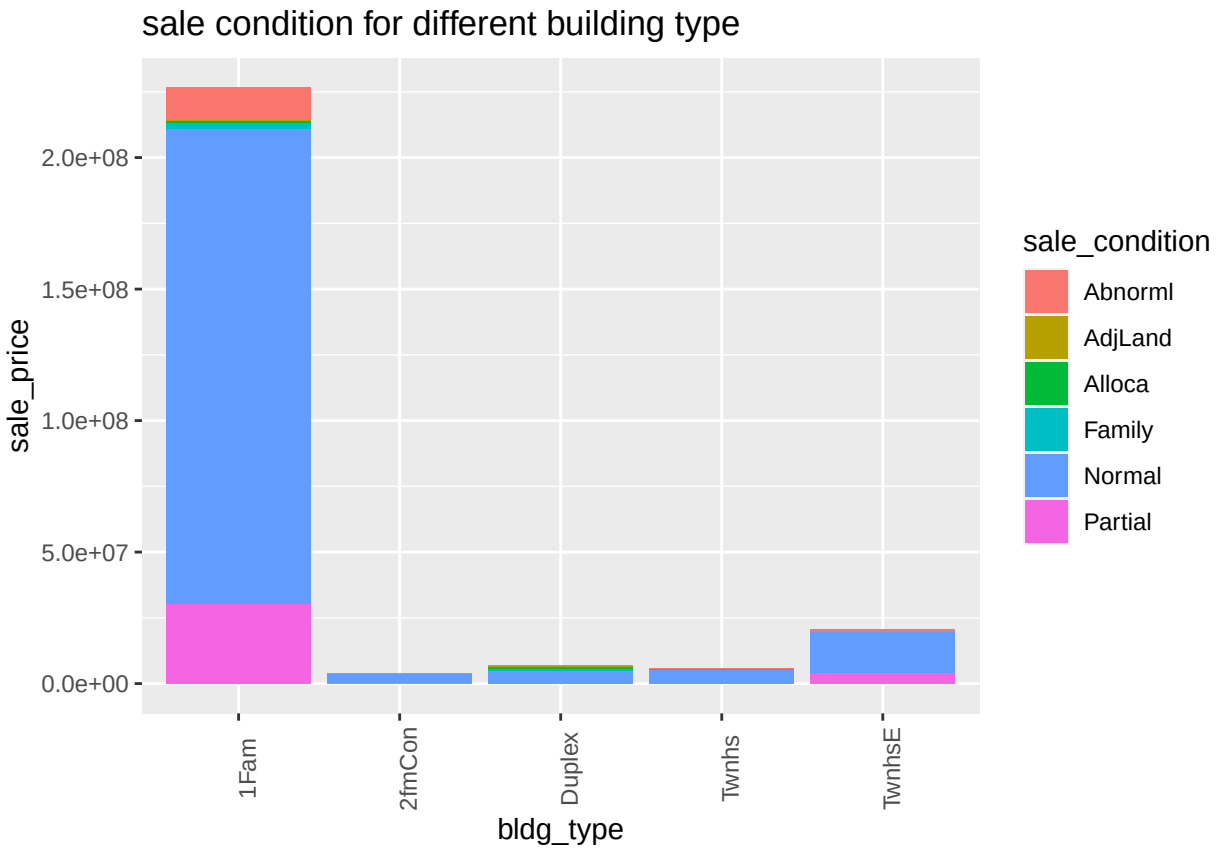
Sale type for different building type

```
ggplot(data, aes(fill=sale_type, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "sale type for different building type")
```



Sale condition for different building type

```
ggplot(data, aes(fill=sale_condition, y=sale_price, x=bldg_type)) +
  geom_bar(position="stack", stat="identity") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "sale condition for different building type")
```



Therefore from the above observations, houses with single story, double story under residential low density zone with detached single family house with all utilities but without central air conditioning are in high demand.

Other factors that influence house prices are

1. Shape of the property — slightly irregular shape is preferred.
2. Flatness of the property — quick and significant rise from street grade to building is preferred.
3. Foundation — Brick & Tile, cinder block foundation is preferred.
4. Type of heating — Gas forced warm air furnance heating is preferred.
5. Electrical system — Standard Circuit Breakers & Romex, Fuse Box over 60 AMP and all Romex wiring (Average) is preferred.
6. Kitchen quality — Excellent, good kitchen quality is preferred.
7. Home functionality — Typical functionality is preferred.
8. Garage location — houses with attached garage is preferred.
9. Interior finish of the garage — Rough finished, finished are preferred.
10. paved driveway — Dirt/gravel driveway is bit costly than the paved driveway.
11. type of sale — warranty deed- conventional type of sale is preferred.