Descriptive analysis of the dataset of home sales in the city of Ames, Iowa for BAN 502 Project Phase 1.

Prafull Pandey

2023-06-19

# Exploratory/descriptive analysis of of home sales in the city of Ames, Iowa. The response variable in the dataset is “Above\_Median”

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tidymodels)

## ── Attaching packages ────────────────────────────────────── tidymodels 1.1.0 ──  
## ✔ broom 1.0.4 ✔ rsample 1.1.1  
## ✔ dials 1.2.0 ✔ tune 1.1.1  
## ✔ infer 1.0.4 ✔ workflows 1.1.3  
## ✔ modeldata 1.1.0 ✔ workflowsets 1.0.1  
## ✔ parsnip 1.1.0 ✔ yardstick 1.2.0  
## ✔ recipes 1.0.6   
## ── Conflicts ───────────────────────────────────────── tidymodels\_conflicts() ──  
## ✖ scales::discard() masks purrr::discard()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ recipes::fixed() masks stringr::fixed()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ yardstick::spec() masks readr::spec()  
## ✖ recipes::step() masks stats::step()  
## • Use tidymodels\_prefer() to resolve common conflicts.

library(GGally)

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

library(gridExtra) #used for a little fancy arranging of plots

##   
## Attaching package: 'gridExtra'  
##   
## The following object is masked from 'package:dplyr':  
##   
## combine

library(car) #for the VIF function

## Loading required package: carData  
##   
## Attaching package: 'car'  
##   
## The following object is masked from 'package:dplyr':  
##   
## recode  
##   
## The following object is masked from 'package:purrr':  
##   
## some

library(glmnet)

## Loading required package: Matrix  
##   
## Attaching package: 'Matrix'  
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack  
##   
## Loaded glmnet 4.1-7

library(skimr)  
library(ggcorrplot) #create an alternative to ggcorr plots  
library(MASS) #access to forward and backward selection algorithms

##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select

library(leaps) #best subset selection  
library(lmtest) #for the dw test

## Loading required package: zoo  
##   
## Attaching package: 'zoo'  
##   
## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(splines) #for nonlinear fitting  
library(car) #for calculating the variance inflation factor

## Read-in data. For this work we are using the dataset of home sales in the city of Ames, Iowa.

ames=read\_csv("ames\_student-1.csv")

## Rows: 2053 Columns: 81  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (47): MS\_SubClass, MS\_Zoning, Street, Alley, Lot\_Shape, Land\_Contour, Ut...  
## dbl (34): Lot\_Frontage, Lot\_Area, Year\_Built, Year\_Remod\_Add, Mas\_Vnr\_Area, ...  
##   
## 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.

## Examined the ames data frame

str(ames)

## spc\_tbl\_ [2,053 × 81] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ MS\_SubClass : chr [1:2053] "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" ...  
## $ MS\_Zoning : chr [1:2053] "Residential\_Low\_Density" "Residential\_High\_Density" "Residential\_Low\_Density" "Residential\_Low\_Density" ...  
## $ Lot\_Frontage : num [1:2053] 141 80 81 93 74 78 43 39 0 85 ...  
## $ Lot\_Area : num [1:2053] 31770 11622 14267 11160 13830 ...  
## $ Street : chr [1:2053] "Pave" "Pave" "Pave" "Pave" ...  
## $ Alley : chr [1:2053] "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" ...  
## $ Lot\_Shape : chr [1:2053] "Slightly\_Irregular" "Regular" "Slightly\_Irregular" "Regular" ...  
## $ Land\_Contour : chr [1:2053] "Lvl" "Lvl" "Lvl" "Lvl" ...  
## $ Utilities : chr [1:2053] "AllPub" "AllPub" "AllPub" "AllPub" ...  
## $ Lot\_Config : chr [1:2053] "Corner" "Inside" "Corner" "Corner" ...  
## $ Land\_Slope : chr [1:2053] "Gtl" "Gtl" "Gtl" "Gtl" ...  
## $ Neighborhood : chr [1:2053] "North\_Ames" "North\_Ames" "North\_Ames" "North\_Ames" ...  
## $ Condition\_1 : chr [1:2053] "Norm" "Feedr" "Norm" "Norm" ...  
## $ Condition\_2 : chr [1:2053] "Norm" "Norm" "Norm" "Norm" ...  
## $ Bldg\_Type : chr [1:2053] "OneFam" "OneFam" "OneFam" "OneFam" ...  
## $ House\_Style : chr [1:2053] "One\_Story" "One\_Story" "One\_Story" "One\_Story" ...  
## $ Overall\_Qual : chr [1:2053] "Above\_Average" "Average" "Above\_Average" "Good" ...  
## $ Overall\_Cond : chr [1:2053] "Average" "Above\_Average" "Above\_Average" "Average" ...  
## $ Year\_Built : num [1:2053] 1960 1961 1958 1968 1997 ...  
## $ Year\_Remod\_Add : num [1:2053] 1960 1961 1958 1968 1998 ...  
## $ Roof\_Style : chr [1:2053] "Hip" "Gable" "Hip" "Hip" ...  
## $ Roof\_Matl : chr [1:2053] "CompShg" "CompShg" "CompShg" "CompShg" ...  
## $ Exterior\_1st : chr [1:2053] "BrkFace" "VinylSd" "Wd Sdng" "BrkFace" ...  
## $ Exterior\_2nd : chr [1:2053] "Plywood" "VinylSd" "Wd Sdng" "BrkFace" ...  
## $ Mas\_Vnr\_Type : chr [1:2053] "Stone" "None" "BrkFace" "None" ...  
## $ Mas\_Vnr\_Area : num [1:2053] 112 0 108 0 0 20 0 0 0 0 ...  
## $ Exter\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Good" ...  
## $ Exter\_Cond : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Foundation : chr [1:2053] "CBlock" "CBlock" "CBlock" "CBlock" ...  
## $ Bsmt\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Cond : chr [1:2053] "Good" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Exposure : chr [1:2053] "Gd" "No" "No" "No" ...  
## $ BsmtFin\_Type\_1 : chr [1:2053] "BLQ" "Rec" "ALQ" "ALQ" ...  
## $ BsmtFin\_SF\_1 : num [1:2053] 2 6 1 1 3 3 1 3 1 3 ...  
## $ BsmtFin\_Type\_2 : chr [1:2053] "Unf" "LwQ" "Unf" "Unf" ...  
## $ BsmtFin\_SF\_2 : num [1:2053] 0 144 0 0 0 0 0 0 0 0 ...  
## $ Bsmt\_Unf\_SF : num [1:2053] 441 270 406 1045 137 ...  
## $ Total\_Bsmt\_SF : num [1:2053] 1080 882 1329 2110 928 ...  
## $ Heating : chr [1:2053] "GasA" "GasA" "GasA" "GasA" ...  
## $ Heating\_QC : chr [1:2053] "Fair" "Typical" "Typical" "Excellent" ...  
## $ Central\_Air : chr [1:2053] "Y" "Y" "Y" "Y" ...  
## $ Electrical : chr [1:2053] "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...  
## $ First\_Flr\_SF : num [1:2053] 1656 896 1329 2110 928 ...  
## $ Second\_Flr\_SF : num [1:2053] 0 0 0 0 701 678 0 0 0 0 ...  
## $ Low\_Qual\_Fin\_SF : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:2053] 1656 896 1329 2110 1629 ...  
## $ Bsmt\_Full\_Bath : num [1:2053] 1 0 0 1 0 0 0 1 1 1 ...  
## $ Bsmt\_Half\_Bath : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:2053] 1 1 1 2 2 2 2 2 2 1 ...  
## $ Half\_Bath : num [1:2053] 0 0 1 1 1 1 0 0 0 1 ...  
## $ Bedroom\_AbvGr : num [1:2053] 3 2 3 3 3 3 2 2 3 2 ...  
## $ Kitchen\_AbvGr : num [1:2053] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Kitchen\_Qual : chr [1:2053] "Typical" "Typical" "Good" "Excellent" ...  
## $ TotRms\_AbvGrd : num [1:2053] 7 5 6 8 6 7 5 5 6 5 ...  
## $ Functional : chr [1:2053] "Typ" "Typ" "Typ" "Typ" ...  
## $ Fireplaces : num [1:2053] 2 0 0 2 1 1 0 1 0 1 ...  
## $ Fireplace\_Qu : chr [1:2053] "Good" "No\_Fireplace" "No\_Fireplace" "Typical" ...  
## $ Garage\_Type : chr [1:2053] "Attchd" "Attchd" "Attchd" "Attchd" ...  
## $ Garage\_Finish : chr [1:2053] "Fin" "Unf" "Unf" "Fin" ...  
## $ Garage\_Cars : num [1:2053] 2 1 1 2 2 2 2 2 2 2 ...  
## $ Garage\_Area : num [1:2053] 528 730 312 522 482 470 506 608 420 506 ...  
## $ Garage\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Garage\_Cond : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Paved\_Drive : chr [1:2053] "Partial\_Pavement" "Paved" "Paved" "Paved" ...  
## $ Wood\_Deck\_SF : num [1:2053] 210 140 393 0 212 360 0 237 483 192 ...  
## $ Open\_Porch\_SF : num [1:2053] 62 0 36 0 34 36 82 152 21 0 ...  
## $ Enclosed\_Porch : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Three\_season\_porch: num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Screen\_Porch : num [1:2053] 0 120 0 0 0 0 144 0 0 0 ...  
## $ Pool\_Area : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_QC : chr [1:2053] "No\_Pool" "No\_Pool" "No\_Pool" "No\_Pool" ...  
## $ Fence : chr [1:2053] "No\_Fence" "Minimum\_Privacy" "No\_Fence" "No\_Fence" ...  
## $ Misc\_Feature : chr [1:2053] "None" "None" "Gar2" "None" ...  
## $ Misc\_Val : num [1:2053] 0 0 12500 0 0 0 0 0 500 0 ...  
## $ Mo\_Sold : num [1:2053] 5 6 6 4 3 6 1 3 3 2 ...  
## $ Year\_Sold : num [1:2053] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : chr [1:2053] "WD" "WD" "WD" "WD" ...  
## $ Sale\_Condition : chr [1:2053] "Normal" "Normal" "Normal" "Normal" ...  
## $ Longitude : num [1:2053] -93.6 -93.6 -93.6 -93.6 -93.6 ...  
## $ Latitude : num [1:2053] 42.1 42.1 42.1 42.1 42.1 ...  
## $ Above\_Median : chr [1:2053] "Yes" "No" "Yes" "Yes" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. MS\_SubClass = col\_character(),  
## .. MS\_Zoning = col\_character(),  
## .. Lot\_Frontage = col\_double(),  
## .. Lot\_Area = col\_double(),  
## .. Street = col\_character(),  
## .. Alley = col\_character(),  
## .. Lot\_Shape = col\_character(),  
## .. Land\_Contour = col\_character(),  
## .. Utilities = col\_character(),  
## .. Lot\_Config = col\_character(),  
## .. Land\_Slope = col\_character(),  
## .. Neighborhood = col\_character(),  
## .. Condition\_1 = col\_character(),  
## .. Condition\_2 = col\_character(),  
## .. Bldg\_Type = col\_character(),  
## .. House\_Style = col\_character(),  
## .. Overall\_Qual = col\_character(),  
## .. Overall\_Cond = col\_character(),  
## .. Year\_Built = col\_double(),  
## .. Year\_Remod\_Add = col\_double(),  
## .. Roof\_Style = col\_character(),  
## .. Roof\_Matl = col\_character(),  
## .. Exterior\_1st = col\_character(),  
## .. Exterior\_2nd = col\_character(),  
## .. Mas\_Vnr\_Type = col\_character(),  
## .. Mas\_Vnr\_Area = col\_double(),  
## .. Exter\_Qual = col\_character(),  
## .. Exter\_Cond = col\_character(),  
## .. Foundation = col\_character(),  
## .. Bsmt\_Qual = col\_character(),  
## .. Bsmt\_Cond = col\_character(),  
## .. Bsmt\_Exposure = col\_character(),  
## .. BsmtFin\_Type\_1 = col\_character(),  
## .. BsmtFin\_SF\_1 = col\_double(),  
## .. BsmtFin\_Type\_2 = col\_character(),  
## .. BsmtFin\_SF\_2 = col\_double(),  
## .. Bsmt\_Unf\_SF = col\_double(),  
## .. Total\_Bsmt\_SF = col\_double(),  
## .. Heating = col\_character(),  
## .. Heating\_QC = col\_character(),  
## .. Central\_Air = col\_character(),  
## .. Electrical = col\_character(),  
## .. First\_Flr\_SF = col\_double(),  
## .. Second\_Flr\_SF = col\_double(),  
## .. Low\_Qual\_Fin\_SF = col\_double(),  
## .. Gr\_Liv\_Area = col\_double(),  
## .. Bsmt\_Full\_Bath = col\_double(),  
## .. Bsmt\_Half\_Bath = col\_double(),  
## .. Full\_Bath = col\_double(),  
## .. Half\_Bath = col\_double(),  
## .. Bedroom\_AbvGr = col\_double(),  
## .. Kitchen\_AbvGr = col\_double(),  
## .. Kitchen\_Qual = col\_character(),  
## .. TotRms\_AbvGrd = col\_double(),  
## .. Functional = col\_character(),  
## .. Fireplaces = col\_double(),  
## .. Fireplace\_Qu = col\_character(),  
## .. Garage\_Type = col\_character(),  
## .. Garage\_Finish = col\_character(),  
## .. Garage\_Cars = col\_double(),  
## .. Garage\_Area = col\_double(),  
## .. Garage\_Qual = col\_character(),  
## .. Garage\_Cond = col\_character(),  
## .. Paved\_Drive = col\_character(),  
## .. Wood\_Deck\_SF = col\_double(),  
## .. Open\_Porch\_SF = col\_double(),  
## .. Enclosed\_Porch = col\_double(),  
## .. Three\_season\_porch = col\_double(),  
## .. Screen\_Porch = col\_double(),  
## .. Pool\_Area = col\_double(),  
## .. Pool\_QC = col\_character(),  
## .. Fence = col\_character(),  
## .. Misc\_Feature = col\_character(),  
## .. Misc\_Val = col\_double(),  
## .. Mo\_Sold = col\_double(),  
## .. Year\_Sold = col\_double(),  
## .. Sale\_Type = col\_character(),  
## .. Sale\_Condition = col\_character(),  
## .. Longitude = col\_double(),  
## .. Latitude = col\_double(),  
## .. Above\_Median = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

#glimpse(ames)  
skim(ames)

Data summary

|  |  |
| --- | --- |
| Name | ames |
| Number of rows | 2053 |
| Number of columns | 81 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 47 |
| numeric | 34 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| MS\_SubClass | 0 | 1 | 11 | 41 | 0 | 16 | 0 |
| MS\_Zoning | 0 | 1 | 5 | 28 | 0 | 7 | 0 |
| Street | 0 | 1 | 4 | 4 | 0 | 2 | 0 |
| Alley | 0 | 1 | 5 | 15 | 0 | 3 | 0 |
| Lot\_Shape | 0 | 1 | 7 | 20 | 0 | 4 | 0 |
| Land\_Contour | 0 | 1 | 3 | 3 | 0 | 4 | 0 |
| Utilities | 0 | 1 | 6 | 6 | 0 | 2 | 0 |
| Lot\_Config | 0 | 1 | 3 | 7 | 0 | 5 | 0 |
| Land\_Slope | 0 | 1 | 3 | 3 | 0 | 3 | 0 |
| Neighborhood | 0 | 1 | 6 | 39 | 0 | 28 | 0 |
| Condition\_1 | 0 | 1 | 4 | 6 | 0 | 9 | 0 |
| Condition\_2 | 0 | 1 | 4 | 6 | 0 | 8 | 0 |
| Bldg\_Type | 0 | 1 | 5 | 8 | 0 | 5 | 0 |
| House\_Style | 0 | 1 | 4 | 16 | 0 | 8 | 0 |
| Overall\_Qual | 0 | 1 | 4 | 14 | 0 | 10 | 0 |
| Overall\_Cond | 0 | 1 | 4 | 13 | 0 | 9 | 0 |
| Roof\_Style | 0 | 1 | 3 | 7 | 0 | 6 | 0 |
| Roof\_Matl | 0 | 1 | 4 | 7 | 0 | 6 | 0 |
| Exterior\_1st | 0 | 1 | 5 | 7 | 0 | 16 | 0 |
| Exterior\_2nd | 0 | 1 | 5 | 7 | 0 | 17 | 0 |
| Mas\_Vnr\_Type | 0 | 1 | 4 | 7 | 0 | 5 | 0 |
| Exter\_Qual | 0 | 1 | 4 | 9 | 0 | 4 | 0 |
| Exter\_Cond | 0 | 1 | 4 | 9 | 0 | 5 | 0 |
| Foundation | 0 | 1 | 4 | 6 | 0 | 6 | 0 |
| Bsmt\_Qual | 0 | 1 | 4 | 11 | 0 | 6 | 0 |
| Bsmt\_Cond | 0 | 1 | 4 | 11 | 0 | 6 | 0 |
| Bsmt\_Exposure | 0 | 1 | 2 | 11 | 0 | 5 | 0 |
| BsmtFin\_Type\_1 | 0 | 1 | 3 | 11 | 0 | 7 | 0 |
| BsmtFin\_Type\_2 | 0 | 1 | 3 | 11 | 0 | 7 | 0 |
| Heating | 0 | 1 | 4 | 5 | 0 | 6 | 0 |
| Heating\_QC | 0 | 1 | 4 | 9 | 0 | 5 | 0 |
| Central\_Air | 0 | 1 | 1 | 1 | 0 | 2 | 0 |
| Electrical | 0 | 1 | 5 | 7 | 0 | 5 | 0 |
| Kitchen\_Qual | 0 | 1 | 4 | 9 | 0 | 5 | 0 |
| Functional | 0 | 1 | 3 | 4 | 0 | 8 | 0 |
| Fireplace\_Qu | 0 | 1 | 4 | 12 | 0 | 6 | 0 |
| Garage\_Type | 0 | 1 | 6 | 19 | 0 | 7 | 0 |
| Garage\_Finish | 0 | 1 | 3 | 9 | 0 | 4 | 0 |
| Garage\_Qual | 0 | 1 | 4 | 9 | 0 | 6 | 0 |
| Garage\_Cond | 0 | 1 | 4 | 9 | 0 | 6 | 0 |
| Paved\_Drive | 0 | 1 | 5 | 16 | 0 | 3 | 0 |
| Pool\_QC | 0 | 1 | 4 | 9 | 0 | 5 | 0 |
| Fence | 0 | 1 | 8 | 17 | 0 | 5 | 0 |
| Misc\_Feature | 0 | 1 | 4 | 4 | 0 | 5 | 0 |
| Sale\_Type | 0 | 1 | 2 | 5 | 0 | 10 | 0 |
| Sale\_Condition | 0 | 1 | 6 | 7 | 0 | 6 | 0 |
| Above\_Median | 0 | 1 | 2 | 3 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Lot\_Frontage | 0 | 1 | 57.38 | 33.20 | 0.00 | 43.00 | 62.00 | 78.00 | 313.00 | ▇▇▁▁▁ |
| Lot\_Area | 0 | 1 | 10258.40 | 8427.38 | 1300.00 | 7500.00 | 9548.00 | 11600.00 | 215245.00 | ▇▁▁▁▁ |
| Year\_Built | 0 | 1 | 1970.64 | 30.40 | 1875.00 | 1953.00 | 1972.00 | 2000.00 | 2010.00 | ▁▂▃▆▇ |
| Year\_Remod\_Add | 0 | 1 | 1984.08 | 20.96 | 1950.00 | 1965.00 | 1993.00 | 2004.00 | 2010.00 | ▅▂▂▃▇ |
| Mas\_Vnr\_Area | 0 | 1 | 103.75 | 183.59 | 0.00 | 0.00 | 0.00 | 164.00 | 1600.00 | ▇▁▁▁▁ |
| BsmtFin\_SF\_1 | 0 | 1 | 4.21 | 2.24 | 1.00 | 3.00 | 3.00 | 7.00 | 7.00 | ▅▆▁▁▇ |
| BsmtFin\_SF\_2 | 0 | 1 | 52.57 | 175.99 | 0.00 | 0.00 | 0.00 | 0.00 | 1526.00 | ▇▁▁▁▁ |
| Bsmt\_Unf\_SF | 0 | 1 | 561.19 | 441.72 | 0.00 | 226.00 | 460.00 | 801.00 | 2336.00 | ▇▅▂▁▁ |
| Total\_Bsmt\_SF | 0 | 1 | 1054.57 | 435.33 | 0.00 | 793.00 | 988.00 | 1304.00 | 5095.00 | ▇▇▁▁▁ |
| First\_Flr\_SF | 0 | 1 | 1167.52 | 391.79 | 432.00 | 882.00 | 1088.00 | 1402.00 | 5095.00 | ▇▃▁▁▁ |
| Second\_Flr\_SF | 0 | 1 | 326.07 | 422.44 | 0.00 | 0.00 | 0.00 | 701.00 | 1862.00 | ▇▂▂▁▁ |
| Low\_Qual\_Fin\_SF | 0 | 1 | 4.97 | 49.09 | 0.00 | 0.00 | 0.00 | 0.00 | 1064.00 | ▇▁▁▁▁ |
| Gr\_Liv\_Area | 0 | 1 | 1498.56 | 487.84 | 480.00 | 1137.00 | 1447.00 | 1737.00 | 5095.00 | ▇▇▁▁▁ |
| Bsmt\_Full\_Bath | 0 | 1 | 0.43 | 0.53 | 0.00 | 0.00 | 0.00 | 1.00 | 3.00 | ▇▆▁▁▁ |
| Bsmt\_Half\_Bath | 0 | 1 | 0.06 | 0.24 | 0.00 | 0.00 | 0.00 | 0.00 | 2.00 | ▇▁▁▁▁ |
| Full\_Bath | 0 | 1 | 1.56 | 0.55 | 0.00 | 1.00 | 2.00 | 2.00 | 4.00 | ▁▇▇▁▁ |
| Half\_Bath | 0 | 1 | 0.38 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 2.00 | ▇▁▅▁▁ |
| Bedroom\_AbvGr | 0 | 1 | 2.86 | 0.82 | 0.00 | 2.00 | 3.00 | 3.00 | 6.00 | ▁▃▇▂▁ |
| Kitchen\_AbvGr | 0 | 1 | 1.05 | 0.22 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | ▇▁▁▁▁ |
| TotRms\_AbvGrd | 0 | 1 | 6.44 | 1.54 | 3.00 | 5.00 | 6.00 | 7.00 | 15.00 | ▅▇▃▁▁ |
| Fireplaces | 0 | 1 | 0.60 | 0.65 | 0.00 | 0.00 | 1.00 | 1.00 | 4.00 | ▇▇▁▁▁ |
| Garage\_Cars | 0 | 1 | 1.77 | 0.76 | 0.00 | 1.00 | 2.00 | 2.00 | 5.00 | ▅▇▂▁▁ |
| Garage\_Area | 0 | 1 | 471.96 | 213.43 | 0.00 | 320.00 | 478.00 | 576.00 | 1488.00 | ▃▇▂▁▁ |
| Wood\_Deck\_SF | 0 | 1 | 93.52 | 127.71 | 0.00 | 0.00 | 0.00 | 168.00 | 1424.00 | ▇▁▁▁▁ |
| Open\_Porch\_SF | 0 | 1 | 48.17 | 69.51 | 0.00 | 0.00 | 27.00 | 72.00 | 742.00 | ▇▁▁▁▁ |
| Enclosed\_Porch | 0 | 1 | 23.02 | 60.59 | 0.00 | 0.00 | 0.00 | 0.00 | 584.00 | ▇▁▁▁▁ |
| Three\_season\_porch | 0 | 1 | 2.80 | 25.65 | 0.00 | 0.00 | 0.00 | 0.00 | 407.00 | ▇▁▁▁▁ |
| Screen\_Porch | 0 | 1 | 16.68 | 57.94 | 0.00 | 0.00 | 0.00 | 0.00 | 576.00 | ▇▁▁▁▁ |
| Pool\_Area | 0 | 1 | 1.34 | 27.74 | 0.00 | 0.00 | 0.00 | 0.00 | 800.00 | ▇▁▁▁▁ |
| Misc\_Val | 0 | 1 | 60.12 | 662.76 | 0.00 | 0.00 | 0.00 | 0.00 | 17000.00 | ▇▁▁▁▁ |
| Mo\_Sold | 0 | 1 | 6.19 | 2.70 | 1.00 | 4.00 | 6.00 | 8.00 | 12.00 | ▅▆▇▃▃ |
| Year\_Sold | 0 | 1 | 2007.75 | 1.30 | 2006.00 | 2007.00 | 2008.00 | 2009.00 | 2010.00 | ▇▇▇▇▃ |
| Longitude | 0 | 1 | -93.64 | 0.03 | -93.69 | -93.66 | -93.64 | -93.62 | -93.58 | ▅▅▇▇▁ |
| Latitude | 0 | 1 | 42.03 | 0.02 | 41.99 | 42.02 | 42.03 | 42.05 | 42.06 | ▂▂▇▇▇ |

## Examin summary of dataset and Perform mutation of response variable into numeric for quantitative variables corelation analysis and translate all character variables into factors.

summary(ames)

## MS\_SubClass MS\_Zoning Lot\_Frontage Lot\_Area   
## Length:2053 Length:2053 Min. : 0.00 Min. : 1300   
## Class :character Class :character 1st Qu.: 43.00 1st Qu.: 7500   
## Mode :character Mode :character Median : 62.00 Median : 9548   
## Mean : 57.38 Mean : 10258   
## 3rd Qu.: 78.00 3rd Qu.: 11600   
## Max. :313.00 Max. :215245   
## Street Alley Lot\_Shape Land\_Contour   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Utilities Lot\_Config Land\_Slope Neighborhood   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Condition\_1 Condition\_2 Bldg\_Type House\_Style   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Overall\_Qual Overall\_Cond Year\_Built Year\_Remod\_Add  
## Length:2053 Length:2053 Min. :1875 Min. :1950   
## Class :character Class :character 1st Qu.:1953 1st Qu.:1965   
## Mode :character Mode :character Median :1972 Median :1993   
## Mean :1971 Mean :1984   
## 3rd Qu.:2000 3rd Qu.:2004   
## Max. :2010 Max. :2010   
## Roof\_Style Roof\_Matl Exterior\_1st Exterior\_2nd   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual Exter\_Cond   
## Length:2053 Min. : 0.0 Length:2053 Length:2053   
## Class :character 1st Qu.: 0.0 Class :character Class :character   
## Mode :character Median : 0.0 Mode :character Mode :character   
## Mean : 103.8   
## 3rd Qu.: 164.0   
## Max. :1600.0   
## Foundation Bsmt\_Qual Bsmt\_Cond Bsmt\_Exposure   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2 BsmtFin\_SF\_2   
## Length:2053 Min. :1.00 Length:2053 Min. : 0.00   
## Class :character 1st Qu.:3.00 Class :character 1st Qu.: 0.00   
## Mode :character Median :3.00 Mode :character Median : 0.00   
## Mean :4.21 Mean : 52.57   
## 3rd Qu.:7.00 3rd Qu.: 0.00   
## Max. :7.00 Max. :1526.00   
## Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating Heating\_QC   
## Min. : 0.0 Min. : 0 Length:2053 Length:2053   
## 1st Qu.: 226.0 1st Qu.: 793 Class :character Class :character   
## Median : 460.0 Median : 988 Mode :character Mode :character   
## Mean : 561.2 Mean :1055   
## 3rd Qu.: 801.0 3rd Qu.:1304   
## Max. :2336.0 Max. :5095   
## Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## Length:2053 Length:2053 Min. : 432 Min. : 0.0   
## Class :character Class :character 1st Qu.: 882 1st Qu.: 0.0   
## Mode :character Mode :character Median :1088 Median : 0.0   
## Mean :1168 Mean : 326.1   
## 3rd Qu.:1402 3rd Qu.: 701.0   
## Max. :5095 Max. :1862.0   
## Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath   
## Min. : 0.000 Min. : 480 Min. :0.0000 Min. :0.00000   
## 1st Qu.: 0.000 1st Qu.:1137 1st Qu.:0.0000 1st Qu.:0.00000   
## Median : 0.000 Median :1447 Median :0.0000 Median :0.00000   
## Mean : 4.973 Mean :1499 Mean :0.4301 Mean :0.05796   
## 3rd Qu.: 0.000 3rd Qu.:1737 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1064.000 Max. :5095 Max. :3.0000 Max. :2.00000   
## Full\_Bath Half\_Bath Bedroom\_AbvGr Kitchen\_AbvGr   
## Min. :0.000 Min. :0.0000 Min. :0.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:1.000   
## Median :2.000 Median :0.0000 Median :3.000 Median :1.000   
## Mean :1.564 Mean :0.3751 Mean :2.855 Mean :1.047   
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.:1.000   
## Max. :4.000 Max. :2.0000 Max. :6.000 Max. :3.000   
## Kitchen\_Qual TotRms\_AbvGrd Functional Fireplaces   
## Length:2053 Min. : 3.000 Length:2053 Min. :0.000   
## Class :character 1st Qu.: 5.000 Class :character 1st Qu.:0.000   
## Mode :character Median : 6.000 Mode :character Median :1.000   
## Mean : 6.442 Mean :0.603   
## 3rd Qu.: 7.000 3rd Qu.:1.000   
## Max. :15.000 Max. :4.000   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## Length:2053 Length:2053 Length:2053 Min. :0.000   
## Class :character Class :character Class :character 1st Qu.:1.000   
## Mode :character Mode :character Mode :character Median :2.000   
## Mean :1.774   
## 3rd Qu.:2.000   
## Max. :5.000   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## Min. : 0 Length:2053 Length:2053 Length:2053   
## 1st Qu.: 320 Class :character Class :character Class :character   
## Median : 478 Mode :character Mode :character Mode :character   
## Mean : 472   
## 3rd Qu.: 576   
## Max. :1488   
## Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch Three\_season\_porch  
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 27.00 Median : 0.00 Median : 0.000   
## Mean : 93.52 Mean : 48.17 Mean : 23.02 Mean : 2.799   
## 3rd Qu.: 168.00 3rd Qu.: 72.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :1424.00 Max. :742.00 Max. :584.00 Max. :407.000   
## Screen\_Porch Pool\_Area Pool\_QC Fence   
## Min. : 0.00 Min. : 0.000 Length:2053 Length:2053   
## 1st Qu.: 0.00 1st Qu.: 0.000 Class :character Class :character   
## Median : 0.00 Median : 0.000 Mode :character Mode :character   
## Mean : 16.68 Mean : 1.339   
## 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :576.00 Max. :800.000   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold   
## Length:2053 Min. : 0.00 Min. : 1.000 Min. :2006   
## Class :character 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007   
## Mode :character Median : 0.00 Median : 6.000 Median :2008   
## Mean : 60.12 Mean : 6.189 Mean :2008   
## 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009   
## Max. :17000.00 Max. :12.000 Max. :2010   
## Sale\_Type Sale\_Condition Longitude Latitude   
## Length:2053 Length:2053 Min. :-93.69 Min. :41.99   
## Class :character Class :character 1st Qu.:-93.66 1st Qu.:42.02   
## Mode :character Mode :character 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   
## Above\_Median   
## Length:2053   
## Class :character   
## Mode :character   
##   
##   
##

ames2= ames %>%  
 mutate(Above\_Median\_Numeric = ifelse(Above\_Median == "Yes", 1, 0)) # to estimate correlation with quantitative or continuous variables.  
#ames2 = ames2 %>% filter(ames$Year\_Built >= 1965) # Decided to consume full data  
  
ames2 = ames2 %>% mutate\_if(is.character, as\_factor) # translate all character variables to factor variables.

## Examin dataset after translation and mutation.

summary(ames2)

## MS\_SubClass MS\_Zoning   
## One\_Story\_1946\_and\_Newer\_All\_Styles :772 Residential\_Low\_Density :1600   
## Two\_Story\_1946\_and\_Newer :383 Residential\_High\_Density : 20   
## One\_and\_Half\_Story\_Finished\_All\_Ages:204 Floating\_Village\_Residential: 87   
## One\_Story\_PUD\_1946\_and\_Newer :129 Residential\_Medium\_Density : 326   
## One\_Story\_1945\_and\_Older : 98 C\_all : 17   
## Two\_Story\_1945\_and\_Older : 95 A\_agr : 2   
## (Other) :372 I\_all : 1   
## Lot\_Frontage Lot\_Area Street Alley   
## Min. : 0.00 Min. : 1300 Pave:2046 No\_Alley\_Access:1914   
## 1st Qu.: 43.00 1st Qu.: 7500 Grvl: 7 Paved : 45   
## Median : 62.00 Median : 9548 Gravel : 94   
## Mean : 57.38 Mean : 10258   
## 3rd Qu.: 78.00 3rd Qu.: 11600   
## Max. :313.00 Max. :215245   
##   
## Lot\_Shape Land\_Contour Utilities Lot\_Config   
## Slightly\_Irregular : 714 Lvl:1833 AllPub:2052 Corner : 359   
## Regular :1275 HLS: 94 NoSewr: 1 Inside :1495   
## Moderately\_Irregular: 53 Bnk: 81 CulDSac: 135   
## Irregular : 11 Low: 45 FR2 : 56   
## FR3 : 8   
##   
##   
## Land\_Slope Neighborhood Condition\_1 Condition\_2 Bldg\_Type   
## Gtl:1951 North\_Ames : 327 Norm :1771 Norm :2027 OneFam :1706   
## Mod: 89 College\_Creek: 183 Feedr : 113 Feedr : 12 TwnhsE : 157   
## Sev: 13 Old\_Town : 181 Artery : 67 PosA : 4 Twnhs : 67   
## Edwards : 129 RRAn : 35 Artery : 4 Duplex : 76   
## Somerset : 119 PosN : 24 PosN : 3 TwoFmCon: 47   
## Gilbert : 109 RRAe : 19 RRNn : 1   
## (Other) :1005 (Other): 24 (Other): 2   
## House\_Style Overall\_Qual Overall\_Cond   
## One\_Story :1052 Average :587 Average :1143   
## Two\_Story : 590 Above\_Average:518 Above\_Average: 376   
## One\_and\_Half\_Fin: 225 Good :411 Good : 286   
## SLvl : 90 Very\_Good :237 Very\_Good : 98   
## SFoyer : 56 Below\_Average:169 Below\_Average: 73   
## Two\_and\_Half\_Unf: 19 Excellent : 70 Fair : 35   
## (Other) : 21 (Other) : 61 (Other) : 42   
## Year\_Built Year\_Remod\_Add Roof\_Style Roof\_Matl Exterior\_1st  
## Min. :1875 Min. :1950 Hip : 404 CompShg:2023 VinylSd:705   
## 1st Qu.:1953 1st Qu.:1965 Gable :1607 WdShake: 8 MetalSd:319   
## Median :1972 Median :1993 Mansard: 9 Tar&Grv: 17 Wd Sdng:313   
## Mean :1971 Mean :1984 Gambrel: 14 WdShngl: 3 HdBoard:303   
## 3rd Qu.:2000 3rd Qu.:2004 Shed : 5 Roll : 1 Plywood:151   
## Max. :2010 Max. :2010 Flat : 14 Metal : 1 CemntBd: 90   
## (Other):172   
## Exterior\_2nd Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual   
## VinylSd:699 Stone : 166 Min. : 0.0 Typical :1272   
## MetalSd:317 None :1231 1st Qu.: 0.0 Good : 682   
## Wd Sdng:302 BrkFace: 638 Median : 0.0 Excellent: 78   
## HdBoard:277 BrkCmn : 17 Mean : 103.8 Fair : 21   
## Plywood:190 CBlock : 1 3rd Qu.: 164.0   
## CmentBd: 90 Max. :1600.0   
## (Other):178   
## Exter\_Cond Foundation Bsmt\_Qual Bsmt\_Cond   
## Typical :1787 CBlock:880 Typical :911 Good : 80   
## Good : 213 PConc :911 Good :849 Typical :1833   
## Fair : 43 Wood : 4 Excellent :178 Poor : 4   
## Excellent: 9 BrkTil:216 No\_Basement: 57 No\_Basement: 57   
## Poor : 1 Slab : 36 Fair : 57 Fair : 76   
## Stone : 6 Poor : 1 Excellent : 3   
##   
## Bsmt\_Exposure BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2  
## Gd : 199 BLQ :196 Min. :1.00 Unf :1740   
## No :1331 Rec :216 1st Qu.:3.00 LwQ : 64   
## Av : 284 ALQ :298 Median :3.00 BLQ : 47   
## Mn : 179 GLQ :578 Mean :4.21 Rec : 79   
## No\_Basement: 60 Unf :602 3rd Qu.:7.00 GLQ : 23   
## LwQ :106 Max. :7.00 No\_Basement: 58   
## No\_Basement: 57 ALQ : 42   
## BsmtFin\_SF\_2 Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating   
## Min. : 0.00 Min. : 0.0 Min. : 0 GasA :2019   
## 1st Qu.: 0.00 1st Qu.: 226.0 1st Qu.: 793 GasW : 21   
## Median : 0.00 Median : 460.0 Median : 988 Grav : 6   
## Mean : 52.57 Mean : 561.2 Mean :1055 Wall : 5   
## 3rd Qu.: 0.00 3rd Qu.: 801.0 3rd Qu.:1304 Floor: 1   
## Max. :1526.00 Max. :2336.0 Max. :5095 OthW : 1   
##   
## Heating\_QC Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## Fair : 61 Y:1916 SBrkr :1887 Min. : 432 Min. : 0.0   
## Typical : 618 N: 137 FuseA : 126 1st Qu.: 882 1st Qu.: 0.0   
## Excellent:1040 FuseF : 33 Median :1088 Median : 0.0   
## Good : 333 FuseP : 6 Mean :1168 Mean : 326.1   
## Poor : 1 Unknown: 1 3rd Qu.:1402 3rd Qu.: 701.0   
## Max. :5095 Max. :1862.0   
##   
## Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath   
## Min. : 0.000 Min. : 480 Min. :0.0000 Min. :0.00000   
## 1st Qu.: 0.000 1st Qu.:1137 1st Qu.:0.0000 1st Qu.:0.00000   
## Median : 0.000 Median :1447 Median :0.0000 Median :0.00000   
## Mean : 4.973 Mean :1499 Mean :0.4301 Mean :0.05796   
## 3rd Qu.: 0.000 3rd Qu.:1737 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1064.000 Max. :5095 Max. :3.0000 Max. :2.00000   
##   
## Full\_Bath Half\_Bath Bedroom\_AbvGr Kitchen\_AbvGr   
## Min. :0.000 Min. :0.0000 Min. :0.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:1.000   
## Median :2.000 Median :0.0000 Median :3.000 Median :1.000   
## Mean :1.564 Mean :0.3751 Mean :2.855 Mean :1.047   
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.:1.000   
## Max. :4.000 Max. :2.0000 Max. :6.000 Max. :3.000   
##   
## Kitchen\_Qual TotRms\_AbvGrd Functional Fireplaces   
## Typical :1070 Min. : 3.000 Typ :1896 Min. :0.000   
## Good : 790 1st Qu.: 5.000 Min2 : 54 1st Qu.:0.000   
## Excellent: 142 Median : 6.000 Min1 : 51 Median :1.000   
## Fair : 50 Mean : 6.442 Mod : 27 Mean :0.603   
## Poor : 1 3rd Qu.: 7.000 Maj1 : 15 3rd Qu.:1.000   
## Max. :15.000 Maj2 : 6 Max. :4.000   
## (Other): 4   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## Good :538 Attchd :1204 Fin :509 Min. :0.000   
## No\_Fireplace:993 BuiltIn : 127 Unf :872 1st Qu.:1.000   
## Typical :409 Basment : 29 RFn :563 Median :2.000   
## Poor : 36 Detchd : 549 No\_Garage:109 Mean :1.774   
## Excellent : 21 No\_Garage : 108 3rd Qu.:2.000   
## Fair : 56 CarPort : 15 Max. :5.000   
## More\_Than\_Two\_Types: 21   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## Min. : 0 Typical :1839 Typical :1872 Partial\_Pavement: 42   
## 1st Qu.: 320 No\_Garage: 109 No\_Garage: 109 Paved :1848   
## Median : 478 Fair : 85 Fair : 53 Dirt\_Gravel : 163   
## Mean : 472 Good : 16 Excellent: 1   
## 3rd Qu.: 576 Excellent: 2 Poor : 8   
## Max. :1488 Poor : 2 Good : 10   
##   
## Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch Three\_season\_porch  
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 27.00 Median : 0.00 Median : 0.000   
## Mean : 93.52 Mean : 48.17 Mean : 23.02 Mean : 2.799   
## 3rd Qu.: 168.00 3rd Qu.: 72.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :1424.00 Max. :742.00 Max. :584.00 Max. :407.000   
##   
## Screen\_Porch Pool\_Area Pool\_QC Fence   
## Min. : 0.00 Min. : 0.000 No\_Pool :2047 No\_Fence :1661   
## 1st Qu.: 0.00 1st Qu.: 0.000 Excellent: 2 Minimum\_Privacy : 225   
## Median : 0.00 Median : 0.000 Typical : 2 Good\_Privacy : 81   
## Mean : 16.68 Mean : 1.339 Fair : 1 Good\_Wood : 77   
## 3rd Qu.: 0.00 3rd Qu.: 0.000 Good : 1 Minimum\_Wood\_Wire: 9   
## Max. :576.00 Max. :800.000   
##   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold Sale\_Type   
## None:1978 Min. : 0.00 Min. : 1.000 Min. :2006 WD :1789   
## Gar2: 5 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007 New : 163   
## Shed: 66 Median : 0.00 Median : 6.000 Median :2008 COD : 54   
## Othr: 3 Mean : 60.12 Mean : 6.189 Mean :2008 ConLD : 16   
## Elev: 1 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009 ConLI : 8   
## Max. :17000.00 Max. :12.000 Max. :2010 CWD : 8   
## (Other): 15   
## Sale\_Condition Longitude Latitude Above\_Median  
## Normal :1712 Min. :-93.69 Min. :41.99 Yes:1043   
## Partial: 169 1st Qu.:-93.66 1st Qu.:42.02 No :1010   
## Family : 30 Median :-93.64 Median :42.03   
## Abnorml: 121 Mean :-93.64 Mean :42.03   
## Alloca : 16 3rd Qu.:-93.62 3rd Qu.:42.05   
## AdjLand: 5 Max. :-93.58 Max. :42.06   
##   
## Above\_Median\_Numeric  
## Min. :0.000   
## 1st Qu.:0.000   
## Median :1.000   
## Mean :0.508   
## 3rd Qu.:1.000   
## Max. :1.000   
##

skim(ames2)

Data summary

|  |  |
| --- | --- |
| Name | ames2 |
| Number of rows | 2053 |
| Number of columns | 82 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 47 |
| numeric | 35 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| MS\_SubClass | 0 | 1 | FALSE | 16 | One: 772, Two: 383, One: 204, One: 129 |
| MS\_Zoning | 0 | 1 | FALSE | 7 | Res: 1600, Res: 326, Flo: 87, Res: 20 |
| Street | 0 | 1 | FALSE | 2 | Pav: 2046, Grv: 7 |
| Alley | 0 | 1 | FALSE | 3 | No\_: 1914, Gra: 94, Pav: 45 |
| Lot\_Shape | 0 | 1 | FALSE | 4 | Reg: 1275, Sli: 714, Mod: 53, Irr: 11 |
| Land\_Contour | 0 | 1 | FALSE | 4 | Lvl: 1833, HLS: 94, Bnk: 81, Low: 45 |
| Utilities | 0 | 1 | FALSE | 2 | All: 2052, NoS: 1 |
| Lot\_Config | 0 | 1 | FALSE | 5 | Ins: 1495, Cor: 359, Cul: 135, FR2: 56 |
| Land\_Slope | 0 | 1 | FALSE | 3 | Gtl: 1951, Mod: 89, Sev: 13 |
| Neighborhood | 0 | 1 | FALSE | 28 | Nor: 327, Col: 183, Old: 181, Edw: 129 |
| Condition\_1 | 0 | 1 | FALSE | 9 | Nor: 1771, Fee: 113, Art: 67, RRA: 35 |
| Condition\_2 | 0 | 1 | FALSE | 8 | Nor: 2027, Fee: 12, Pos: 4, Art: 4 |
| Bldg\_Type | 0 | 1 | FALSE | 5 | One: 1706, Twn: 157, Dup: 76, Twn: 67 |
| House\_Style | 0 | 1 | FALSE | 8 | One: 1052, Two: 590, One: 225, SLv: 90 |
| Overall\_Qual | 0 | 1 | FALSE | 10 | Ave: 587, Abo: 518, Goo: 411, Ver: 237 |
| Overall\_Cond | 0 | 1 | FALSE | 9 | Ave: 1143, Abo: 376, Goo: 286, Ver: 98 |
| Roof\_Style | 0 | 1 | FALSE | 6 | Gab: 1607, Hip: 404, Gam: 14, Fla: 14 |
| Roof\_Matl | 0 | 1 | FALSE | 6 | Com: 2023, Tar: 17, WdS: 8, WdS: 3 |
| Exterior\_1st | 0 | 1 | FALSE | 16 | Vin: 705, Met: 319, Wd : 313, HdB: 303 |
| Exterior\_2nd | 0 | 1 | FALSE | 17 | Vin: 699, Met: 317, Wd : 302, HdB: 277 |
| Mas\_Vnr\_Type | 0 | 1 | FALSE | 5 | Non: 1231, Brk: 638, Sto: 166, Brk: 17 |
| Exter\_Qual | 0 | 1 | FALSE | 4 | Typ: 1272, Goo: 682, Exc: 78, Fai: 21 |
| Exter\_Cond | 0 | 1 | FALSE | 5 | Typ: 1787, Goo: 213, Fai: 43, Exc: 9 |
| Foundation | 0 | 1 | FALSE | 6 | PCo: 911, CBl: 880, Brk: 216, Sla: 36 |
| Bsmt\_Qual | 0 | 1 | FALSE | 6 | Typ: 911, Goo: 849, Exc: 178, No\_: 57 |
| Bsmt\_Cond | 0 | 1 | FALSE | 6 | Typ: 1833, Goo: 80, Fai: 76, No\_: 57 |
| Bsmt\_Exposure | 0 | 1 | FALSE | 5 | No: 1331, Av: 284, Gd: 199, Mn: 179 |
| BsmtFin\_Type\_1 | 0 | 1 | FALSE | 7 | Unf: 602, GLQ: 578, ALQ: 298, Rec: 216 |
| BsmtFin\_Type\_2 | 0 | 1 | FALSE | 7 | Unf: 1740, Rec: 79, LwQ: 64, No\_: 58 |
| Heating | 0 | 1 | FALSE | 6 | Gas: 2019, Gas: 21, Gra: 6, Wal: 5 |
| Heating\_QC | 0 | 1 | FALSE | 5 | Exc: 1040, Typ: 618, Goo: 333, Fai: 61 |
| Central\_Air | 0 | 1 | FALSE | 2 | Y: 1916, N: 137 |
| Electrical | 0 | 1 | FALSE | 5 | SBr: 1887, Fus: 126, Fus: 33, Fus: 6 |
| Kitchen\_Qual | 0 | 1 | FALSE | 5 | Typ: 1070, Goo: 790, Exc: 142, Fai: 50 |
| Functional | 0 | 1 | FALSE | 8 | Typ: 1896, Min: 54, Min: 51, Mod: 27 |
| Fireplace\_Qu | 0 | 1 | FALSE | 6 | No\_: 993, Goo: 538, Typ: 409, Fai: 56 |
| Garage\_Type | 0 | 1 | FALSE | 7 | Att: 1204, Det: 549, Bui: 127, No\_: 108 |
| Garage\_Finish | 0 | 1 | FALSE | 4 | Unf: 872, RFn: 563, Fin: 509, No\_: 109 |
| Garage\_Qual | 0 | 1 | FALSE | 6 | Typ: 1839, No\_: 109, Fai: 85, Goo: 16 |
| Garage\_Cond | 0 | 1 | FALSE | 6 | Typ: 1872, No\_: 109, Fai: 53, Goo: 10 |
| Paved\_Drive | 0 | 1 | FALSE | 3 | Pav: 1848, Dir: 163, Par: 42 |
| Pool\_QC | 0 | 1 | FALSE | 5 | No\_: 2047, Exc: 2, Typ: 2, Fai: 1 |
| Fence | 0 | 1 | FALSE | 5 | No\_: 1661, Min: 225, Goo: 81, Goo: 77 |
| Misc\_Feature | 0 | 1 | FALSE | 5 | Non: 1978, She: 66, Gar: 5, Oth: 3 |
| Sale\_Type | 0 | 1 | FALSE | 10 | WD: 1789, New: 163, COD: 54, Con: 16 |
| Sale\_Condition | 0 | 1 | FALSE | 6 | Nor: 1712, Par: 169, Abn: 121, Fam: 30 |
| Above\_Median | 0 | 1 | FALSE | 2 | Yes: 1043, No: 1010 |

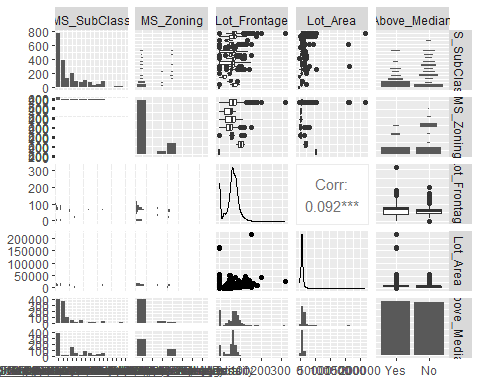
**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Lot\_Frontage | 0 | 1 | 57.38 | 33.20 | 0.00 | 43.00 | 62.00 | 78.00 | 313.00 | ▇▇▁▁▁ |
| Lot\_Area | 0 | 1 | 10258.40 | 8427.38 | 1300.00 | 7500.00 | 9548.00 | 11600.00 | 215245.00 | ▇▁▁▁▁ |
| Year\_Built | 0 | 1 | 1970.64 | 30.40 | 1875.00 | 1953.00 | 1972.00 | 2000.00 | 2010.00 | ▁▂▃▆▇ |
| Year\_Remod\_Add | 0 | 1 | 1984.08 | 20.96 | 1950.00 | 1965.00 | 1993.00 | 2004.00 | 2010.00 | ▅▂▂▃▇ |
| Mas\_Vnr\_Area | 0 | 1 | 103.75 | 183.59 | 0.00 | 0.00 | 0.00 | 164.00 | 1600.00 | ▇▁▁▁▁ |
| BsmtFin\_SF\_1 | 0 | 1 | 4.21 | 2.24 | 1.00 | 3.00 | 3.00 | 7.00 | 7.00 | ▅▆▁▁▇ |
| BsmtFin\_SF\_2 | 0 | 1 | 52.57 | 175.99 | 0.00 | 0.00 | 0.00 | 0.00 | 1526.00 | ▇▁▁▁▁ |
| Bsmt\_Unf\_SF | 0 | 1 | 561.19 | 441.72 | 0.00 | 226.00 | 460.00 | 801.00 | 2336.00 | ▇▅▂▁▁ |
| Total\_Bsmt\_SF | 0 | 1 | 1054.57 | 435.33 | 0.00 | 793.00 | 988.00 | 1304.00 | 5095.00 | ▇▇▁▁▁ |
| First\_Flr\_SF | 0 | 1 | 1167.52 | 391.79 | 432.00 | 882.00 | 1088.00 | 1402.00 | 5095.00 | ▇▃▁▁▁ |
| Second\_Flr\_SF | 0 | 1 | 326.07 | 422.44 | 0.00 | 0.00 | 0.00 | 701.00 | 1862.00 | ▇▂▂▁▁ |
| Low\_Qual\_Fin\_SF | 0 | 1 | 4.97 | 49.09 | 0.00 | 0.00 | 0.00 | 0.00 | 1064.00 | ▇▁▁▁▁ |
| Gr\_Liv\_Area | 0 | 1 | 1498.56 | 487.84 | 480.00 | 1137.00 | 1447.00 | 1737.00 | 5095.00 | ▇▇▁▁▁ |
| Bsmt\_Full\_Bath | 0 | 1 | 0.43 | 0.53 | 0.00 | 0.00 | 0.00 | 1.00 | 3.00 | ▇▆▁▁▁ |
| Bsmt\_Half\_Bath | 0 | 1 | 0.06 | 0.24 | 0.00 | 0.00 | 0.00 | 0.00 | 2.00 | ▇▁▁▁▁ |
| Full\_Bath | 0 | 1 | 1.56 | 0.55 | 0.00 | 1.00 | 2.00 | 2.00 | 4.00 | ▁▇▇▁▁ |
| Half\_Bath | 0 | 1 | 0.38 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 2.00 | ▇▁▅▁▁ |
| Bedroom\_AbvGr | 0 | 1 | 2.86 | 0.82 | 0.00 | 2.00 | 3.00 | 3.00 | 6.00 | ▁▃▇▂▁ |
| Kitchen\_AbvGr | 0 | 1 | 1.05 | 0.22 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | ▇▁▁▁▁ |
| TotRms\_AbvGrd | 0 | 1 | 6.44 | 1.54 | 3.00 | 5.00 | 6.00 | 7.00 | 15.00 | ▅▇▃▁▁ |
| Fireplaces | 0 | 1 | 0.60 | 0.65 | 0.00 | 0.00 | 1.00 | 1.00 | 4.00 | ▇▇▁▁▁ |
| Garage\_Cars | 0 | 1 | 1.77 | 0.76 | 0.00 | 1.00 | 2.00 | 2.00 | 5.00 | ▅▇▂▁▁ |
| Garage\_Area | 0 | 1 | 471.96 | 213.43 | 0.00 | 320.00 | 478.00 | 576.00 | 1488.00 | ▃▇▂▁▁ |
| Wood\_Deck\_SF | 0 | 1 | 93.52 | 127.71 | 0.00 | 0.00 | 0.00 | 168.00 | 1424.00 | ▇▁▁▁▁ |
| Open\_Porch\_SF | 0 | 1 | 48.17 | 69.51 | 0.00 | 0.00 | 27.00 | 72.00 | 742.00 | ▇▁▁▁▁ |
| Enclosed\_Porch | 0 | 1 | 23.02 | 60.59 | 0.00 | 0.00 | 0.00 | 0.00 | 584.00 | ▇▁▁▁▁ |
| Three\_season\_porch | 0 | 1 | 2.80 | 25.65 | 0.00 | 0.00 | 0.00 | 0.00 | 407.00 | ▇▁▁▁▁ |
| Screen\_Porch | 0 | 1 | 16.68 | 57.94 | 0.00 | 0.00 | 0.00 | 0.00 | 576.00 | ▇▁▁▁▁ |
| Pool\_Area | 0 | 1 | 1.34 | 27.74 | 0.00 | 0.00 | 0.00 | 0.00 | 800.00 | ▇▁▁▁▁ |
| Misc\_Val | 0 | 1 | 60.12 | 662.76 | 0.00 | 0.00 | 0.00 | 0.00 | 17000.00 | ▇▁▁▁▁ |
| Mo\_Sold | 0 | 1 | 6.19 | 2.70 | 1.00 | 4.00 | 6.00 | 8.00 | 12.00 | ▅▆▇▃▃ |
| Year\_Sold | 0 | 1 | 2007.75 | 1.30 | 2006.00 | 2007.00 | 2008.00 | 2009.00 | 2010.00 | ▇▇▇▇▃ |
| Longitude | 0 | 1 | -93.64 | 0.03 | -93.69 | -93.66 | -93.64 | -93.62 | -93.58 | ▅▅▇▇▁ |
| Latitude | 0 | 1 | 42.03 | 0.02 | 41.99 | 42.02 | 42.03 | 42.05 | 42.06 | ▂▂▇▇▇ |
| Above\_Median\_Numeric | 0 | 1 | 0.51 | 0.50 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | ▇▁▁▁▇ |

## Visualize relationsship of all variables with response variable Above\_Median to identify required variables for future model with expected positive relationship

ggpairs(ames2, columns = c("MS\_SubClass", "MS\_Zoning", "Lot\_Frontage", "Lot\_Area","Above\_Median"),cardinality\_threshold = 50)

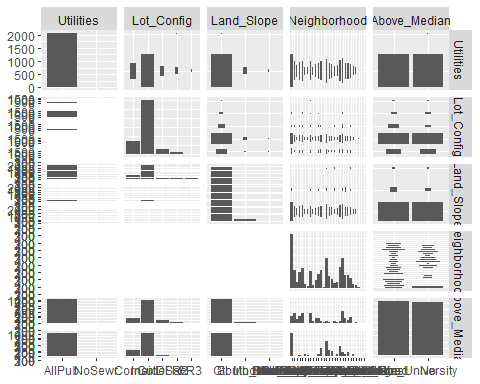
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



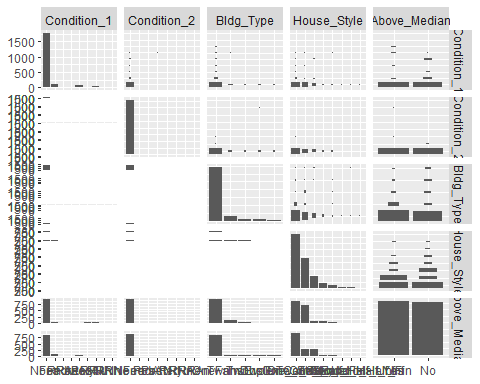
ggpairs(ames2, columns = c("Street", "Alley", "Lot\_Shape", "Land\_Contour","Above\_Median"))



ggpairs(ames2, columns = c("Utilities", "Lot\_Config", "Land\_Slope", "Neighborhood","Above\_Median"),cardinality\_threshold = 50)

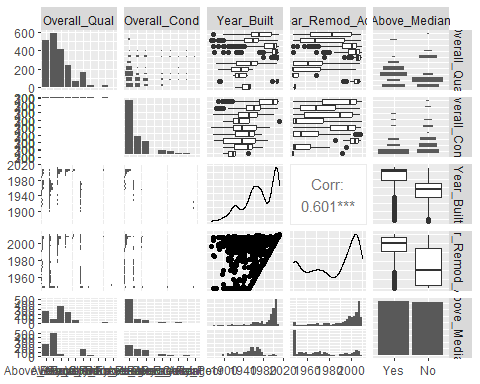


ggpairs(ames2, columns = c("Condition\_1", "Condition\_2", "Bldg\_Type", "House\_Style","Above\_Median"))

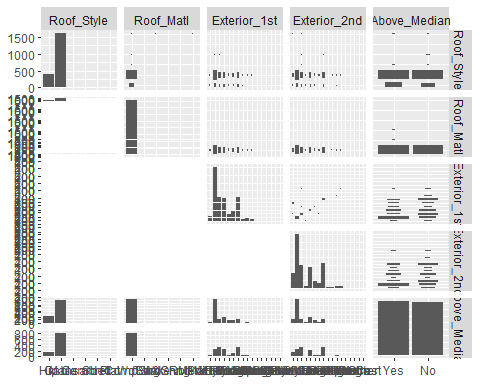


ggpairs(ames2, columns = c("Overall\_Qual", "Overall\_Cond", "Year\_Built", "Year\_Remod\_Add","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

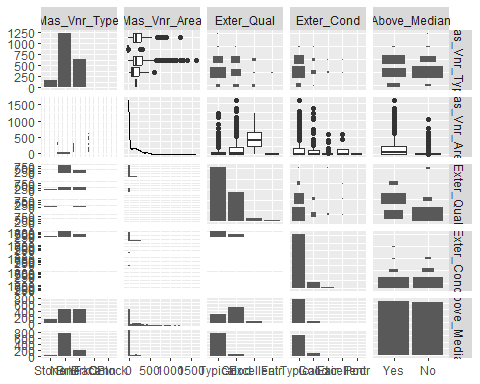


ggpairs(ames2, columns = c("Roof\_Style", "Roof\_Matl", "Exterior\_1st", "Exterior\_2nd","Above\_Median"),cardinality\_threshold = 50)

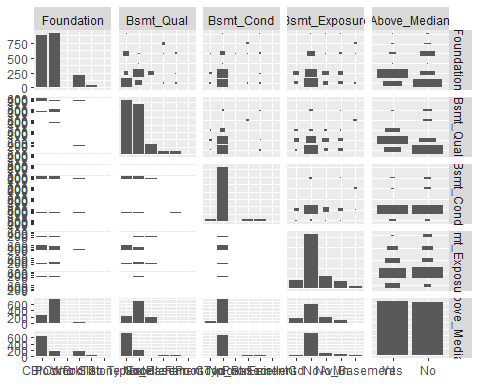


ggpairs(ames2, columns = c("Mas\_Vnr\_Type", "Mas\_Vnr\_Area", "Exter\_Qual", "Exter\_Cond","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

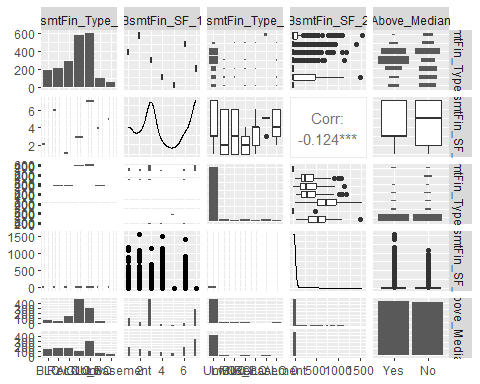


ggpairs(ames2, columns = c("Foundation", "Bsmt\_Qual", "Bsmt\_Cond", "Bsmt\_Exposure","Above\_Median"))



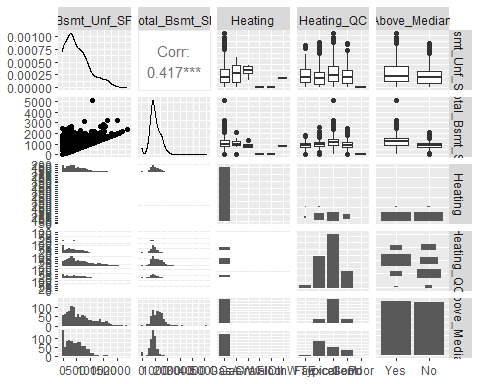
ggpairs(ames2, columns = c("BsmtFin\_Type\_1", "BsmtFin\_SF\_1", "BsmtFin\_Type\_2", "BsmtFin\_SF\_2","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



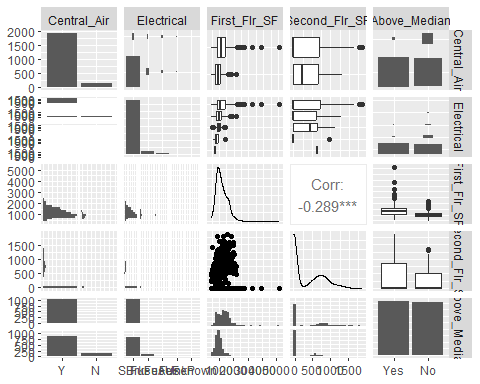
ggpairs(ames2, columns = c("Bsmt\_Unf\_SF", "Total\_Bsmt\_SF", "Heating", "Heating\_QC","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



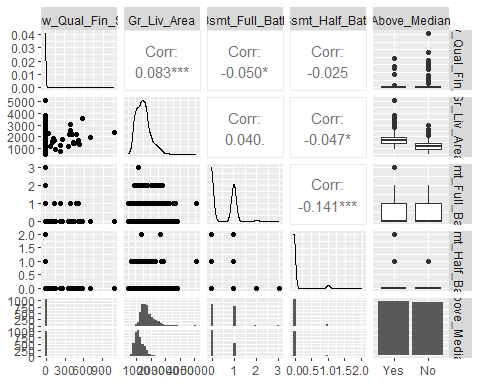
ggpairs(ames2, columns = c("Central\_Air", "Electrical", "First\_Flr\_SF", "Second\_Flr\_SF","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



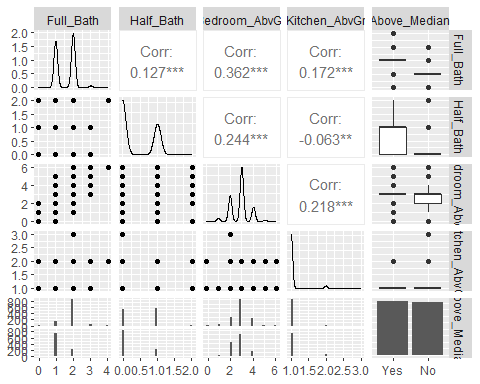
ggpairs(ames2, columns = c("Low\_Qual\_Fin\_SF", "Gr\_Liv\_Area", "Bsmt\_Full\_Bath", "Bsmt\_Half\_Bath","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



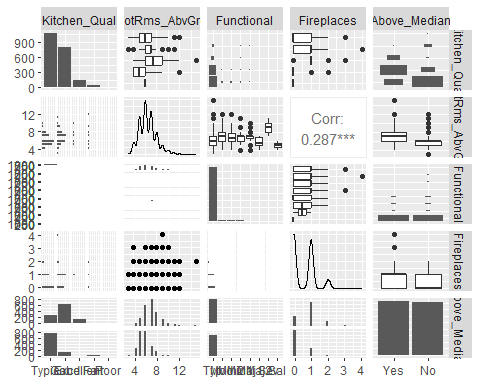
ggpairs(ames2, columns = c("Full\_Bath", "Half\_Bath", "Bedroom\_AbvGr", "Kitchen\_AbvGr","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



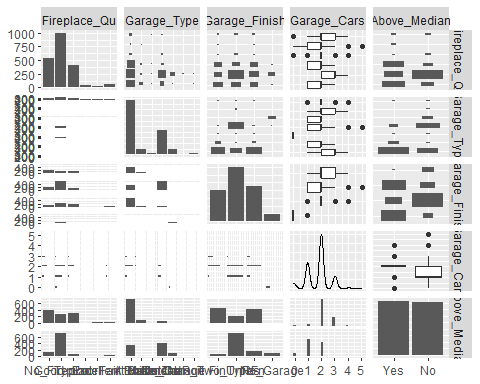
ggpairs(ames2, columns = c("Kitchen\_Qual", "TotRms\_AbvGrd", "Functional", "Fireplaces","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



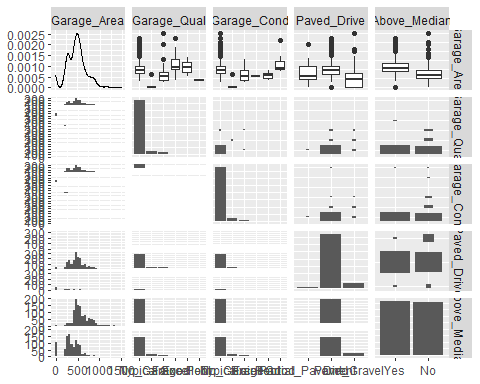
ggpairs(ames2, columns = c("Fireplace\_Qu", "Garage\_Type", "Garage\_Finish", "Garage\_Cars","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



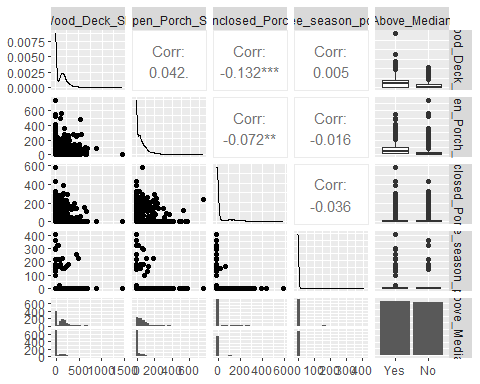
ggpairs(ames2, columns = c("Garage\_Area", "Garage\_Qual", "Garage\_Cond", "Paved\_Drive","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



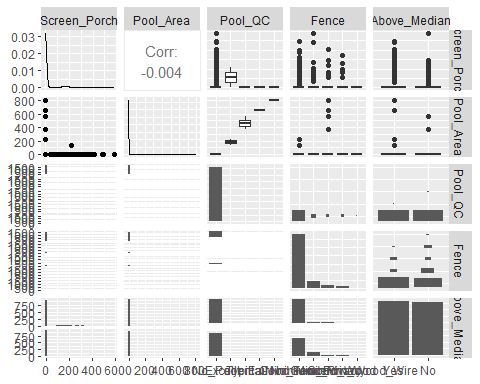
ggpairs(ames2, columns = c("Wood\_Deck\_SF", "Open\_Porch\_SF", "Enclosed\_Porch", "Three\_season\_porch","Above\_Median"),cardinality\_threshold = 251)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



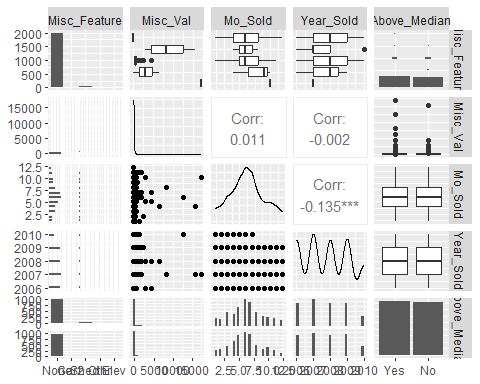
ggpairs(ames2, columns = c("Screen\_Porch", "Pool\_Area", "Pool\_QC", "Fence","Above\_Median"),cardinality\_threshold = 251)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



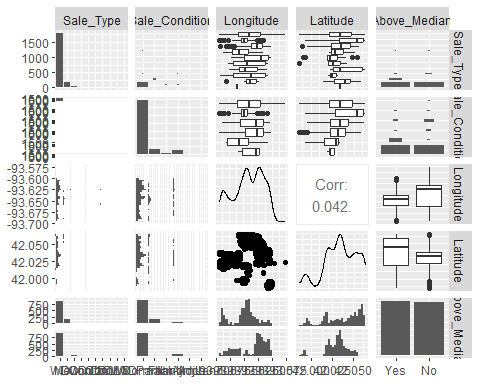
ggpairs(ames2, columns = c("Misc\_Feature", "Misc\_Val", "Mo\_Sold", "Year\_Sold","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggpairs(ames2, columns = c("Sale\_Type", "Sale\_Condition", "Longitude", "Latitude","Above\_Median"))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 ### Analyze all variables and segrigating in two differfent groups to examin them as factor or quantitative variables.

ames3= ames2[, sapply(ames2, is.factor)]  
ames4= ames2[, !sapply(ames2, is.factor)]

### Analyze contineous quantitative variables with closer look on data patterns

#colnames(ames4)  
glimpse((ames4))

## Rows: 2,053  
## Columns: 35  
## $ Lot\_Frontage <dbl> 141, 80, 81, 93, 74, 78, 43, 39, 0, 85, 0, 47, 15…  
## $ Lot\_Area <dbl> 31770, 11622, 14267, 11160, 13830, 9978, 5005, 53…  
## $ Year\_Built <dbl> 1960, 1961, 1958, 1968, 1997, 1998, 1992, 1995, 1…  
## $ Year\_Remod\_Add <dbl> 1960, 1961, 1958, 1968, 1998, 1998, 1992, 1996, 2…  
## $ Mas\_Vnr\_Area <dbl> 112, 0, 108, 0, 0, 20, 0, 0, 0, 0, 0, 603, 0, 350…  
## $ BsmtFin\_SF\_1 <dbl> 2, 6, 1, 1, 3, 3, 1, 3, 1, 3, 3, 1, 3, 3, 2, 3, 1…  
## $ BsmtFin\_SF\_2 <dbl> 0, 144, 0, 0, 0, 0, 0, 0, 0, 0, 1120, 0, 0, 0, 0,…  
## $ Bsmt\_Unf\_SF <dbl> 441, 270, 406, 1045, 137, 324, 1017, 415, 233, 66…  
## $ Total\_Bsmt\_SF <dbl> 1080, 882, 1329, 2110, 928, 926, 1280, 1595, 1168…  
## $ First\_Flr\_SF <dbl> 1656, 896, 1329, 2110, 928, 926, 1280, 1616, 1187…  
## $ Second\_Flr\_SF <dbl> 0, 0, 0, 0, 701, 678, 0, 0, 0, 0, 0, 1589, 672, 0…  
## $ Low\_Qual\_Fin\_SF <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ Gr\_Liv\_Area <dbl> 1656, 896, 1329, 2110, 1629, 1604, 1280, 1616, 11…  
## $ Bsmt\_Full\_Bath <dbl> 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1…  
## $ Bsmt\_Half\_Bath <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ Full\_Bath <dbl> 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 1, 3, 2, 1, 2, 2, 1…  
## $ Half\_Bath <dbl> 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0…  
## $ Bedroom\_AbvGr <dbl> 3, 2, 3, 3, 3, 3, 2, 2, 3, 2, 1, 4, 4, 1, 3, 3, 2…  
## $ Kitchen\_AbvGr <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…  
## $ TotRms\_AbvGrd <dbl> 7, 5, 6, 8, 6, 7, 5, 5, 6, 5, 4, 12, 8, 8, 7, 7, …  
## $ Fireplaces <dbl> 2, 0, 0, 2, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1…  
## $ Garage\_Cars <dbl> 2, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 3, 2, 2, 2…  
## $ Garage\_Area <dbl> 528, 730, 312, 522, 482, 470, 506, 608, 420, 506,…  
## $ Wood\_Deck\_SF <dbl> 210, 140, 393, 0, 212, 360, 0, 237, 483, 192, 0, …  
## $ Open\_Porch\_SF <dbl> 62, 0, 36, 0, 34, 36, 82, 152, 21, 0, 54, 36, 12,…  
## $ Enclosed\_Porch <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ Three\_season\_porch <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ Screen\_Porch <dbl> 0, 120, 0, 0, 0, 0, 144, 0, 0, 0, 140, 210, 0, 0,…  
## $ Pool\_Area <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ Misc\_Val <dbl> 0, 0, 12500, 0, 0, 0, 0, 0, 500, 0, 0, 0, 0, 0, 0…  
## $ Mo\_Sold <dbl> 5, 6, 6, 4, 3, 6, 1, 3, 3, 2, 6, 6, 6, 6, 1, 1, 3…  
## $ Year\_Sold <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2…  
## $ Longitude <dbl> -93.61975, -93.61976, -93.61939, -93.61732, -93.6…  
## $ Latitude <dbl> 42.05403, 42.05301, 42.05266, 42.05125, 42.06090,…  
## $ Above\_Median\_Numeric <dbl> 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0…

skim(ames4)

Data summary

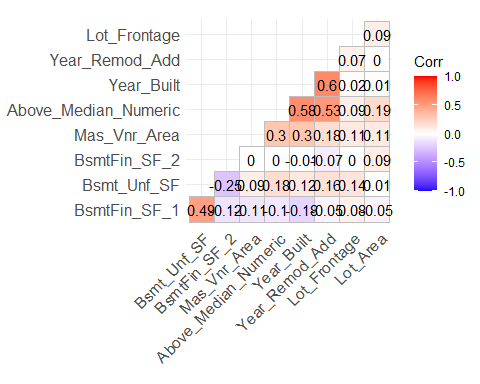
|  |  |
| --- | --- |
| Name | ames4 |
| Number of rows | 2053 |
| Number of columns | 35 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 35 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

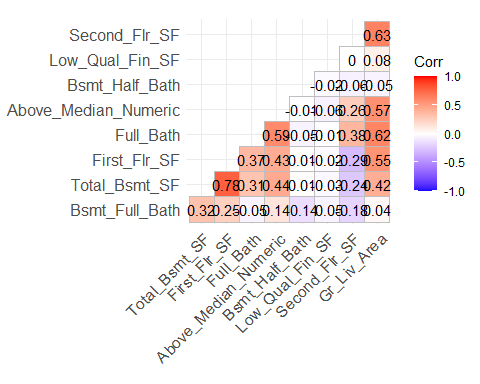
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Lot\_Frontage | 0 | 1 | 57.38 | 33.20 | 0.00 | 43.00 | 62.00 | 78.00 | 313.00 | ▇▇▁▁▁ |
| Lot\_Area | 0 | 1 | 10258.40 | 8427.38 | 1300.00 | 7500.00 | 9548.00 | 11600.00 | 215245.00 | ▇▁▁▁▁ |
| Year\_Built | 0 | 1 | 1970.64 | 30.40 | 1875.00 | 1953.00 | 1972.00 | 2000.00 | 2010.00 | ▁▂▃▆▇ |
| Year\_Remod\_Add | 0 | 1 | 1984.08 | 20.96 | 1950.00 | 1965.00 | 1993.00 | 2004.00 | 2010.00 | ▅▂▂▃▇ |
| Mas\_Vnr\_Area | 0 | 1 | 103.75 | 183.59 | 0.00 | 0.00 | 0.00 | 164.00 | 1600.00 | ▇▁▁▁▁ |
| BsmtFin\_SF\_1 | 0 | 1 | 4.21 | 2.24 | 1.00 | 3.00 | 3.00 | 7.00 | 7.00 | ▅▆▁▁▇ |
| BsmtFin\_SF\_2 | 0 | 1 | 52.57 | 175.99 | 0.00 | 0.00 | 0.00 | 0.00 | 1526.00 | ▇▁▁▁▁ |
| Bsmt\_Unf\_SF | 0 | 1 | 561.19 | 441.72 | 0.00 | 226.00 | 460.00 | 801.00 | 2336.00 | ▇▅▂▁▁ |
| Total\_Bsmt\_SF | 0 | 1 | 1054.57 | 435.33 | 0.00 | 793.00 | 988.00 | 1304.00 | 5095.00 | ▇▇▁▁▁ |
| First\_Flr\_SF | 0 | 1 | 1167.52 | 391.79 | 432.00 | 882.00 | 1088.00 | 1402.00 | 5095.00 | ▇▃▁▁▁ |
| Second\_Flr\_SF | 0 | 1 | 326.07 | 422.44 | 0.00 | 0.00 | 0.00 | 701.00 | 1862.00 | ▇▂▂▁▁ |
| Low\_Qual\_Fin\_SF | 0 | 1 | 4.97 | 49.09 | 0.00 | 0.00 | 0.00 | 0.00 | 1064.00 | ▇▁▁▁▁ |
| Gr\_Liv\_Area | 0 | 1 | 1498.56 | 487.84 | 480.00 | 1137.00 | 1447.00 | 1737.00 | 5095.00 | ▇▇▁▁▁ |
| Bsmt\_Full\_Bath | 0 | 1 | 0.43 | 0.53 | 0.00 | 0.00 | 0.00 | 1.00 | 3.00 | ▇▆▁▁▁ |
| Bsmt\_Half\_Bath | 0 | 1 | 0.06 | 0.24 | 0.00 | 0.00 | 0.00 | 0.00 | 2.00 | ▇▁▁▁▁ |
| Full\_Bath | 0 | 1 | 1.56 | 0.55 | 0.00 | 1.00 | 2.00 | 2.00 | 4.00 | ▁▇▇▁▁ |
| Half\_Bath | 0 | 1 | 0.38 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 2.00 | ▇▁▅▁▁ |
| Bedroom\_AbvGr | 0 | 1 | 2.86 | 0.82 | 0.00 | 2.00 | 3.00 | 3.00 | 6.00 | ▁▃▇▂▁ |
| Kitchen\_AbvGr | 0 | 1 | 1.05 | 0.22 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | ▇▁▁▁▁ |
| TotRms\_AbvGrd | 0 | 1 | 6.44 | 1.54 | 3.00 | 5.00 | 6.00 | 7.00 | 15.00 | ▅▇▃▁▁ |
| Fireplaces | 0 | 1 | 0.60 | 0.65 | 0.00 | 0.00 | 1.00 | 1.00 | 4.00 | ▇▇▁▁▁ |
| Garage\_Cars | 0 | 1 | 1.77 | 0.76 | 0.00 | 1.00 | 2.00 | 2.00 | 5.00 | ▅▇▂▁▁ |
| Garage\_Area | 0 | 1 | 471.96 | 213.43 | 0.00 | 320.00 | 478.00 | 576.00 | 1488.00 | ▃▇▂▁▁ |
| Wood\_Deck\_SF | 0 | 1 | 93.52 | 127.71 | 0.00 | 0.00 | 0.00 | 168.00 | 1424.00 | ▇▁▁▁▁ |
| Open\_Porch\_SF | 0 | 1 | 48.17 | 69.51 | 0.00 | 0.00 | 27.00 | 72.00 | 742.00 | ▇▁▁▁▁ |
| Enclosed\_Porch | 0 | 1 | 23.02 | 60.59 | 0.00 | 0.00 | 0.00 | 0.00 | 584.00 | ▇▁▁▁▁ |
| Three\_season\_porch | 0 | 1 | 2.80 | 25.65 | 0.00 | 0.00 | 0.00 | 0.00 | 407.00 | ▇▁▁▁▁ |
| Screen\_Porch | 0 | 1 | 16.68 | 57.94 | 0.00 | 0.00 | 0.00 | 0.00 | 576.00 | ▇▁▁▁▁ |
| Pool\_Area | 0 | 1 | 1.34 | 27.74 | 0.00 | 0.00 | 0.00 | 0.00 | 800.00 | ▇▁▁▁▁ |
| Misc\_Val | 0 | 1 | 60.12 | 662.76 | 0.00 | 0.00 | 0.00 | 0.00 | 17000.00 | ▇▁▁▁▁ |
| Mo\_Sold | 0 | 1 | 6.19 | 2.70 | 1.00 | 4.00 | 6.00 | 8.00 | 12.00 | ▅▆▇▃▃ |
| Year\_Sold | 0 | 1 | 2007.75 | 1.30 | 2006.00 | 2007.00 | 2008.00 | 2009.00 | 2010.00 | ▇▇▇▇▃ |
| Longitude | 0 | 1 | -93.64 | 0.03 | -93.69 | -93.66 | -93.64 | -93.62 | -93.58 | ▅▅▇▇▁ |
| Latitude | 0 | 1 | 42.03 | 0.02 | 41.99 | 42.02 | 42.03 | 42.05 | 42.06 | ▂▂▇▇▇ |
| Above\_Median\_Numeric | 0 | 1 | 0.51 | 0.50 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | ▇▁▁▁▇ |

## We look at correlation for continuous quantitative variables.

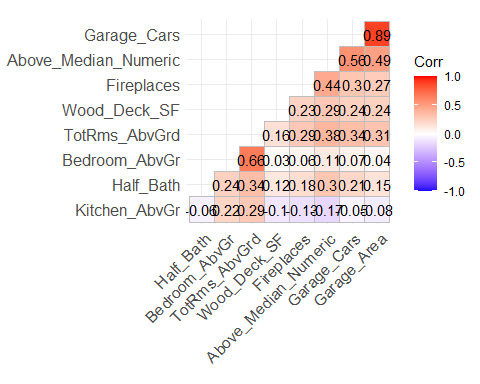
ames2 %>% select\_if(is.numeric) %>% dplyr::select("Lot\_Frontage","Lot\_Area", "Year\_Built","Year\_Remod\_Add","Mas\_Vnr\_Area","BsmtFin\_SF\_1","BsmtFin\_SF\_2","Bsmt\_Unf\_SF","Above\_Median\_Numeric") %>% cor() %>%  
ggcorrplot(hc.order = TRUE, type = "lower", lab = TRUE)



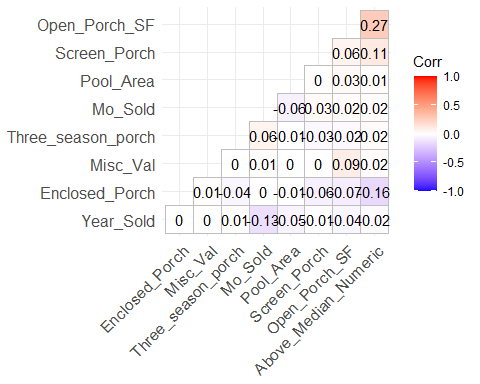
ames2 %>% select\_if(is.numeric) %>% dplyr::select("Total\_Bsmt\_SF","First\_Flr\_SF","Second\_Flr\_SF","Low\_Qual\_Fin\_SF","Gr\_Liv\_Area","Bsmt\_Full\_Bath","Bsmt\_Half\_Bath","Full\_Bath","Above\_Median\_Numeric") %>% cor() %>%  
ggcorrplot(hc.order = TRUE, type = "lower",lab = TRUE)



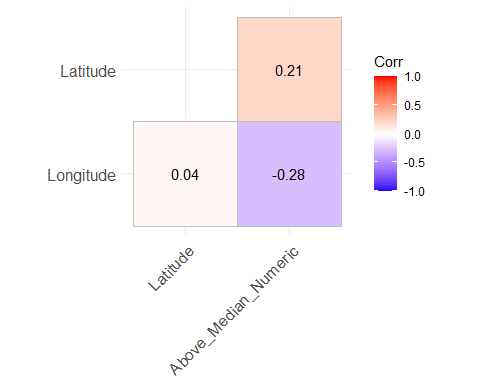
ames2 %>% select\_if(is.numeric) %>% dplyr::select("Half\_Bath","Bedroom\_AbvGr","Kitchen\_AbvGr","TotRms\_AbvGrd","Fireplaces","Garage\_Cars","Garage\_Area","Wood\_Deck\_SF","Above\_Median\_Numeric") %>% cor() %>%  
ggcorrplot(hc.order = TRUE, type = "lower",lab = TRUE)



ames2 %>% select\_if(is.numeric) %>% dplyr::select("Open\_Porch\_SF","Enclosed\_Porch","Three\_season\_porch","Screen\_Porch","Pool\_Area","Misc\_Val","Mo\_Sold","Year\_Sold","Above\_Median\_Numeric") %>% cor() %>%  
ggcorrplot(hc.order = TRUE, type = "lower",lab = TRUE)



ames2 %>% select\_if(is.numeric) %>% dplyr::select("Longitude","Latitude","Above\_Median\_Numeric") %>% cor() %>%  
ggcorrplot(hc.order = TRUE, type = "lower",lab = TRUE)



### Analyze nominal variables with closer look on data patterns

#colnames(ames3)  
glimpse((ames3))

## Rows: 2,053  
## Columns: 47  
## $ MS\_SubClass <fct> One\_Story\_1946\_and\_Newer\_All\_Styles, One\_Story\_1946\_and…  
## $ MS\_Zoning <fct> Residential\_Low\_Density, Residential\_High\_Density, Resi…  
## $ Street <fct> Pave, Pave, Pave, Pave, Pave, Pave, Pave, Pave, Pave, P…  
## $ Alley <fct> No\_Alley\_Access, No\_Alley\_Access, No\_Alley\_Access, No\_A…  
## $ Lot\_Shape <fct> Slightly\_Irregular, Regular, Slightly\_Irregular, Regula…  
## $ Land\_Contour <fct> Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, HLS, Lvl, Lvl, Lvl, Lvl, …  
## $ Utilities <fct> AllPub, AllPub, AllPub, AllPub, AllPub, AllPub, AllPub,…  
## $ Lot\_Config <fct> Corner, Inside, Corner, Corner, Inside, Inside, Inside,…  
## $ Land\_Slope <fct> Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, …  
## $ Neighborhood <fct> North\_Ames, North\_Ames, North\_Ames, North\_Ames, Gilbert…  
## $ Condition\_1 <fct> Norm, Feedr, Norm, Norm, Norm, Norm, Norm, Norm, Norm, …  
## $ Condition\_2 <fct> Norm, Norm, Norm, Norm, Norm, Norm, Norm, Norm, Norm, N…  
## $ Bldg\_Type <fct> OneFam, OneFam, OneFam, OneFam, OneFam, OneFam, TwnhsE,…  
## $ House\_Style <fct> One\_Story, One\_Story, One\_Story, One\_Story, Two\_Story, …  
## $ Overall\_Qual <fct> Above\_Average, Average, Above\_Average, Good, Average, A…  
## $ Overall\_Cond <fct> Average, Above\_Average, Above\_Average, Average, Average…  
## $ Roof\_Style <fct> Hip, Gable, Hip, Hip, Gable, Gable, Gable, Gable, Gable…  
## $ Roof\_Matl <fct> CompShg, CompShg, CompShg, CompShg, CompShg, CompShg, C…  
## $ Exterior\_1st <fct> BrkFace, VinylSd, Wd Sdng, BrkFace, VinylSd, VinylSd, H…  
## $ Exterior\_2nd <fct> Plywood, VinylSd, Wd Sdng, BrkFace, VinylSd, VinylSd, H…  
## $ Mas\_Vnr\_Type <fct> Stone, None, BrkFace, None, None, BrkFace, None, None, …  
## $ Exter\_Qual <fct> Typical, Typical, Typical, Good, Typical, Typical, Good…  
## $ Exter\_Cond <fct> Typical, Typical, Typical, Typical, Typical, Typical, T…  
## $ Foundation <fct> CBlock, CBlock, CBlock, CBlock, PConc, PConc, PConc, PC…  
## $ Bsmt\_Qual <fct> Typical, Typical, Typical, Typical, Good, Typical, Good…  
## $ Bsmt\_Cond <fct> Good, Typical, Typical, Typical, Typical, Typical, Typi…  
## $ Bsmt\_Exposure <fct> Gd, No, No, No, No, No, No, No, No, Gd, Av, Gd, Av, Av,…  
## $ BsmtFin\_Type\_1 <fct> BLQ, Rec, ALQ, ALQ, GLQ, GLQ, ALQ, GLQ, ALQ, GLQ, GLQ, …  
## $ BsmtFin\_Type\_2 <fct> Unf, LwQ, Unf, Unf, Unf, Unf, Unf, Unf, Unf, Unf, BLQ, …  
## $ Heating <fct> GasA, GasA, GasA, GasA, GasA, GasA, GasA, GasA, GasA, G…  
## $ Heating\_QC <fct> Fair, Typical, Typical, Excellent, Good, Excellent, Exc…  
## $ Central\_Air <fct> Y, 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,…  
## $ Kitchen\_Qual <fct> Typical, Typical, Good, Excellent, Typical, Good, Good,…  
## $ Functional <fct> Typ, Typ, Typ, Typ, Typ, Typ, Typ, Typ, Typ, Typ, Typ, …  
## $ Fireplace\_Qu <fct> Good, No\_Fireplace, No\_Fireplace, Typical, Typical, Goo…  
## $ Garage\_Type <fct> Attchd, Attchd, Attchd, Attchd, Attchd, Attchd, Attchd,…  
## $ Garage\_Finish <fct> Fin, Unf, Unf, Fin, Fin, Fin, RFn, RFn, Fin, Unf, RFn, …  
## $ Garage\_Qual <fct> Typical, Typical, Typical, Typical, Typical, Typical, T…  
## $ Garage\_Cond <fct> Typical, Typical, Typical, Typical, Typical, Typical, T…  
## $ Paved\_Drive <fct> Partial\_Pavement, Paved, Paved, Paved, Paved, Paved, Pa…  
## $ Pool\_QC <fct> No\_Pool, No\_Pool, No\_Pool, No\_Pool, No\_Pool, No\_Pool, N…  
## $ Fence <fct> No\_Fence, Minimum\_Privacy, No\_Fence, No\_Fence, Minimum\_…  
## $ Misc\_Feature <fct> None, None, Gar2, None, None, None, None, None, Shed, N…  
## $ Sale\_Type <fct> WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, New…  
## $ Sale\_Condition <fct> Normal, Normal, Normal, Normal, Normal, Normal, Normal,…  
## $ Above\_Median <fct> Yes, No, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Y…

#labels(ames3)  
skim(ames3)

Data summary

|  |  |
| --- | --- |
| Name | ames3 |
| Number of rows | 2053 |
| Number of columns | 47 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 47 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| MS\_SubClass | 0 | 1 | FALSE | 16 | One: 772, Two: 383, One: 204, One: 129 |
| MS\_Zoning | 0 | 1 | FALSE | 7 | Res: 1600, Res: 326, Flo: 87, Res: 20 |
| Street | 0 | 1 | FALSE | 2 | Pav: 2046, Grv: 7 |
| Alley | 0 | 1 | FALSE | 3 | No\_: 1914, Gra: 94, Pav: 45 |
| Lot\_Shape | 0 | 1 | FALSE | 4 | Reg: 1275, Sli: 714, Mod: 53, Irr: 11 |
| Land\_Contour | 0 | 1 | FALSE | 4 | Lvl: 1833, HLS: 94, Bnk: 81, Low: 45 |
| Utilities | 0 | 1 | FALSE | 2 | All: 2052, NoS: 1 |
| Lot\_Config | 0 | 1 | FALSE | 5 | Ins: 1495, Cor: 359, Cul: 135, FR2: 56 |
| Land\_Slope | 0 | 1 | FALSE | 3 | Gtl: 1951, Mod: 89, Sev: 13 |
| Neighborhood | 0 | 1 | FALSE | 28 | Nor: 327, Col: 183, Old: 181, Edw: 129 |
| Condition\_1 | 0 | 1 | FALSE | 9 | Nor: 1771, Fee: 113, Art: 67, RRA: 35 |
| Condition\_2 | 0 | 1 | FALSE | 8 | Nor: 2027, Fee: 12, Pos: 4, Art: 4 |
| Bldg\_Type | 0 | 1 | FALSE | 5 | One: 1706, Twn: 157, Dup: 76, Twn: 67 |
| House\_Style | 0 | 1 | FALSE | 8 | One: 1052, Two: 590, One: 225, SLv: 90 |
| Overall\_Qual | 0 | 1 | FALSE | 10 | Ave: 587, Abo: 518, Goo: 411, Ver: 237 |
| Overall\_Cond | 0 | 1 | FALSE | 9 | Ave: 1143, Abo: 376, Goo: 286, Ver: 98 |
| Roof\_Style | 0 | 1 | FALSE | 6 | Gab: 1607, Hip: 404, Gam: 14, Fla: 14 |
| Roof\_Matl | 0 | 1 | FALSE | 6 | Com: 2023, Tar: 17, WdS: 8, WdS: 3 |
| Exterior\_1st | 0 | 1 | FALSE | 16 | Vin: 705, Met: 319, Wd : 313, HdB: 303 |
| Exterior\_2nd | 0 | 1 | FALSE | 17 | Vin: 699, Met: 317, Wd : 302, HdB: 277 |
| Mas\_Vnr\_Type | 0 | 1 | FALSE | 5 | Non: 1231, Brk: 638, Sto: 166, Brk: 17 |
| Exter\_Qual | 0 | 1 | FALSE | 4 | Typ: 1272, Goo: 682, Exc: 78, Fai: 21 |
| Exter\_Cond | 0 | 1 | FALSE | 5 | Typ: 1787, Goo: 213, Fai: 43, Exc: 9 |
| Foundation | 0 | 1 | FALSE | 6 | PCo: 911, CBl: 880, Brk: 216, Sla: 36 |
| Bsmt\_Qual | 0 | 1 | FALSE | 6 | Typ: 911, Goo: 849, Exc: 178, No\_: 57 |
| Bsmt\_Cond | 0 | 1 | FALSE | 6 | Typ: 1833, Goo: 80, Fai: 76, No\_: 57 |
| Bsmt\_Exposure | 0 | 1 | FALSE | 5 | No: 1331, Av: 284, Gd: 199, Mn: 179 |
| BsmtFin\_Type\_1 | 0 | 1 | FALSE | 7 | Unf: 602, GLQ: 578, ALQ: 298, Rec: 216 |
| BsmtFin\_Type\_2 | 0 | 1 | FALSE | 7 | Unf: 1740, Rec: 79, LwQ: 64, No\_: 58 |
| Heating | 0 | 1 | FALSE | 6 | Gas: 2019, Gas: 21, Gra: 6, Wal: 5 |
| Heating\_QC | 0 | 1 | FALSE | 5 | Exc: 1040, Typ: 618, Goo: 333, Fai: 61 |
| Central\_Air | 0 | 1 | FALSE | 2 | Y: 1916, N: 137 |
| Electrical | 0 | 1 | FALSE | 5 | SBr: 1887, Fus: 126, Fus: 33, Fus: 6 |
| Kitchen\_Qual | 0 | 1 | FALSE | 5 | Typ: 1070, Goo: 790, Exc: 142, Fai: 50 |
| Functional | 0 | 1 | FALSE | 8 | Typ: 1896, Min: 54, Min: 51, Mod: 27 |
| Fireplace\_Qu | 0 | 1 | FALSE | 6 | No\_: 993, Goo: 538, Typ: 409, Fai: 56 |
| Garage\_Type | 0 | 1 | FALSE | 7 | Att: 1204, Det: 549, Bui: 127, No\_: 108 |
| Garage\_Finish | 0 | 1 | FALSE | 4 | Unf: 872, RFn: 563, Fin: 509, No\_: 109 |
| Garage\_Qual | 0 | 1 | FALSE | 6 | Typ: 1839, No\_: 109, Fai: 85, Goo: 16 |
| Garage\_Cond | 0 | 1 | FALSE | 6 | Typ: 1872, No\_: 109, Fai: 53, Goo: 10 |
| Paved\_Drive | 0 | 1 | FALSE | 3 | Pav: 1848, Dir: 163, Par: 42 |
| Pool\_QC | 0 | 1 | FALSE | 5 | No\_: 2047, Exc: 2, Typ: 2, Fai: 1 |
| Fence | 0 | 1 | FALSE | 5 | No\_: 1661, Min: 225, Goo: 81, Goo: 77 |
| Misc\_Feature | 0 | 1 | FALSE | 5 | Non: 1978, She: 66, Gar: 5, Oth: 3 |
| Sale\_Type | 0 | 1 | FALSE | 10 | WD: 1789, New: 163, COD: 54, Con: 16 |
| Sale\_Condition | 0 | 1 | FALSE | 6 | Nor: 1712, Par: 169, Abn: 121, Fam: 30 |
| Above\_Median | 0 | 1 | FALSE | 2 | Yes: 1043, No: 1010 |

## Identifying variables for final dataset

homes\_data=ames2 %>% dplyr::select("Above\_Median", "Lot\_Config","Overall\_Qual","Neighborhood","Bldg\_Type", "House\_Style","Exterior\_2nd","Bsmt\_Exposure","BsmtFin\_Type\_1", "Heating\_QC","Garage\_Finish","Garage\_Qual", "Year\_Built", "Year\_Remod\_Add","Gr\_Liv\_Area" ,"Fireplaces","Garage\_Cars","Garage\_Area")  
skim(homes\_data)

Data summary

|  |  |
| --- | --- |
| Name | homes\_data |
| Number of rows | 2053 |
| Number of columns | 18 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 12 |
| numeric | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| Above\_Median | 0 | 1 | FALSE | 2 | Yes: 1043, No: 1010 |
| Lot\_Config | 0 | 1 | FALSE | 5 | Ins: 1495, Cor: 359, Cul: 135, FR2: 56 |
| Overall\_Qual | 0 | 1 | FALSE | 10 | Ave: 587, Abo: 518, Goo: 411, Ver: 237 |
| Neighborhood | 0 | 1 | FALSE | 28 | Nor: 327, Col: 183, Old: 181, Edw: 129 |
| Bldg\_Type | 0 | 1 | FALSE | 5 | One: 1706, Twn: 157, Dup: 76, Twn: 67 |
| House\_Style | 0 | 1 | FALSE | 8 | One: 1052, Two: 590, One: 225, SLv: 90 |
| Exterior\_2nd | 0 | 1 | FALSE | 17 | Vin: 699, Met: 317, Wd : 302, HdB: 277 |
| Bsmt\_Exposure | 0 | 1 | FALSE | 5 | No: 1331, Av: 284, Gd: 199, Mn: 179 |
| BsmtFin\_Type\_1 | 0 | 1 | FALSE | 7 | Unf: 602, GLQ: 578, ALQ: 298, Rec: 216 |
| Heating\_QC | 0 | 1 | FALSE | 5 | Exc: 1040, Typ: 618, Goo: 333, Fai: 61 |
| Garage\_Finish | 0 | 1 | FALSE | 4 | Unf: 872, RFn: 563, Fin: 509, No\_: 109 |
| Garage\_Qual | 0 | 1 | FALSE | 6 | Typ: 1839, No\_: 109, Fai: 85, Goo: 16 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year\_Built | 0 | 1 | 1970.64 | 30.40 | 1875 | 1953 | 1972 | 2000 | 2010 | ▁▂▃▆▇ |
| Year\_Remod\_Add | 0 | 1 | 1984.08 | 20.96 | 1950 | 1965 | 1993 | 2004 | 2010 | ▅▂▂▃▇ |
| Gr\_Liv\_Area | 0 | 1 | 1498.56 | 487.84 | 480 | 1137 | 1447 | 1737 | 5095 | ▇▇▁▁▁ |
| Fireplaces | 0 | 1 | 0.60 | 0.65 | 0 | 0 | 1 | 1 | 4 | ▇▇▁▁▁ |
| Garage\_Cars | 0 | 1 | 1.77 | 0.76 | 0 | 1 | 2 | 2 | 5 | ▅▇▂▁▁ |
| Garage\_Area | 0 | 1 | 471.96 | 213.43 | 0 | 320 | 478 | 576 | 1488 | ▃▇▂▁▁ |

glimpse(homes\_data)

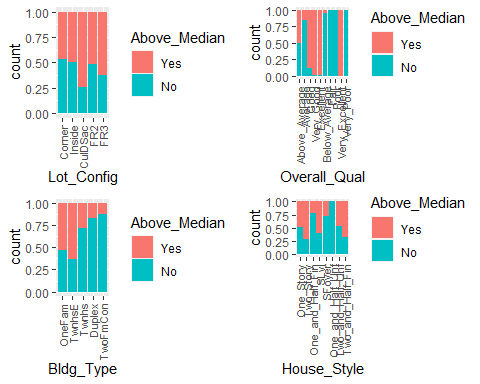
## Rows: 2,053  
## Columns: 18  
## $ Above\_Median <fct> Yes, No, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Y…  
## $ Lot\_Config <fct> Corner, Inside, Corner, Corner, Inside, Inside, Inside,…  
## $ Overall\_Qual <fct> Above\_Average, Average, Above\_Average, Good, Average, A…  
## $ Neighborhood <fct> North\_Ames, North\_Ames, North\_Ames, North\_Ames, Gilbert…  
## $ Bldg\_Type <fct> OneFam, OneFam, OneFam, OneFam, OneFam, OneFam, TwnhsE,…  
## $ House\_Style <fct> One\_Story, One\_Story, One\_Story, One\_Story, Two\_Story, …  
## $ Exterior\_2nd <fct> Plywood, VinylSd, Wd Sdng, BrkFace, VinylSd, VinylSd, H…  
## $ Bsmt\_Exposure <fct> Gd, No, No, No, No, No, No, No, No, Gd, Av, Gd, Av, Av,…  
## $ BsmtFin\_Type\_1 <fct> BLQ, Rec, ALQ, ALQ, GLQ, GLQ, ALQ, GLQ, ALQ, GLQ, GLQ, …  
## $ Heating\_QC <fct> Fair, Typical, Typical, Excellent, Good, Excellent, Exc…  
## $ Garage\_Finish <fct> Fin, Unf, Unf, Fin, Fin, Fin, RFn, RFn, Fin, Unf, RFn, …  
## $ Garage\_Qual <fct> Typical, Typical, Typical, Typical, Typical, Typical, T…  
## $ Year\_Built <dbl> 1960, 1961, 1958, 1968, 1997, 1998, 1992, 1995, 1992, 1…  
## $ Year\_Remod\_Add <dbl> 1960, 1961, 1958, 1968, 1998, 1998, 1992, 1996, 2007, 1…  
## $ Gr\_Liv\_Area <dbl> 1656, 896, 1329, 2110, 1629, 1604, 1280, 1616, 1187, 13…  
## $ Fireplaces <dbl> 2, 0, 0, 2, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1…  
## $ Garage\_Cars <dbl> 2, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 3, 2, 2, 2, 2, 1…  
## $ Garage\_Area <dbl> 528, 730, 312, 522, 482, 470, 506, 608, 420, 506, 528, …

str(homes\_data)

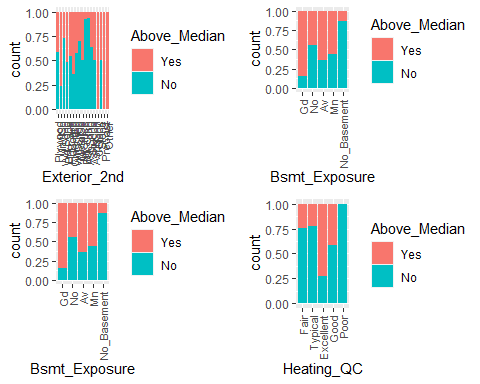
## tibble [2,053 × 18] (S3: tbl\_df/tbl/data.frame)  
## $ Above\_Median : Factor w/ 2 levels "Yes","No": 1 2 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Config : Factor w/ 5 levels "Corner","Inside",..: 1 2 1 1 2 2 2 2 2 2 ...  
## $ Overall\_Qual : Factor w/ 10 levels "Above\_Average",..: 1 2 1 3 2 1 4 4 1 3 ...  
## $ Neighborhood : Factor w/ 28 levels "North\_Ames","Gilbert",..: 1 1 1 1 2 2 3 3 2 2 ...  
## $ Bldg\_Type : Factor w/ 5 levels "OneFam","TwnhsE",..: 1 1 1 1 1 1 2 2 1 1 ...  
## $ House\_Style : Factor w/ 8 levels "One\_Story","Two\_Story",..: 1 1 1 1 2 2 1 1 1 1 ...  
## $ Exterior\_2nd : Factor w/ 17 levels "Plywood","VinylSd",..: 1 2 3 4 2 2 5 6 5 5 ...  
## $ Bsmt\_Exposure : Factor w/ 5 levels "Gd","No","Av",..: 1 2 2 2 2 2 2 2 2 1 ...  
## $ BsmtFin\_Type\_1: Factor w/ 7 levels "BLQ","Rec","ALQ",..: 1 2 3 3 4 4 3 4 3 4 ...  
## $ Heating\_QC : Factor w/ 5 levels "Fair","Typical",..: 1 2 2 3 4 3 3 3 3 4 ...  
## $ Garage\_Finish : Factor w/ 4 levels "Fin","Unf","RFn",..: 1 2 2 1 1 1 3 3 1 2 ...  
## $ Garage\_Qual : Factor w/ 6 levels "Typical","No\_Garage",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Year\_Built : num [1:2053] 1960 1961 1958 1968 1997 ...  
## $ Year\_Remod\_Add: num [1:2053] 1960 1961 1958 1968 1998 ...  
## $ Gr\_Liv\_Area : num [1:2053] 1656 896 1329 2110 1629 ...  
## $ Fireplaces : num [1:2053] 2 0 0 2 1 1 0 1 0 1 ...  
## $ Garage\_Cars : num [1:2053] 2 1 1 2 2 2 2 2 2 2 ...  
## $ Garage\_Area : num [1:2053] 528 730 312 522 482 470 506 608 420 506 ...

## Examin Corelation of nominal variables using focused plots since we had identified them during correlation plots in begining

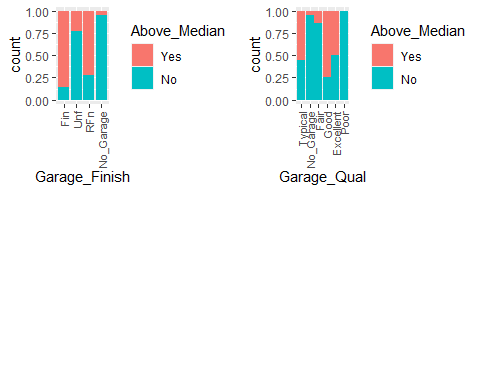
p1 = ggplot(homes\_data, aes(x = Lot\_Config, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p2 = ggplot(homes\_data, aes(x = Overall\_Qual, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p3 = ggplot(homes\_data, aes(x = Bldg\_Type, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p4 = ggplot(homes\_data, aes(x = House\_Style, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p5 = ggplot(homes\_data, aes(x = Exterior\_2nd, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p6 = ggplot(homes\_data, aes(x = Bsmt\_Exposure, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p7 = ggplot(homes\_data, aes(x = BsmtFin\_Type\_1, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p8 = ggplot(homes\_data, aes(x = Heating\_QC, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p9 = ggplot(homes\_data, aes(x = Garage\_Finish, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
p10 = ggplot(homes\_data, aes(x = Garage\_Qual, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 90, size = 8, vjust = 0.5, hjust=1))  
  
grid.arrange(p1,p2,p3,p4, ncol = 2, heights=c(4,4))



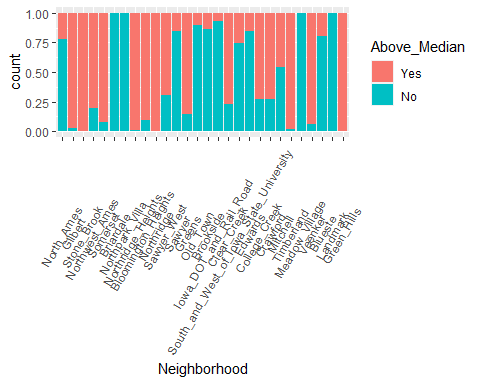
grid.arrange(p5,p6,p6,p8, ncol = 2, heights=c(4,4))



grid.arrange(p9,p10, ncol = 2, heights=c(4,4))

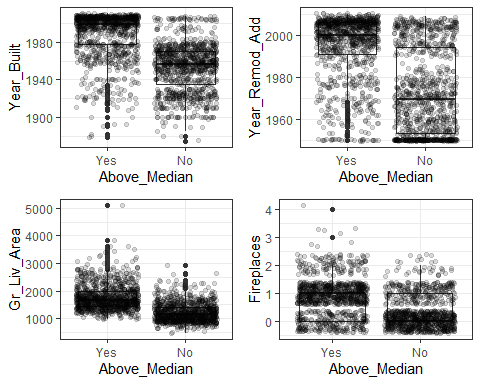


ggplot(homes\_data, aes(x = Neighborhood, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x = element\_text(angle = 60, size = 9, vjust = 0.6, hjust=0.6))



### Examing datapoint of Continuous variables in relation with response variable : Year\_Built, Year\_Remod\_Add, Mas\_Vnr\_Area, First\_Flr\_SF, Fireplaces, Garage\_Cars, Garage\_Area

p11 = ggplot(homes\_data,aes(x=Above\_Median,y=Year\_Built)) + geom\_boxplot() + geom\_jitter(alpha = 0.15) + theme\_bw()   
p12 = ggplot(homes\_data,aes(x=Above\_Median,y=Year\_Remod\_Add)) + geom\_boxplot() + geom\_jitter(alpha = 0.15) + theme\_bw()   
p13 = ggplot(homes\_data,aes(x=Above\_Median,y=Gr\_Liv\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.15) + theme\_bw()  
p14 = ggplot(homes\_data,aes(x=Above\_Median,y=Fireplaces)) + geom\_boxplot() + geom\_jitter(alpha = 0.15) + theme\_bw()   
p15 = ggplot(homes\_data,aes(x=Above\_Median,y=Garage\_Cars)) + geom\_boxplot() + geom\_jitter(alpha = 0.15) + theme\_bw()  
p16 = ggplot(homes\_data,aes(x=Above\_Median,y=Garage\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.15) + theme\_bw()   
  
grid.arrange(p11,p12,p13,p14, ncol = 2)



grid.arrange(p15, p16, ncol = 2)

