ITERA

Modul 1 Praktikum
Statistika Sains Data

Eksplorasi Data

Program Studi Sains Data
Fakultas Sains
Institut Teknologi Sumatera

2024

A. Tujuan Praktikum

1. Mahasiswa mampu mengoperasikan dan melakukan eksplorasi data menggunakan software RStudio.
2. Mahasiswa mampu menganalisis dan menginterpetasi hasil dari eksplorasi data yang telah diolah dengan RStudio

B. Teori Dasar

I. Eksplorasi Data

Data yang akan dianalisis harus sudah siap terlebih dahulu dalam format yang sesuai dengan paket komputer yang akan dipakai. Sebelum memilih metode yang akan dipergunakan untuk menguji data, terlebih dahulu perlu dilakukan eksplorasi data untuk mengetahui informasi penting dan sifat-sifat khas dari data (banyaknya dan jenis peubah, sebaran data dan sebagainya) sehingga prasarat penggunaan metode statistika terpenuhi. Selain itu data juga harus disajikan dalam bentuk yang mudah untuk dilihat karakteristik umumnya (misalnya dalam bentuk tabel atau grafik). Kegiatan ini menjadi bagian dari Statistika Deskriptif yang biasa dilakukan sebelum melakukan uji data lebih jauh.

II. Data Kuantitatif dan Data Kualitatif

- **Pendahuluan**
  - Di R data umumnya disimpan dalam bentuk vektor atau data frame.
  - Data kualitatif, di dalam statistika dikenal sebagai data kategorikal.
  - Data kualitatif dapat disimpan dalam bentuk Factors.
  - Data kuantitatif, di dalam statistika dikenal sebagai data kontinu atau data numerik.
  - Data kuantitatif dapat disimpan dalam bentuk Numerik.
- **Data kualitatif**

  Data kualitatif merupakan data non-statistik yang umumnya bersifat tidak terstruktur atau semi-terstruktur.
  - Data kualitatif tidak melulu berasal dari pengukuran.
  - Data kualitatif dikategorikan berdasarkan sifat - sifat, atribut, label, dll.
  - Data ini digunakan untuk interpretasi dan pembuatan hipotesis.
  - Data ini tidak dapat dikumpulkan dan dianalisa menggunakan metode - metode konvensional.
  - Contoh - contoh data kualitatif: Jenis kelamin, Ukuran sepatu, Rating.

Contoh data kualitatif dalam pemrograman R

```
ukuranBaju <- c('S', 'M', 'L', 'XL',
                'XXL', 'M', 'L', 'XL',
                'XXL', 'S', 'M')
ukuranBaju
```

1. 'S'
2. 'M'
3. 'L'
4. 'XL'
5. 'XXL'
6. 'M'
7. 'L'
8. 'XL'

9. 'XXL'
10. 'S'
11. 'M'

```
ukuran_baju <- factor(ukuranBaju) # dijadikan dalam bentuk Factor
ukuran_baju
```

1. S
2. M
3. L
4. XL
5. XXL
6. M
7. L
8. XL
9. XXL
10. S
11. M

▶**Levels**:

```
str(ukuran_baju)
```

```
 Factor w/ 5 levels "L","M","S","XL",..: 3 2 1 4 5 2 1 4 5
3 ...
```

```
summary(ukuran_baju)
```

```
L
        2
M
        3
S
        2
XL
        2
XXL
        2
```

```
levels(ukuranBaju)
```

```
NULL
```

```
levels(ukuran_baju)
```

1. 'L'
2. 'M'
3. 'S'
4. 'XL'
5. 'XXL'

- **Visualisasi data kualitatif**

  Gunakan: Diagram batang dan Diagram lingkaran.

```
ukuran_baju
```

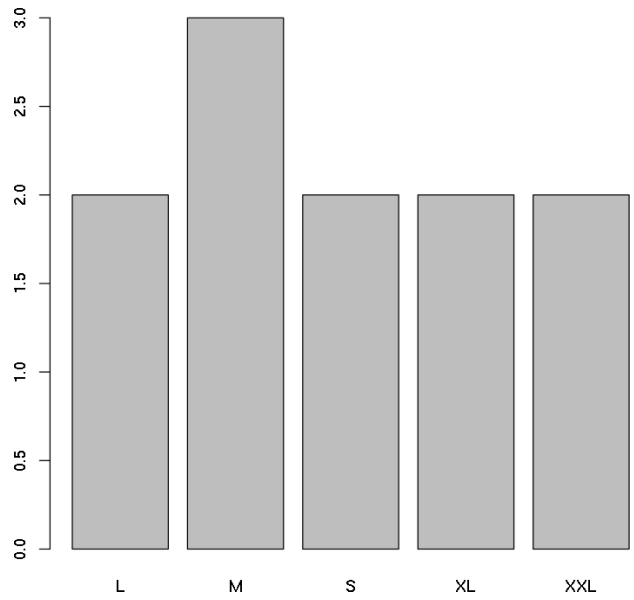
1. S
2. M
3. L
4. XL
5. XXL
6. M
7. L
8. XL
9. XXL
10. S
11. M

▶ **Levels**:

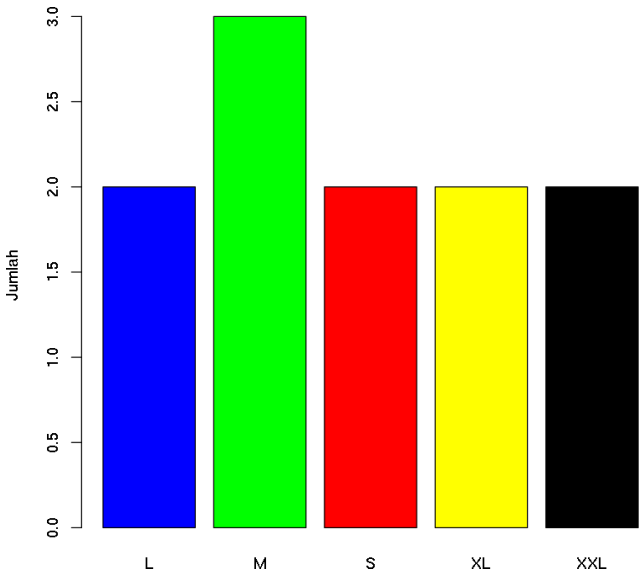
```
tabelUkuranBaju <- table(ukuran_baju)
tabelUkuranBaju
```

```
ukuran_baju
 L  M  S XL XXL 
 2  3  2  2   2 
```

```
barplot(tabelUkuranBaju)
```




```
# kostumisasi diagram batang
barplot(tabelUkuranBaju,
        col = c('blue', 'green', 'red', 'yellow','black'),
        ylab = 'Jumlah')
```

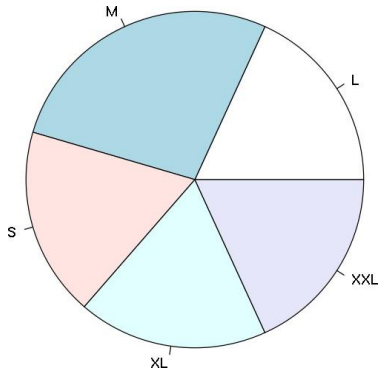



```
jmlh_ukuranM = sum(ukuran_baju == 'M')
jmlh_ukuranM
```

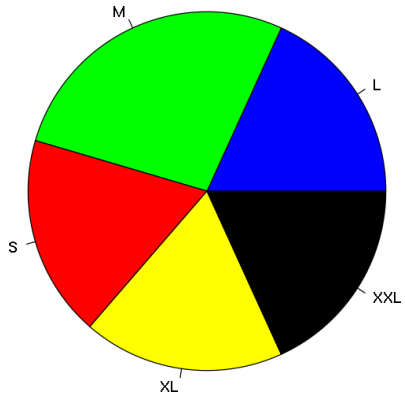
1. FALSE
2. TRUE
3. FALSE
4. FALSE
5. FALSE
6. TRUE

7. FALSE
8. FALSE
9. FALSE
10. FALSE
11. TRUE

```
# Penggunaan diagram lingkaran
pie(tabelUkuranBaju)
```




```
# kostumisasi diagram lingkaran pie(tabelUkuranBaju, col =
c('blue', 'green', 'red',
'yellow','black'))
```




```
kategori_usia <- factor(c(2,4,3,3,2,1,1,1,2,1,1,4,3,4,2,2,2,1,4,4,4,4,3,2,2
,1))
```

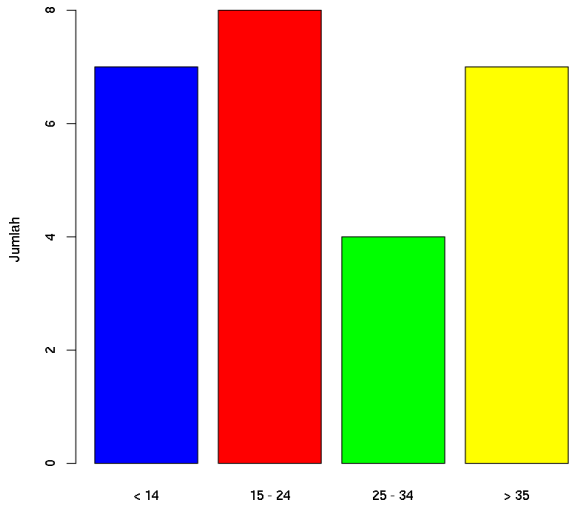
```
levels(kategori_usia)
```

1. '1'
2. '2'
3. '3'
4. '4'

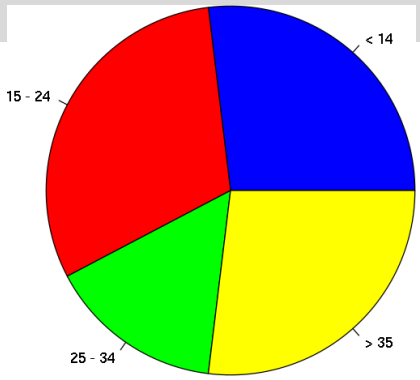
```
# mengubah level kategori usia
levels(kategori_usia) <- c('< 14', '15 - 24', '25 - 34',
'> 35')
```

```
kategori_usia
< 14 15 - 24 25 - 34 > 35
   7     8     4     7
```

```
barplot(tabelKategoriUsia,
        col=c('blue', 'red', 'green', 'yellow'),ylab = 'Jumlah')
```




```
pie(tabelKategoriUsia, col=c('blue', 'red', 'green',
'yellow'))
```



- **Data kuantitatif**
  - Data kuantitatif dapat dihitung, diukur, dan diekspresikan secara numerik.
  - Data kualitatif sendiri bersifat deskriptif dan konseptual.
  - Data kuantitatif bersifat terstruktur.
  - Contoh: Pengukuran temperatur udara, Harga saham, dll.

Contoh data kuantitatif dalam pemrograman R

```
# panjang lagu (dalam menit)
lagu <- c(5.3,3.6,5.5,4.7,6.7,4.3,6.2,4.3,4.9,5.1,5.8,4.4)
lagu
```

1. 5.3
2. 3.6
3. 5.5
4. 4.7
5. 6.7
6. 4.3
7. 6.2
8. 4.3
9. 4.9
10. 5.1
11. 5.8

- **Visualisasi data kuantitatif**
  - Histogram
  - *Boxplot*
  - *strip-chart :* alternatif *boxplot* ketika ukuran sampel kecil.

Histogram

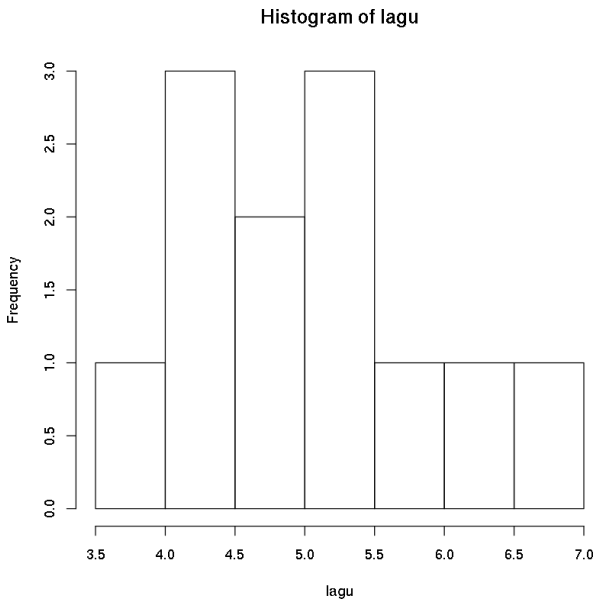
```
length(lagu) # jumlah elemen di dalam vektor lagu
```

12

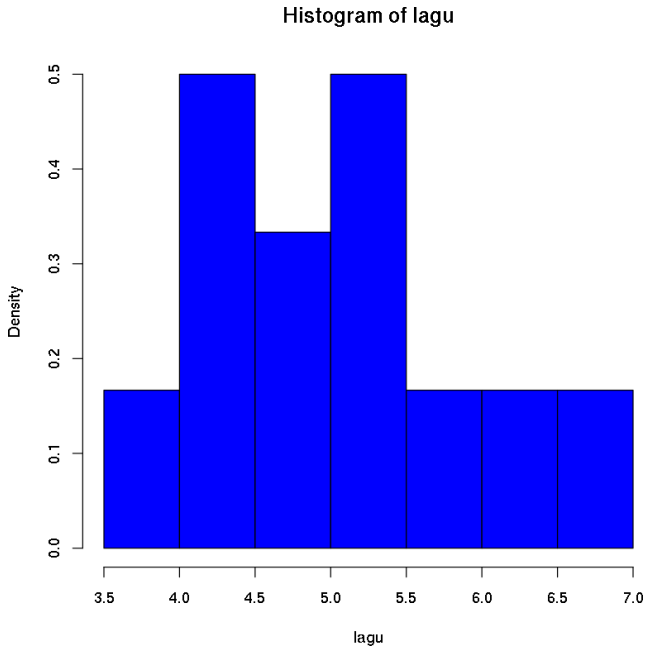
```
summary(lagu)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  3.600   4.375   5.000   5.067   5.575   6.700 
```

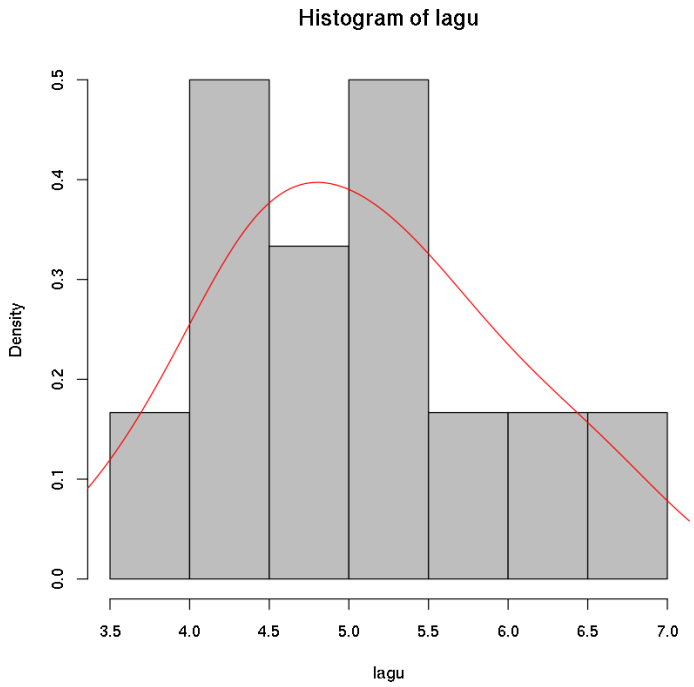
```
hist(lagu)
```




```
hist(lagu, col='blue', prob=T) # pdf: probability density function
```

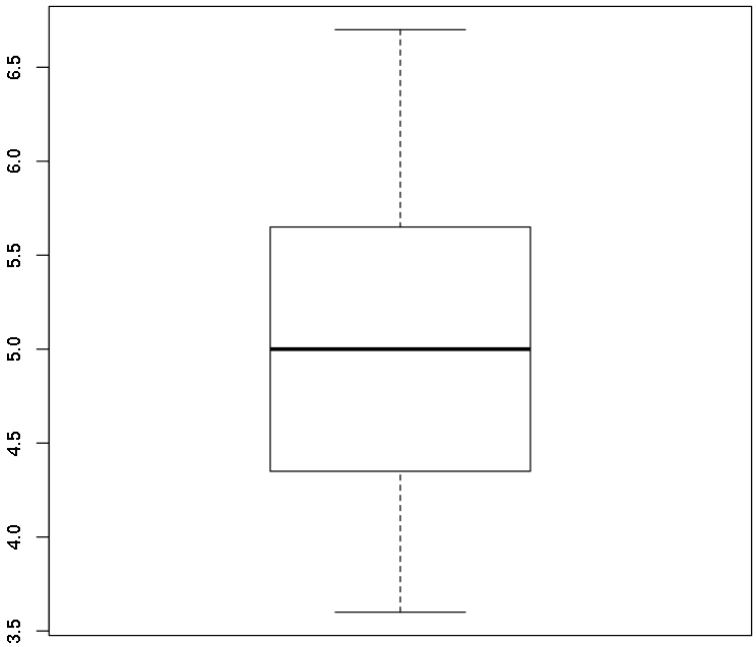



```
hist(lagu, col='grey', prob=T)
lines(density(lagu),  col='red')
```



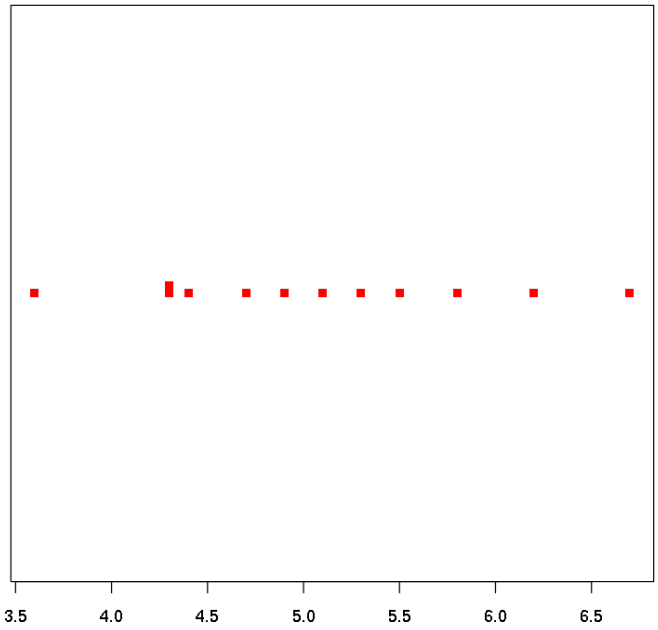
boxplot

```
boxplot(lagu)
```



strip-chart

```
stripchart(lagu,col='red',  pch=15,  method='stack')
```



C. Latihan Praktikum

Dalam hal ini, akan dilakukan praktikum berupa Visualisasi data dengan Eksplorasi data analisis.

- **Tentang Data**

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.

Deskripsi Setiap Kolom dapat dapat diakses pada link berikut ini

Deskripsi Data

Data ini bisa diperoleh di link berikut ini

Download Data

Package

Silahkan install jika belum ada

```
install.packages("tidyverse")
install.packages("DataExplorer")
install.packages("skimr")
```

Memanggil Package

```
library(tidyverse)
library(DataExplorer)
library(skimr)
```

Import Data

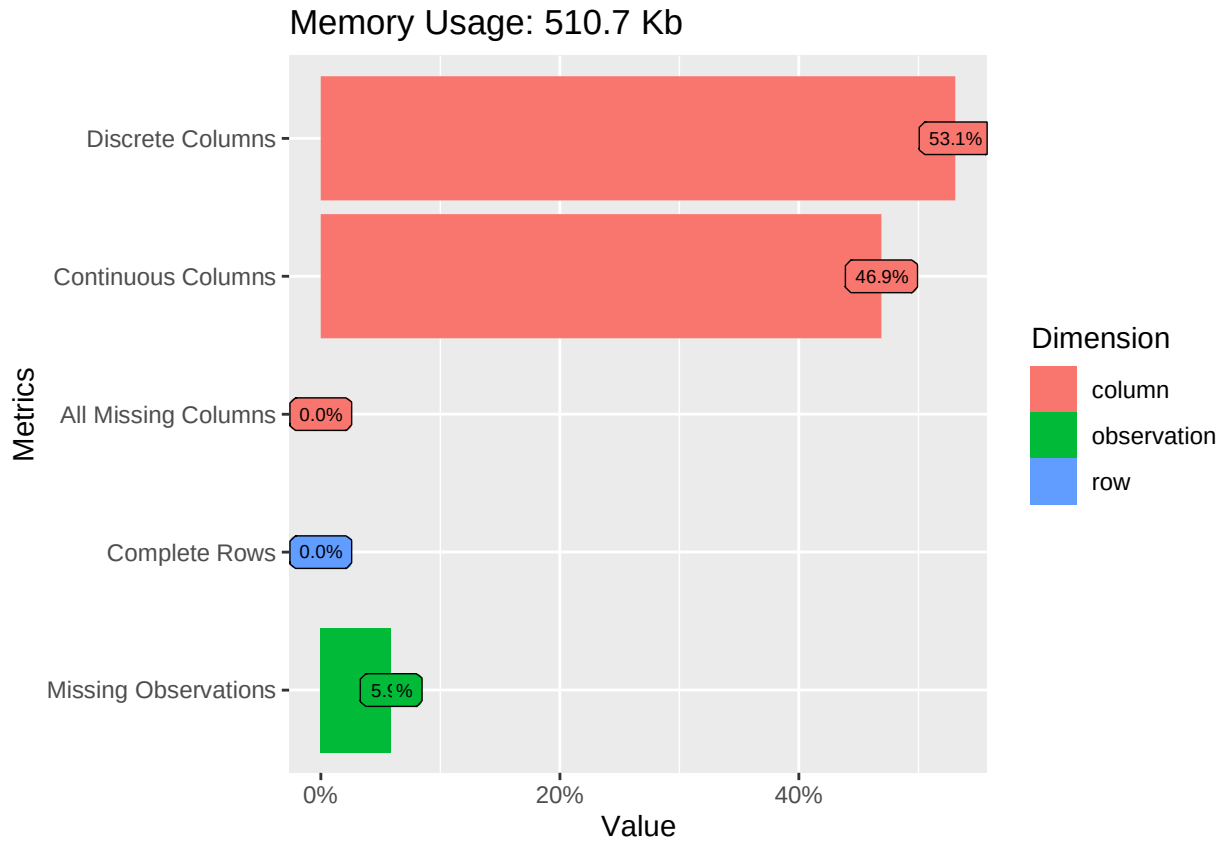
```
data_house <- read.csv("house_price1.csv",stringsAsFactors = TRUE)
glimpse(data_house)
```

```
## Rows: 1,460
## Columns: 81
## $ Id            <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1~
## $ MSSubClass    <int> 60, 20, 60, 70, 60, 50, 20, 60, 50, 190, 20, 60, 20, 20,~
## $ MSZoning      <fct> RL, RL, RL, RL, RL, RL, RL, RL, RM, RL, RL, RL, RL, RL, ~
## $ LotFrontage   <int> 65, 80, 68, 60, 84, 85, 75, NA, 51, 50, 70, 85, NA, 91, ~
## $ LotArea       <int> 8450, 9600, 11250, 9550, 14260, 14115, 10084, 10382, 612~
## $ Street        <fct> Pave, Pave, Pave, Pave, Pave, Pave, Pave, Pave, Pave, Pa~
## $ Alley         <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ LotShape      <fct> Reg, Reg, IR1, IR1, IR1, IR1, Reg, IR1, Reg, Reg, Reg, I~
## $ LandContour   <fct> Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, Lvl, L~
## $ Utilities     <fct> AllPub, AllPub, AllPub, AllPub, AllPub, AllPub, AllPub, ~
## $ LotConfig     <fct> Inside, FR2, Inside, Corner, FR2, Inside, Inside, Corner~
## $ LandSlope     <fct> Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, Gtl, G~
## $ Neighborhood  <fct> CollgCr, Veenker, CollgCr, Crawfor, NoRidge, Mitchel, So~
## $ Condition1    <fct> Norm, Feedr, Norm, Norm, Norm, Norm, Norm, PosN, Artery, ~
## $ Condition2    <fct> Norm, Norm, Norm, Norm, Norm, Norm, Norm, Norm, Norm, Ar~
## $ BldgType      <fct> 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 1Fam, 2f~
## $ HouseStyle    <fct> 2Story, 1Story, 2Story, 2Story, 2Story, 1.5Fin, 1Story, ~
## $ OverallQual   <int> 7, 6, 7, 7, 8, 5, 8, 7, 7, 5, 5, 9, 5, 7, 6, 7, 6, 4, 5, ~
## $ OverallCond   <int> 5, 8, 5, 5, 5, 5, 5, 6, 5, 6, 5, 5, 6, 5, 5, 8, 7, 5, 5, ~
## $ YearBuilt     <int> 2003, 1976, 2001, 1915, 2000, 1993, 2004, 1973, 1931, 19~
## $ YearRemodAdd  <int> 2003, 1976, 2002, 1970, 2000, 1995, 2005, 1973, 1950, 19~
## $ RoofStyle     <fct> Gable, Gable, Gable, Gable, Gable, Gable, Gable, Gable, ~
## $ RoofMatl      <fct> CompShg, CompShg, CompShg, CompShg, CompShg, CompShg, Co~
## $ Exterior1st   <fct> VinylSd, MetalSd, VinylSd, Wd Sdng, VinylSd, VinylSd, Vi~
## $ Exterior2nd   <fct> VinylSd, MetalSd, VinylSd, Wd Shng, VinylSd, VinylSd, Vi~
## $ MasVnrType    <fct> BrkFace, None, BrkFace, None, BrkFace, None, Stone, Ston~
## $ MasVnrArea    <int> 196, 0, 162, 0, 350, 0, 186, 240, 0, 0, 0, 286, 0, 306, ~
## $ ExterQual     <fct> Gd, TA, Gd, TA, Gd, TA, Gd, TA, TA, TA, TA, Ex, TA, Gd, ~
## $ ExterCond     <fct> TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, ~
## $ Foundation    <fct> PConc, CBlock, PConc, BrkTil, PConc, Wood, PConc, CBlock~
## $ BsmtQual      <fct> Gd, Gd, Gd, TA, Gd, Gd, Ex, Gd, TA, TA, TA, Ex, TA, Gd, ~
## $ BsmtCond      <fct> TA, TA, TA, Gd, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, ~
## $ BsmtExposure  <fct> No, Gd, Mn, No, Av, No, Av, Mn, No, No, No, No, No, Av, ~
## $ BsmtFinType1  <fct> GLQ, ALQ, GLQ, ALQ, GLQ, GLQ, GLQ, ALQ, Unf, GLQ, Rec, G~
## $ BsmtFinSF1    <int> 706, 978, 486, 216, 655, 732, 1369, 859, 0, 851, 906, 99~
## $ BsmtFinType2  <fct> Unf, Unf, Unf, Unf, Unf, Unf, Unf, BLQ, Unf, Unf, Unf, U~
## $ BsmtFinSF2    <int> 0, 0, 0, 0, 0, 0, 0, 32, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

```
## $ BsmtUnfSF     <int> 150, 284, 434, 540, 490, 64, 317, 216, 952, 140, 134, 17~
## $ TotalBsmtSF   <int> 856, 1262, 920, 756, 1145, 796, 1686, 1107, 952, 991, 10~
## $ Heating       <fct> GasA, GasA, GasA, GasA, GasA, GasA, GasA, GasA, GasA, Ga~
## $ HeatingQC     <fct> Ex, Ex, Ex, Gd, Ex, Ex, Ex, Ex, Gd, Ex, Ex, Ex, TA, Ex, ~
## $ CentralAir    <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, ~
## $ X1stFlrSF     <int> 856, 1262, 920, 961, 1145, 796, 1694, 1107, 1022, 1077, ~
## $ X2ndFlrSF     <int> 854, 0, 866, 756, 1053, 566, 0, 983, 752, 0, 0, 1142, 0, ~
## $ LowQualFinSF  <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ GrLivArea     <int> 1710, 1262, 1786, 1717, 2198, 1362, 1694, 2090, 1774, 10~
## $ BsmtFullBath  <int> 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, ~
## $ BsmtHalfBath  <int> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ FullBath      <int> 2, 2, 2, 1, 2, 1, 2, 2, 2, 1, 1, 3, 1, 2, 1, 1, 1, 2, 1, ~
## $ HalfBath      <int> 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, ~
## $ BedroomAbvGr  <int> 3, 3, 3, 3, 4, 1, 3, 3, 2, 2, 3, 4, 2, 3, 2, 2, 2, 2, 3, ~
## $ KitchenAbvGr  <int> 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, ~
## $ KitchenQual   <fct> Gd, TA, Gd, Gd, Gd, TA, Gd, TA, TA, TA, TA, Ex, TA, Gd, ~
## $ TotRmsAbvGrd  <int> 8, 6, 6, 7, 9, 5, 7, 7, 8, 5, 5, 11, 4, 7, 5, 5, 5, 6, 6~
## $ Functional    <fct> Typ, Typ, Typ, Typ, Typ, Typ, Typ, Typ, Min1, Typ, Typ, ~
## $ Fireplaces    <int> 0, 1, 1, 1, 1, 0, 1, 2, 2, 2, 0, 2, 0, 1, 1, 0, 1, 0, 0, ~
## $ FireplaceQu   <fct> NA, TA, TA, Gd, TA, NA, Gd, TA, TA, TA, NA, Gd, NA, Gd, ~
## $ GarageType    <fct> Attchd, Attchd, Attchd, Detchd, Attchd, Attchd, Attchd, ~
## $ GarageYrBlt   <int> 2003, 1976, 2001, 1998, 2000, 1993, 2004, 1973, 1931, 19~
## $ GarageFinish  <fct> RFn, RFn, RFn, Unf, RFn, Unf, RFn, RFn, Unf, RFn, Unf, F~
## $ GarageCars    <int> 2, 2, 2, 3, 3, 2, 2, 2, 2, 1, 1, 3, 1, 3, 1, 2, 2, 2, 2, ~
## $ GarageArea    <int> 548, 460, 608, 642, 836, 480, 636, 484, 468, 205, 384, 7~
## $ GarageQual    <fct> TA, TA, TA, TA, TA, TA, TA, TA, Fa, Gd, TA, TA, TA, TA, ~
## $ GarageCond    <fct> TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, TA, ~
## $ PavedDrive    <fct> Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, Y, ~
## $ WoodDeckSF    <int> 0, 298, 0, 0, 192, 40, 255, 235, 90, 0, 0, 147, 140, 160~
## $ OpenPorchSF   <int> 61, 0, 42, 35, 84, 30, 57, 204, 0, 4, 0, 21, 0, 33, 213, ~
## $ EnclosedPorch <int> 0, 0, 0, 272, 0, 0, 0, 228, 205, 0, 0, 0, 0, 0, 176, 0, ~
## $ X3SsnPorch    <int> 0, 0, 0, 0, 0, 320, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ ScreenPorch   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 176, 0, 0, 0, 0, 0, ~
## $ PoolArea      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ PoolQC        <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Fence         <fct> NA, NA, NA, NA, NA, MnPrv, NA, NA, NA, NA, NA, NA, NA, N~
## $ MiscFeature   <fct> NA, NA, NA, NA, NA, Shed, NA, Shed, NA, NA, NA, NA, NA, ~
## $ MiscVal       <int> 0, 0, 0, 0, 0, 700, 0, 350, 0, 0, 0, 0, 0, 0, 0, 0, 700, ~
## $ MoSold        <int> 2, 5, 9, 2, 12, 10, 8, 11, 4, 1, 2, 7, 9, 8, 5, 7, 3, 10~
## $ YrSold        <int> 2008, 2007, 2008, 2006, 2008, 2009, 2007, 2009, 2008, 20~
## $ SaleType      <fct> WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, WD, New, WD, New~
## $ SaleCondition <fct> Normal, Normal, Normal, Abnorml, Normal, Normal, Normal, ~
## $ SalePrice     <int> 208500, 181500, 223500, 140000, 250000, 143000, 307000, ~
```

Memeriksa Gambaran Umum Data

```
plot_intro(data = data_house,
           geom_label_args = list(size=2.5))
```



Catatan:

- `plot_intro` merupakan fungsi yang berasal dari package `DataExplorer` dan argumen utamanya adalah object berbentuk `data.frame`.
- argumen `geom_label_args` bisa diisi dengan opsi-opsi yang ada pada fungsi `geom_label` pada pacakge `ggplot2`.

```
skim_without_charts(data = data_house)
```

Table 1: Data summary

Name	data_house
Number of rows	1460
Number of columns	81
Column type frequency:	
factor	43
numeric	38
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
MSZoning	0	1.00	FALSE	5	RL: 1151, RM: 218, FV: 65, RH: 16
Street	0	1.00	FALSE	2	Pav: 1454, Grv: 6

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
Alley	1369	0.06	FALSE	2	Grv: 50, Pav: 41
LotShape	0	1.00	FALSE	4	Reg: 925, IR1: 484, IR2: 41, IR3: 10
LandContour	0	1.00	FALSE	4	Lvl: 1311, Bnk: 63, HLS: 50, Low: 36
Utilities	0	1.00	FALSE	2	All: 1459, NoS: 1
LotConfig	0	1.00	FALSE	5	Ins: 1052, Cor: 263, Cul: 94, FR2: 47
LandSlope	0	1.00	FALSE	3	Gtl: 1382, Mod: 65, Sev: 13
Neighborhood	0	1.00	FALSE	25	NAm: 225, Col: 150, Old: 113, Edw: 100
Condition1	0	1.00	FALSE	9	Nor: 1260, Fee: 81, Art: 48, RRA: 26
Condition2	0	1.00	FALSE	8	Nor: 1445, Fee: 6, Art: 2, Pos: 2
BldgType	0	1.00	FALSE	5	1Fa: 1220, Twn: 114, Dup: 52, Twn: 43
HouseStyle	0	1.00	FALSE	8	1St: 726, 2St: 445, 1.5: 154, SLv: 65
RoofStyle	0	1.00	FALSE	6	Gab: 1141, Hip: 286, Fla: 13, Gam: 11
RoofMatl	0	1.00	FALSE	8	Com: 1434, Tar: 11, WdS: 6, WdS: 5
Exterior1st	0	1.00	FALSE	15	Vin: 515, HdB: 222, Met: 220, Wd : 206
Exterior2nd	0	1.00	FALSE	16	Vin: 504, Met: 214, HdB: 207, Wd : 197
MasVnrType	8	0.99	FALSE	4	Non: 864, Brk: 445, Sto: 128, Brk: 15
ExterQual	0	1.00	FALSE	4	TA: 906, Gd: 488, Ex: 52, Fa: 14
ExterCond	0	1.00	FALSE	5	TA: 1282, Gd: 146, Fa: 28, Ex: 3
Foundation	0	1.00	FALSE	6	PCo: 647, CBl: 634, Brk: 146, Sla: 24
BsmtQual	37	0.97	FALSE	4	TA: 649, Gd: 618, Ex: 121, Fa: 35
BsmtCond	37	0.97	FALSE	4	TA: 1311, Gd: 65, Fa: 45, Po: 2
BsmtExposure	38	0.97	FALSE	4	No: 953, Av: 221, Gd: 134, Mn: 114
BsmtFinType1	37	0.97	FALSE	6	Unf: 430, GLQ: 418, ALQ: 220, BLQ: 148
BsmtFinType2	38	0.97	FALSE	6	Unf: 1256, Rec: 54, LwQ: 46, BLQ: 33
Heating	0	1.00	FALSE	6	Gas: 1428, Gas: 18, Gra: 7, Wal: 4
HeatingQC	0	1.00	FALSE	5	Ex: 741, TA: 428, Gd: 241, Fa: 49
CentralAir	0	1.00	FALSE	2	Y: 1365, N: 95
Electrical	1	1.00	FALSE	5	SBr: 1334, Fus: 94, Fus: 27, Fus: 3
KitchenQual	0	1.00	FALSE	4	TA: 735, Gd: 586, Ex: 100, Fa: 39
Functional	0	1.00	FALSE	7	Typ: 1360, Min: 34, Min: 31, Mod: 15
FireplaceQu	690	0.53	FALSE	5	Gd: 380, TA: 313, Fa: 33, Ex: 24
GarageType	81	0.94	FALSE	6	Att: 870, Det: 387, Bui: 88, Bas: 19
GarageFinish	81	0.94	FALSE	3	Unf: 605, RFn: 422, Fin: 352
GarageQual	81	0.94	FALSE	5	TA: 1311, Fa: 48, Gd: 14, Ex: 3
GarageCond	81	0.94	FALSE	5	TA: 1326, Fa: 35, Gd: 9, Po: 7
PavedDrive	0	1.00	FALSE	3	Y: 1340, N: 90, P: 30
PoolQC	1453	0.00	FALSE	3	Gd: 3, Ex: 2, Fa: 2
Fence	1179	0.19	FALSE	4	MnP: 157, GdP: 59, GdW: 54, MnW: 11
MiscFeature	1406	0.04	FALSE	4	She: 49, Gar: 2, Oth: 2, Ten: 1
SaleType	0	1.00	FALSE	9	WD: 1267, New: 122, COD: 43, Con: 9
SaleCondition	0	1.00	FALSE	6	Nor: 1198, Par: 125, Abn: 101, Fam: 20

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1.00	730.50	421.61	1	365.75	730.5	1095.25	1460
MSSubClass	0	1.00	56.90	42.30	20	20.00	50.0	70.00	190
LotFrontage	259	0.82	70.05	24.28	21	59.00	69.0	80.00	313
LotArea	0	1.00	10516.83	9981.26	1300	7553.50	9478.5	11601.50	215245
OverallQual	0	1.00	6.10	1.38	1	5.00	6.0	7.00	10
OverallCond	0	1.00	5.58	1.11	1	5.00	5.0	6.00	9
YearBuilt	0	1.00	1971.27	30.20	1872	1954.00	1973.0	2000.00	2010
YearRemodAdd	0	1.00	1984.87	20.65	1950	1967.00	1994.0	2004.00	2010
MasVnrArea	8	0.99	103.69	181.07	0	0.00	0.0	166.00	1600
BsmtFinSF1	0	1.00	443.64	456.10	0	0.00	383.5	712.25	5644
BsmtFinSF2	0	1.00	46.55	161.32	0	0.00	0.0	0.00	1474
BsmtUnfSF	0	1.00	567.24	441.87	0	223.00	477.5	808.00	2336
TotalBsmtSF	0	1.00	1057.43	438.71	0	795.75	991.5	1298.25	6110
X1stFlrSF	0	1.00	1162.63	386.59	334	882.00	1087.0	1391.25	4692
X2ndFlrSF	0	1.00	346.99	436.53	0	0.00	0.0	728.00	2065
LowQualFinSF	0	1.00	5.84	48.62	0	0.00	0.0	0.00	572
GrLivArea	0	1.00	1515.46	525.48	334	1129.50	1464.0	1776.75	5642
BsmtFullBath	0	1.00	0.43	0.52	0	0.00	0.0	1.00	3
BsmtHalfBath	0	1.00	0.06	0.24	0	0.00	0.0	0.00	2
FullBath	0	1.00	1.57	0.55	0	1.00	2.0	2.00	3
HalfBath	0	1.00	0.38	0.50	0	0.00	0.0	1.00	2
BedroomAbvGr	0	1.00	2.87	0.82	0	2.00	3.0	3.00	8
KitchenAbvGr	0	1.00	1.05	0.22	0	1.00	1.0	1.00	3
TotRmsAbvGrd	0	1.00	6.52	1.63	2	5.00	6.0	7.00	14
Fireplaces	0	1.00	0.61	0.64	0	0.00	1.0	1.00	3
GarageYrBlt	81	0.94	1978.51	24.69	1900	1961.00	1980.0	2002.00	2010
GarageCars	0	1.00	1.77	0.75	0	1.00	2.0	2.00	4
GarageArea	0	1.00	472.98	213.80	0	334.50	480.0	576.00	1418
WoodDeckSF	0	1.00	94.24	125.34	0	0.00	0.0	168.00	857
OpenPorchSF	0	1.00	46.66	66.26	0	0.00	25.0	68.00	547
EnclosedPorch	0	1.00	21.95	61.12	0	0.00	0.0	0.00	552
X3SsnPorch	0	1.00	3.41	29.32	0	0.00	0.0	0.00	508
ScreenPorch	0	1.00	15.06	55.76	0	0.00	0.0	0.00	480
PoolArea	0	1.00	2.76	40.18	0	0.00	0.0	0.00	738
MiscVal	0	1.00	43.49	496.12	0	0.00	0.0	0.00	15500
MoSold	0	1.00	6.32	2.70	1	5.00	6.0	8.00	12
YrSold	0	1.00	2007.82	1.33	2006	2007.00	2008.0	2009.00	2010
SalePrice	0	1.00	180921.20	79442.50	34900	129975.00	163000.0	214000.00	755000

Catatan:

- `skim_without_charts` merupakan fungsi yang berasal dari package `skimr` dan argumen utamanya adalah object berbentuk `data.frame`.

Berdasarkan informasi diatas, kita tahu terdata beberapa kolom yang memiliki missing value. Namun hanya 5 kolom saja yang mengalami banyak missing value yaitu `Alley`,`FireplaceQu`,`PoolQX`,`Fence`,`MiscFeature`.

Menangani Missing Value

Dalam kasus ini kita akan menangani missing value dengan dua cara, yaitu

1. Mereplace missing value pada kolom-kolom yang memiliki banyak sekali missing value (diatas 500)

2. Menghapus baris-baris yang mengandung missing value pada kolom-kolom yang memiliki sedikti missing value (dibawah 500)

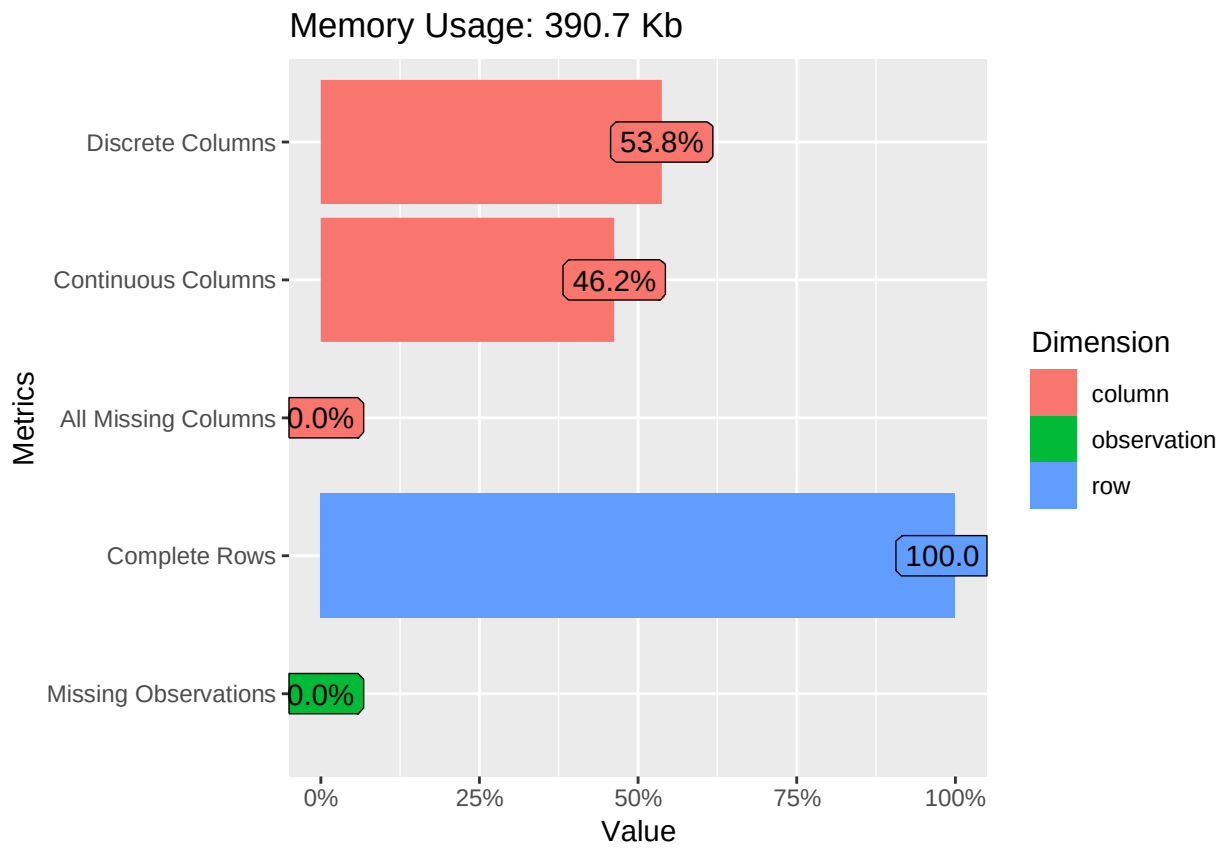
Berikut dibawah ini adalah sintaks untuk melakukan replace missing value, khususnya jika datanya berupa `factor` atau `string`. Kemudian `na.omit` digunakan untuk menghapus semua baris yang mengandung missing value

```
data_house1 <- data_house %>%
  select(-Id) %>%
  mutate(
    Alley = forcats::fct_explicit_na(Alley, na_level = "Ukn"),
    FireplaceQu=forcats::fct_explicit_na(FireplaceQu,
                                         na_level = "Ukn"
                                           ),
    PoolQC = forcats::fct_explicit_na(PoolQC, na_level = "Ukn"),
    Fence = forcats::fct_explicit_na(Fence, na_level = "Ukn"),
    MiscFeature = forcats::fct_explicit_na(MiscFeature, na_level = "Ukn")
  ) %>% na.omit
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `Alley = forcats::fct_explicit_na(Alley, na_level = "Ukn")`.
## Caused by warning:
## ! `fct_explicit_na()` was deprecated in forcats 1.0.0.
## i Please use `fct_na_value_to_level()` instead.
```

Kemudian kita akan lihat kembali data yang sudah kita tangani missing valuenya

```
plot_intro(data = data_house1)
```



```
skim_without_charts(data_house1)
```

Table 4: Data summary

Name	data_house1
Number of rows	1094
Number of columns	80
Column type frequency:	
factor	43
numeric	37
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
MSZoning	0	1	FALSE	5	RL: 850, RM: 173, FV: 54, RH: 9
Street	0	1	FALSE	2	Pav: 1090, Grv: 4
Alley	0	1	FALSE	3	Ukn: 1017, Grv: 41, Pav: 36
LotShape	0	1	FALSE	4	Reg: 760, IR1: 301, IR2: 26, IR3: 7
LandContour	0	1	FALSE	4	Lvl: 991, Bnk: 45, HLS: 44, Low: 14
Utilities	0	1	FALSE	1	All: 1094, NoS: 0
LotConfig	0	1	FALSE	5	Ins: 830, Cor: 187, Cul: 44, FR2: 29
LandSlope	0	1	FALSE	3	Gtl: 1045, Mod: 44, Sev: 5
Neighborhood	0	1	FALSE	25	NAm: 173, Col: 122, Old: 96, Som: 75
Condition1	0	1	FALSE	9	Nor: 950, Fee: 52, Art: 42, RRA: 24
Condition2	0	1	FALSE	6	Nor: 1082, Fee: 5, Art: 2, Pos: 2
BldgType	0	1	FALSE	5	1Fa: 925, Twn: 90, Twn: 35, Dup: 24
HouseStyle	0	1	FALSE	8	1St: 540, 2St: 346, 1.5: 117, SLv: 43
RoofStyle	0	1	FALSE	5	Gab: 843, Hip: 230, Gam: 10, Man: 6
RoofMatl	0	1	FALSE	7	Com: 1078, WdS: 6, Tar: 5, WdS: 2
Exterior1st	0	1	FALSE	14	Vin: 421, Met: 172, HdB: 151, Wd : 149
Exterior2nd	0	1	FALSE	16	Vin: 412, Met: 169, Wd : 145, HdB: 138
MasVnrType	0	1	FALSE	4	Non: 639, Brk: 327, Sto: 119, Brk: 9
ExterQual	0	1	FALSE	4	TA: 646, Gd: 395, Ex: 46, Fa: 7
ExterCond	0	1	FALSE	4	TA: 973, Gd: 104, Fa: 15, Ex: 2
Foundation	0	1	FALSE	5	PCo: 518, CBl: 446, Brk: 122, Sto: 6
BsmtQual	0	1	FALSE	4	TA: 486, Gd: 463, Ex: 113, Fa: 32
BsmtCond	0	1	FALSE	4	TA: 1006, Gd: 51, Fa: 36, Po: 1
BsmtExposure	0	1	FALSE	4	No: 734, Av: 174, Gd: 97, Mn: 89
BsmtFinType1	0	1	FALSE	6	Unf: 343, GLQ: 323, ALQ: 162, BLQ: 105
BsmtFinType2	0	1	FALSE	6	Unf: 972, Rec: 37, LwQ: 35, BLQ: 25
Heating	0	1	FALSE	4	Gas: 1075, Gas: 16, Gra: 2, Oth: 1
HeatingQC	0	1	FALSE	5	Ex: 594, TA: 298, Gd: 174, Fa: 27
CentralAir	0	1	FALSE	2	Y: 1036, N: 58
Electrical	0	1	FALSE	5	SBr: 1009, Fus: 67, Fus: 15, Fus: 2
KitchenQual	0	1	FALSE	4	TA: 528, Gd: 454, Ex: 91, Fa: 21
Functional	0	1	FALSE	6	Typ: 1024, Min: 25, Min: 21, Maj: 10

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
FireplaceQu	0	1	FALSE	6	Ukn: 511, Gd: 315, TA: 212, Fa: 24
GarageType	0	1	FALSE	6	Att: 680, Det: 325, Bui: 63, Bas: 15
GarageFinish	0	1	FALSE	3	Unf: 485, RFn: 333, Fin: 276
GarageQual	0	1	FALSE	5	TA: 1031, Fa: 46, Gd: 11, Ex: 3
GarageCond	0	1	FALSE	5	TA: 1050, Fa: 31, Po: 6, Gd: 5
PavedDrive	0	1	FALSE	3	Y: 1023, N: 48, P: 23
PoolQC	0	1	FALSE	4	Ukn: 1088, Ex: 2, Fa: 2, Gd: 2
Fence	0	1	FALSE	5	Ukn: 882, MnP: 117, GdP: 46, GdW: 39
MiscFeature	0	1	FALSE	4	Ukn: 1059, She: 33, Oth: 1, Ten: 1
SaleType	0	1	FALSE	9	WD: 928, New: 116, COD: 31, Con: 5
SaleCondition	0	1	FALSE	6	Nor: 880, Par: 119, Abn: 70, Fam: 18

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
MSSubClass	0	1	56.13	41.98	20	20.00	50.0	70.00	190
LotFrontage	0	1	70.76	24.51	21	60.00	70.0	80.00	313
LotArea	0	1	10132.35	8212.25	1300	7606.75	9444.5	11387.25	215245
OverallQual	0	1	6.25	1.37	2	5.00	6.0	7.00	10
OverallCond	0	1	5.58	1.07	2	5.00	5.0	6.00	9
YearBuilt	0	1	1972.41	31.19	1880	1953.00	1975.0	2003.00	2010
YearRemodAdd	0	1	1985.92	20.93	1950	1967.00	1995.0	2005.00	2010
MasVnrArea	0	1	109.86	190.67	0	0.00	0.0	171.75	1600
BsmtFinSF1	0	1	448.19	468.73	0	0.00	384.5	712.75	5644
BsmtFinSF2	0	1	45.25	159.08	0	0.00	0.0	0.00	1474
BsmtUnfSF	0	1	606.12	445.83	0	270.00	525.0	846.00	2336
TotalBsmtSF	0	1	1099.56	415.85	105	816.00	1023.0	1345.50	6110
X1stFlrSF	0	1	1173.81	387.68	438	894.00	1097.0	1413.50	4692
X2ndFlrSF	0	1	356.54	439.26	0	0.00	0.0	729.00	2065
LowQualFinSF	0	1	4.68	42.10	0	0.00	0.0	0.00	572
GrLivArea	0	1	1535.03	526.12	438	1164.00	1480.0	1779.00	5642
BsmtFullBath	0	1	0.42	0.51	0	0.00	0.0	1.00	2
BsmtHalfBath	0	1	0.06	0.24	0	0.00	0.0	0.00	2
FullBath	0	1	1.58	0.55	0	1.00	2.0	2.00	3
HalfBath	0	1	0.39	0.50	0	0.00	0.0	1.00	2
BedroomAbvGr	0	1	2.86	0.76	0	2.00	3.0	3.00	6
KitchenAbvGr	0	1	1.03	0.19	1	1.00	1.0	1.00	3
TotRmsAbvGrd	0	1	6.57	1.58	3	5.00	6.0	7.00	12
Fireplaces	0	1	0.61	0.63	0	0.00	1.0	1.00	3
GarageYrBlt	0	1	1978.57	25.93	1900	1960.00	1982.0	2003.00	2010
GarageCars	0	1	1.88	0.66	1	1.00	2.0	2.00	4
GarageArea	0	1	503.76	192.26	160	360.00	484.0	602.50	1418
WoodDeckSF	0	1	94.34	122.62	0	0.00	0.0	169.75	857
OpenPorchSF	0	1	46.95	64.82	0	0.00	28.0	68.00	547
EnclosedPorch	0	1	22.05	61.57	0	0.00	0.0	0.00	552
X3SsnPorch	0	1	3.27	29.66	0	0.00	0.0	0.00	508
ScreenPorch	0	1	16.50	58.46	0	0.00	0.0	0.00	480
PoolArea	0	1	3.01	40.71	0	0.00	0.0	0.00	648
MiscVal	0	1	23.55	167.14	0	0.00	0.0	0.00	2500
MoSold	0	1	6.34	2.69	1	5.00	6.0	8.00	12

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
YrSold	0	1	2007.79	1.33	2006	2007.00	2008.0	2009.00	2010
SalePrice	0	1	187033.26	83165.33	35311	132500.00	165750.0	221000.00	755000

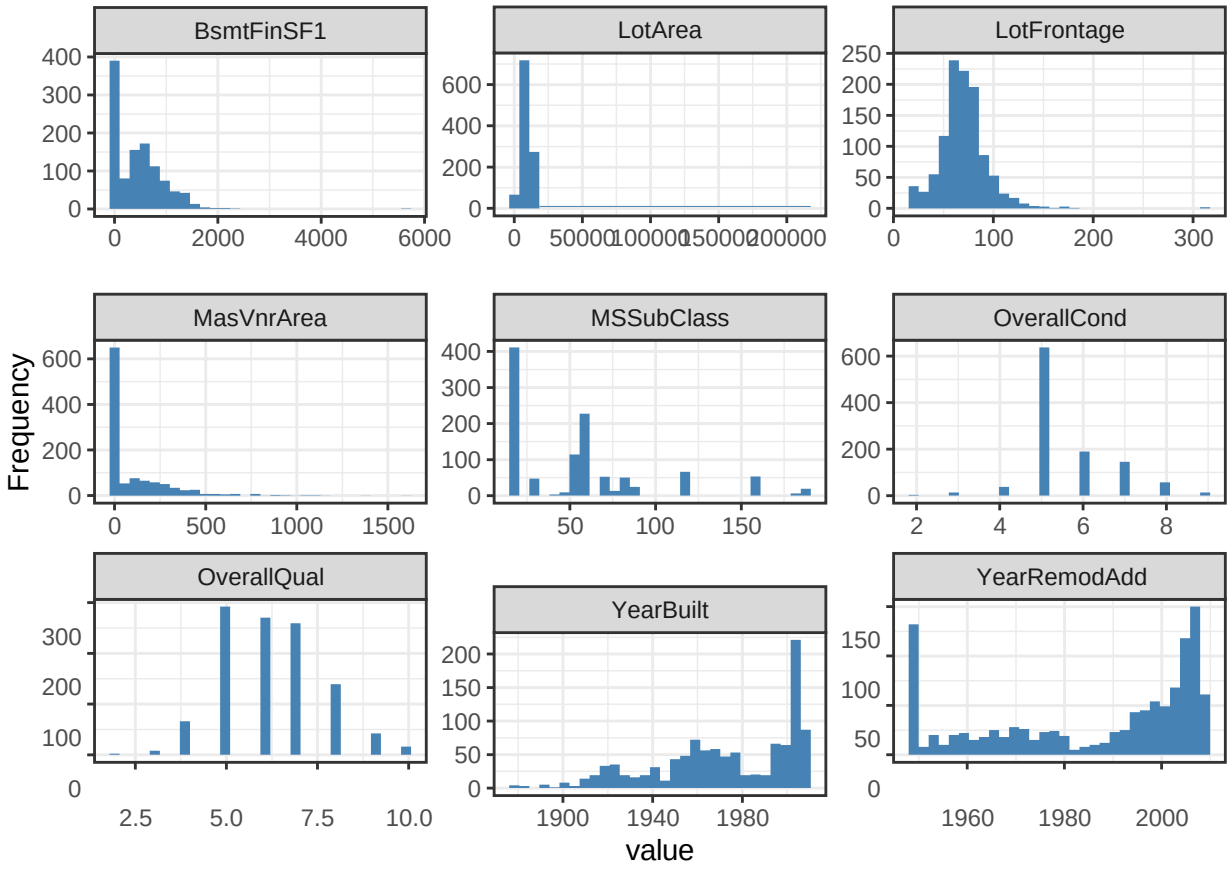
Setelah dilihat kembali ternyata ada kolom yang hanya memiliki satu kategori saja yaitu kolom `Utilities`. Sehingga kita perlu menghapusnya.

```
data_house1 <- data_house1 %>%
  select(-Utilities)
```

Memeriksa Sebaran Data

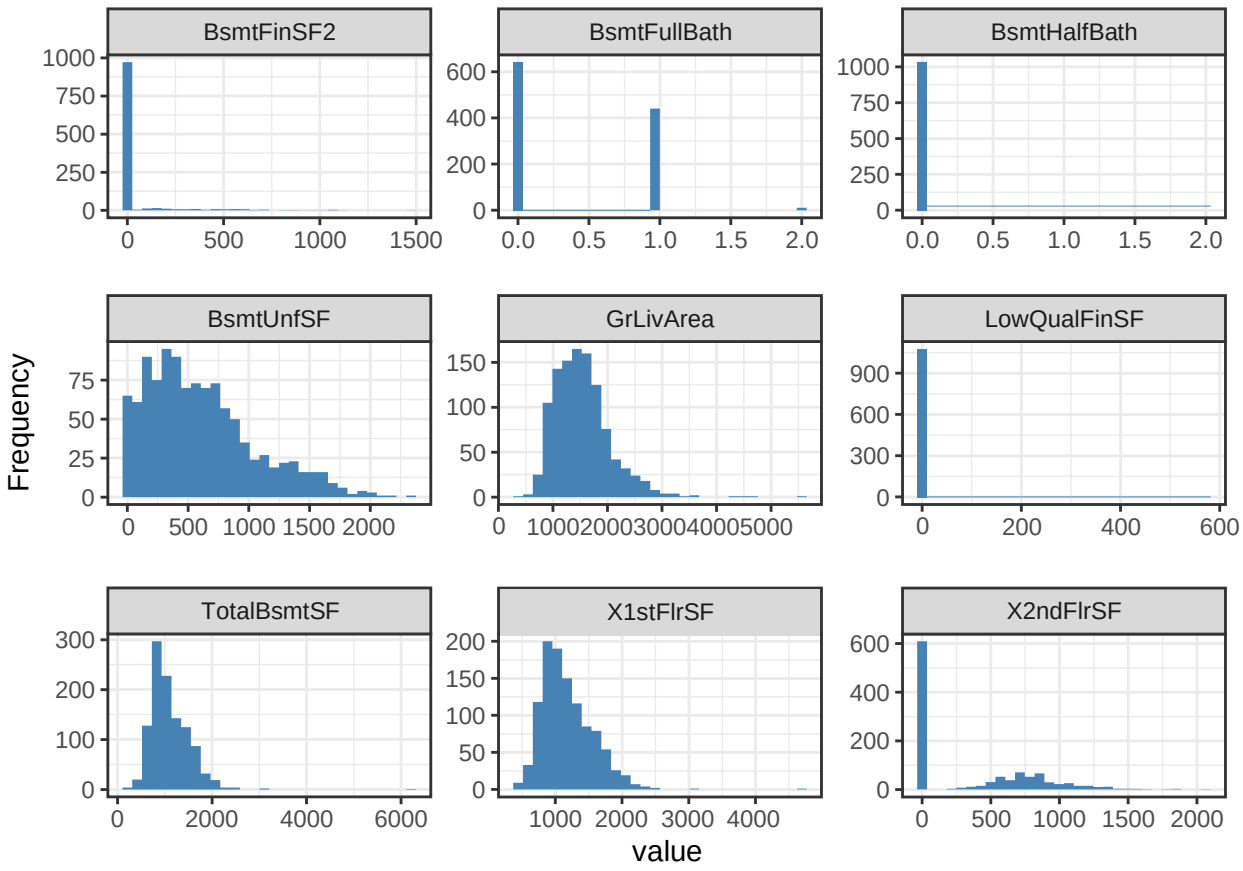
```
plot_histogram(data = data_house1,nrow=3,ncol = 3,
               geom_histogram_args = list(fill="steelblue"),
               ggtheme = theme_bw()
               )
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

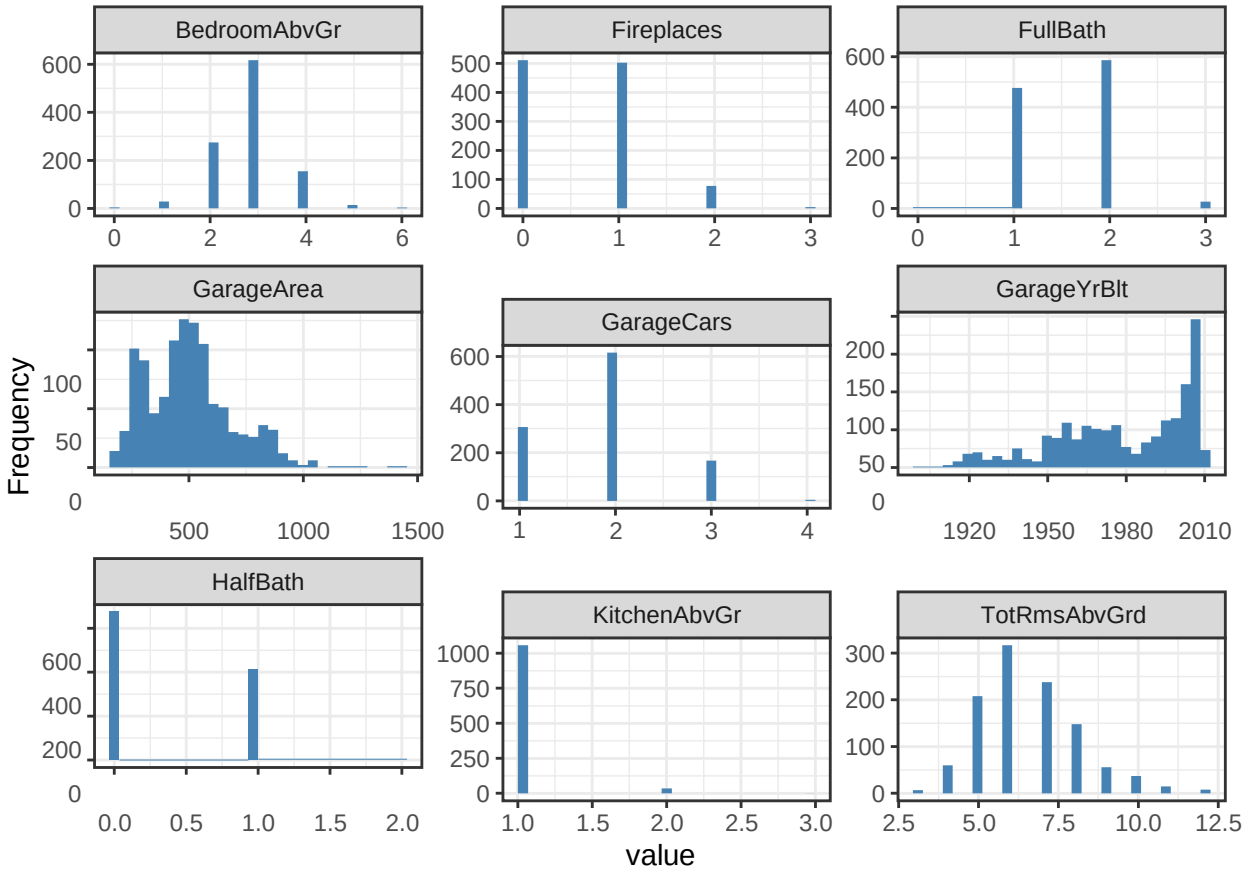


Page 1

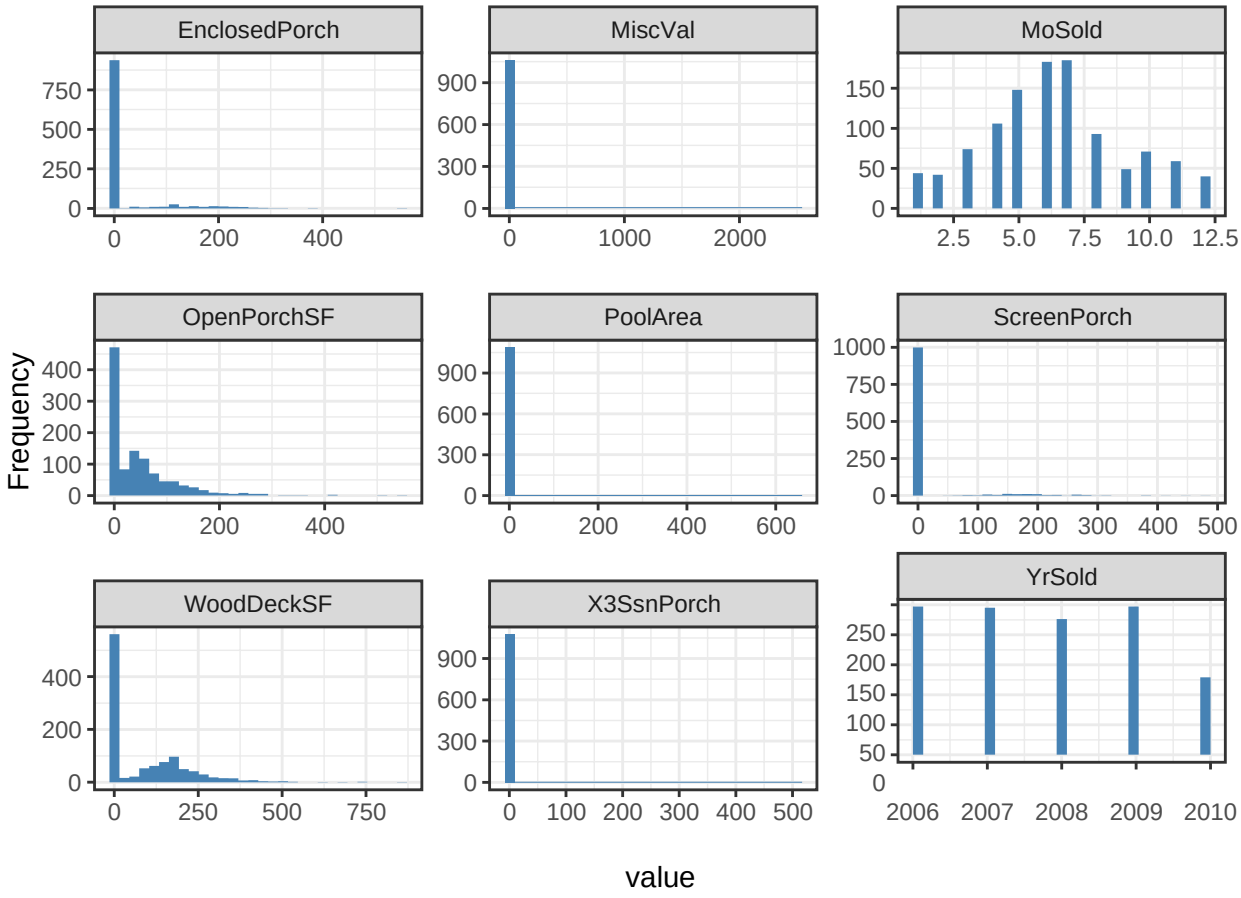
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



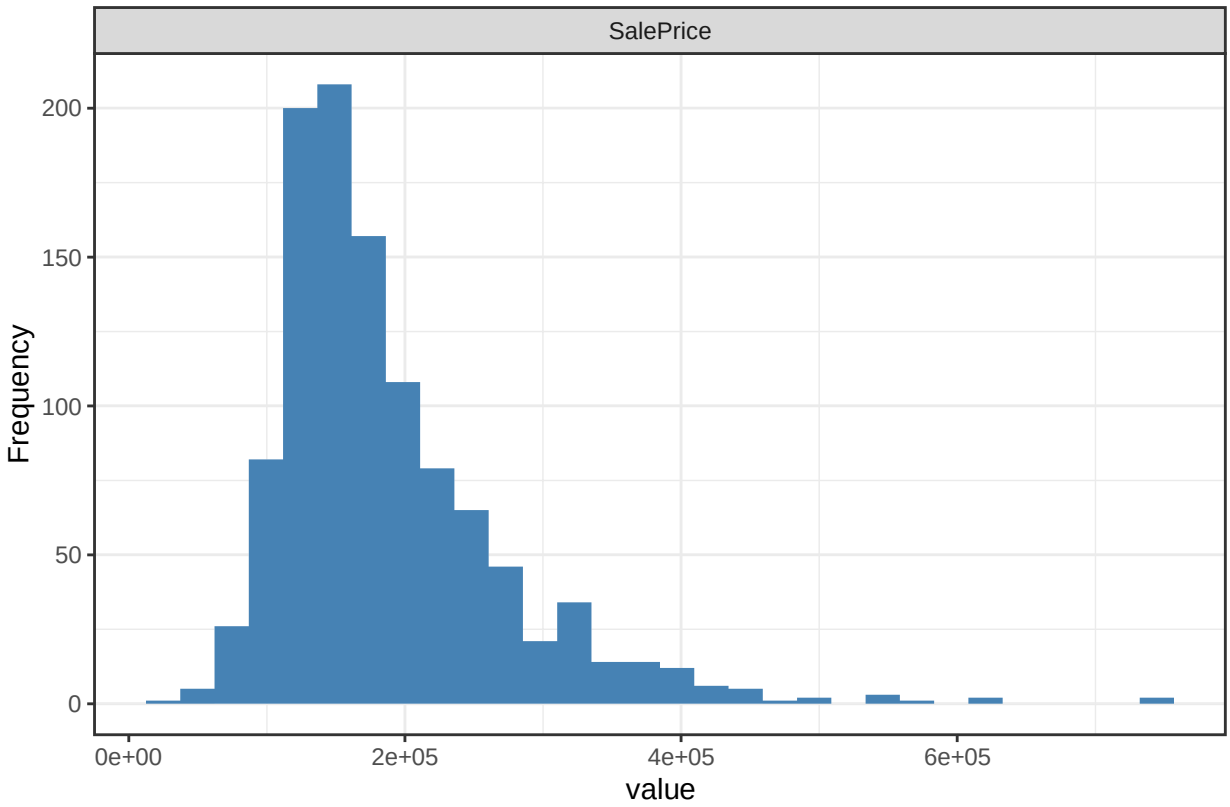

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



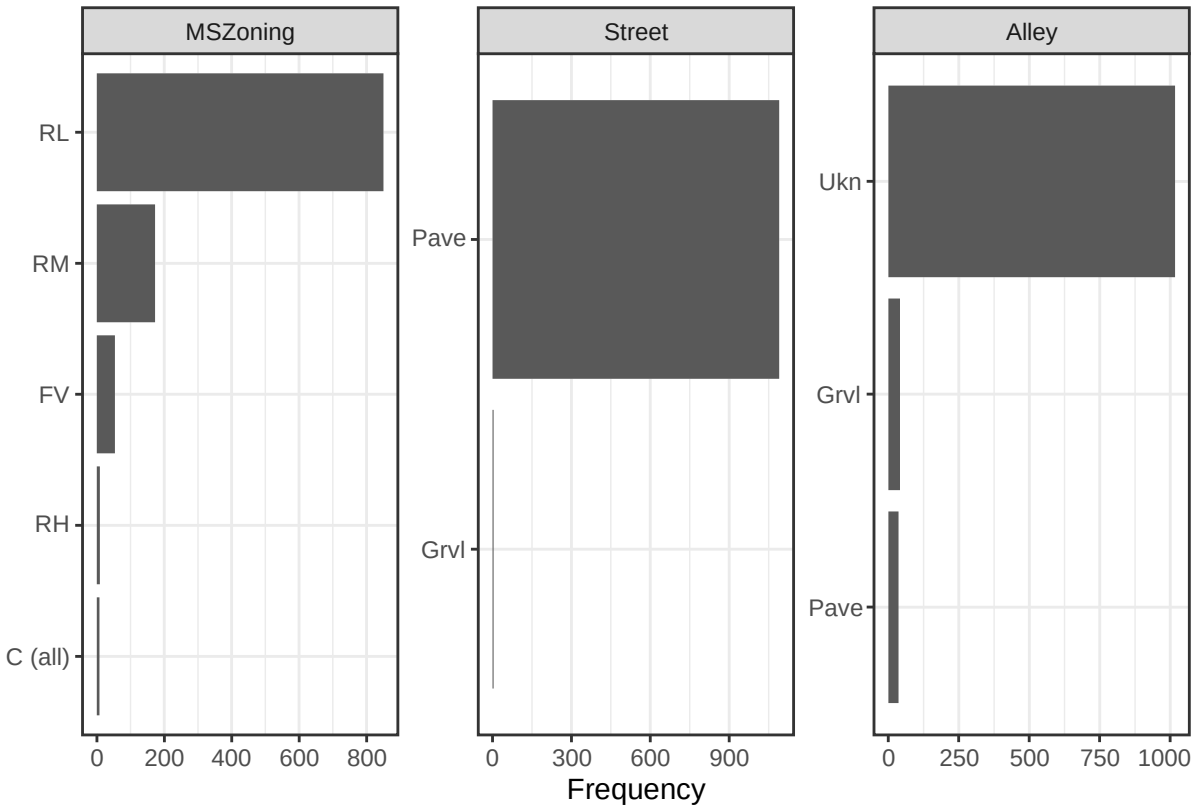

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



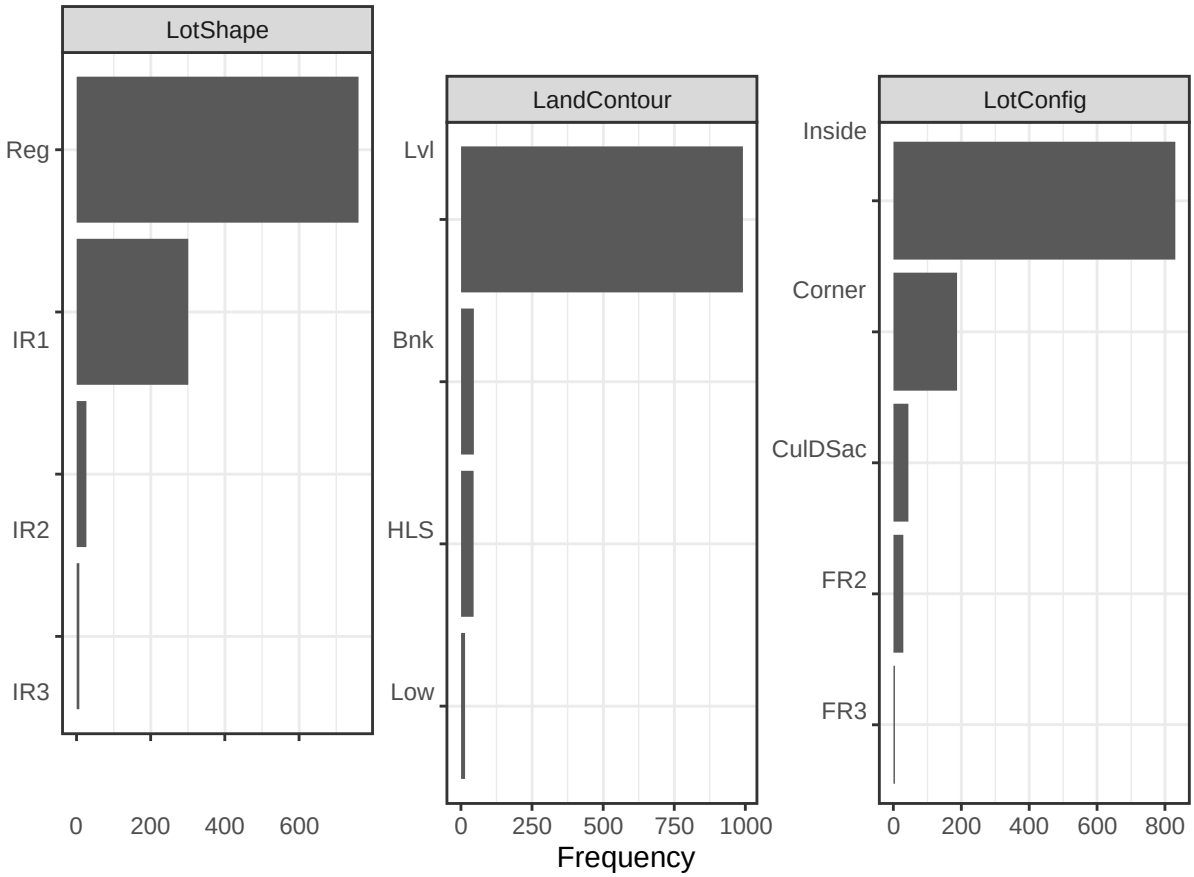

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

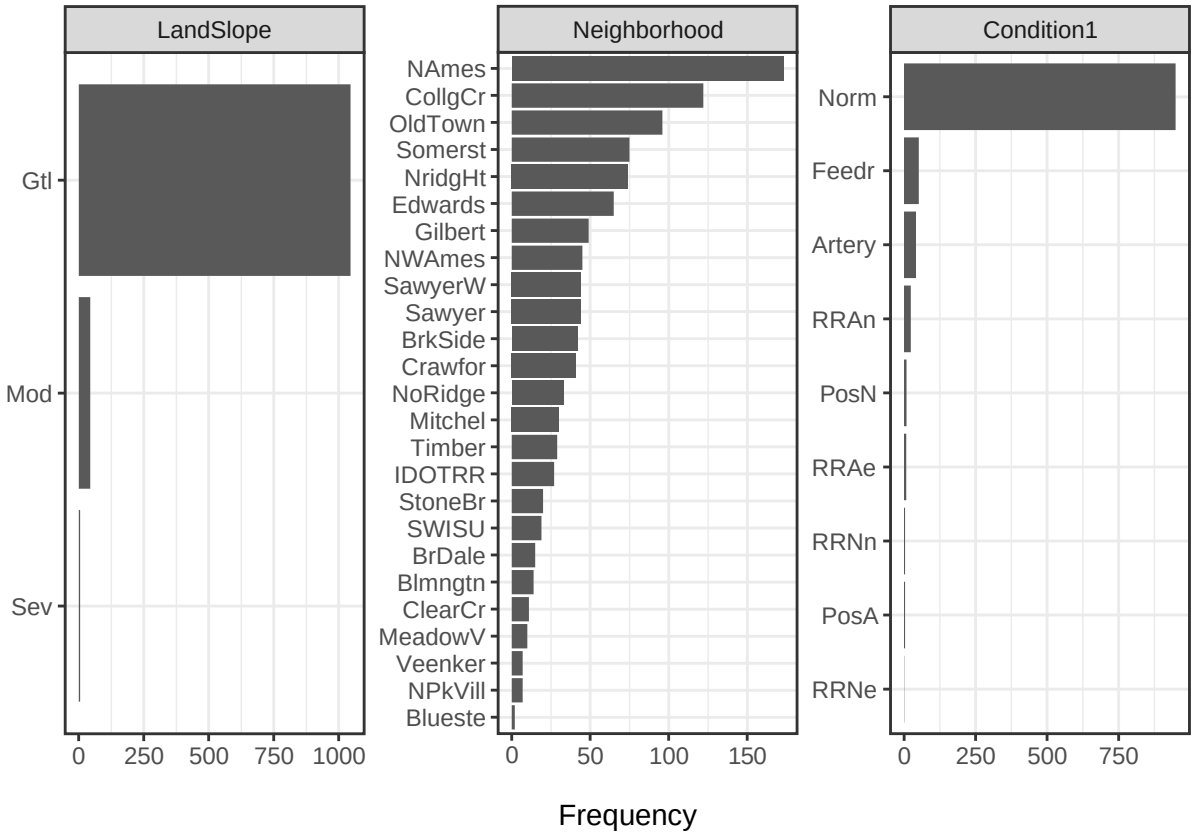


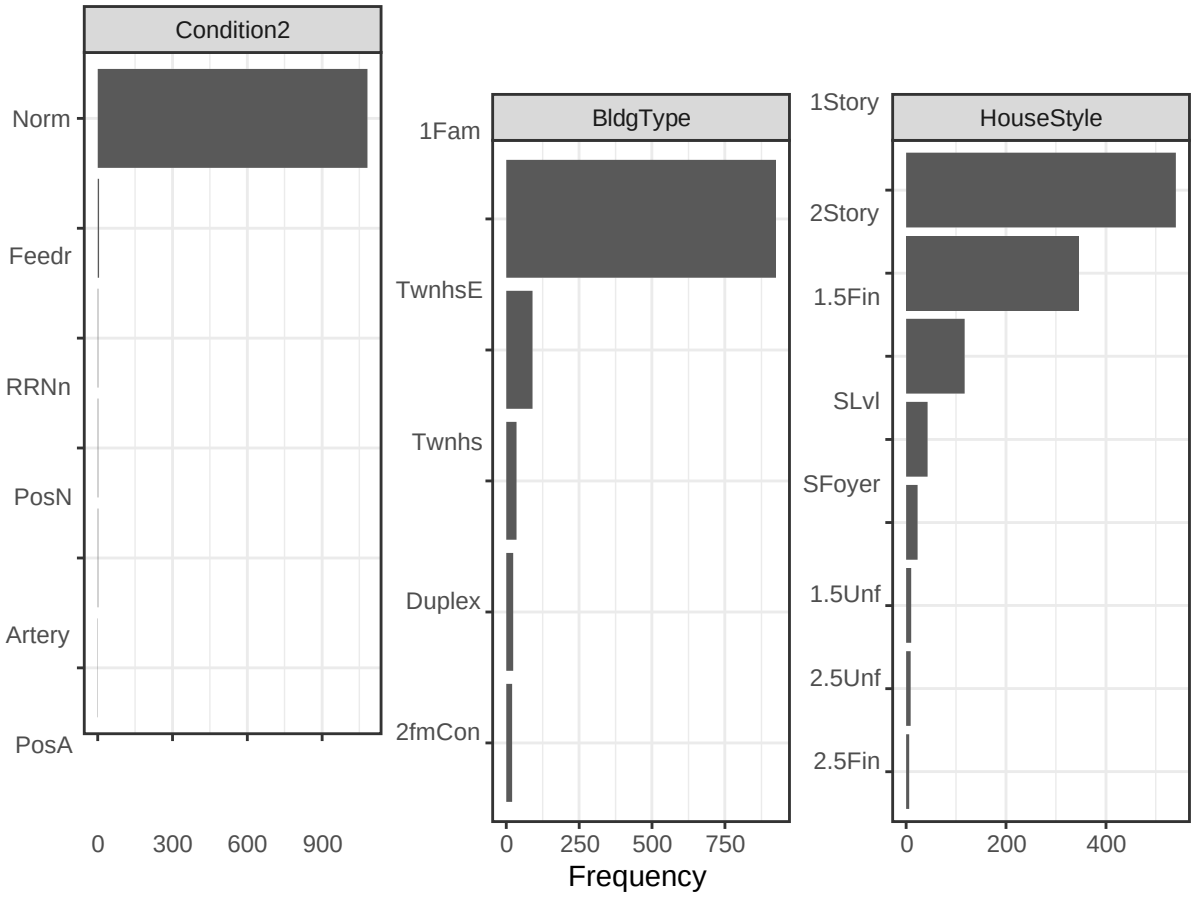

```
plot_bar(data = data_house1,ggtheme =theme_bw(),nrow = 1)
```

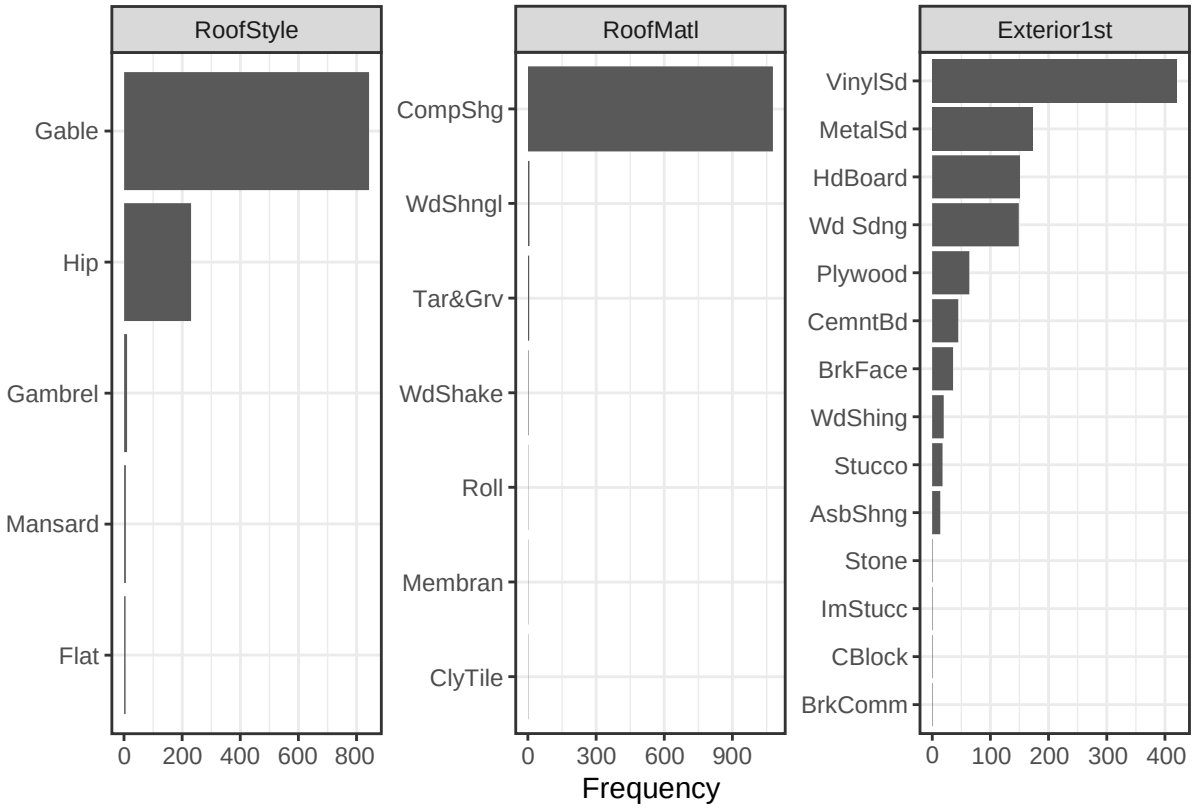

MSZoning
RL
RM
FV
RH
C (all)
0
200
400
600
800
Street
Pave
Grvl
0
300
600
900
Alley
Ukn
Grvl
Pave
0
250
500
750
1000
Frequency

Page 1

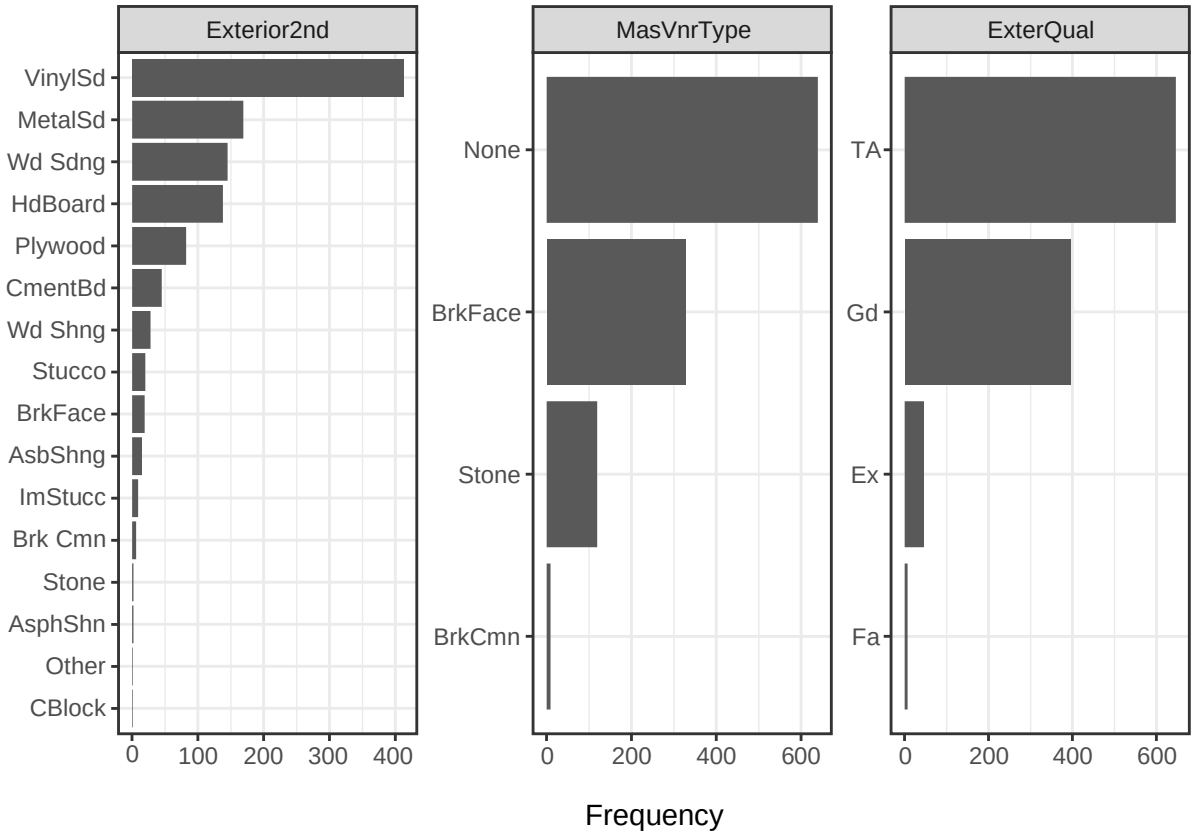

LotShape
Reg
IR1
IR2
IR3
0
200
400
600
LandContour
Lvl
Bnk
HLS
Low
0
250
500
750
1000
LotConfig
Inside
Corner
CulDSac
FR2
FR3
0
200
400
600
800
Frequency

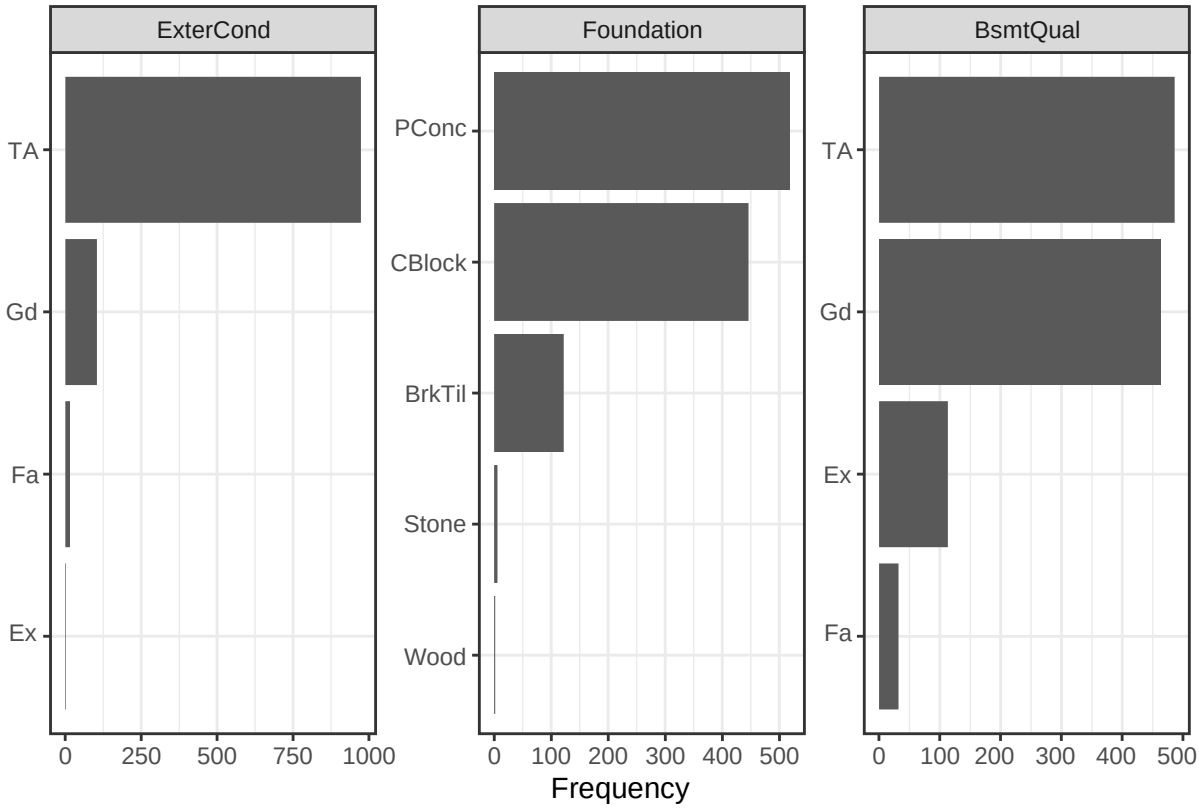

LandSlope
Gtl
Mod
Sev
0
250
500
750
1000
Neighborhood
NAmes
CollgCr
OldTown
Somerst
NridgHt
Edwards
Gilbert
NWAmes
SawyerW
Sawyer
BrkSide
Crawfor
NoRidge
Mitchel
Timber
IDOTRR
StoneBr
SWISU
BrDale
Blmngtn
ClearCr
MeadowV
Veenker
NPkVill
Blueste
0
50
100
150
Condition1
Norm
Feedr
Artery
RRAn
PosN
RRAe
RRNn
PosA
RRNe
0
250
500
750
Frequency

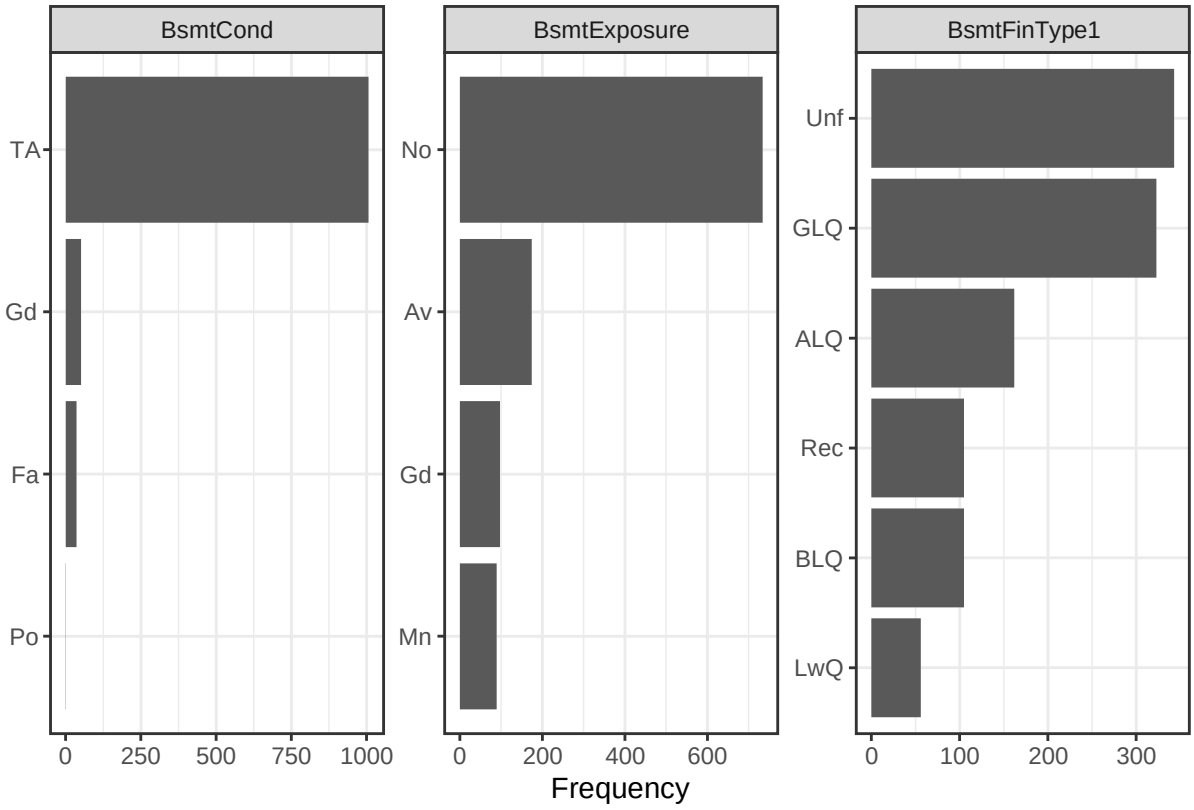

Condition2
Norm
Feedr
RRNn
PosN
Artery
PosA
0
300
600
900
BldgType
1Fam
TwnhsE
Twnhs
Duplex
2fmCon
0
250
500
750
HouseStyle
1Story
2Story
1.5Fin
SLvl
SFoyer
1.5Unf
2.5Unf
2.5Fin
0
200
400
Frequency

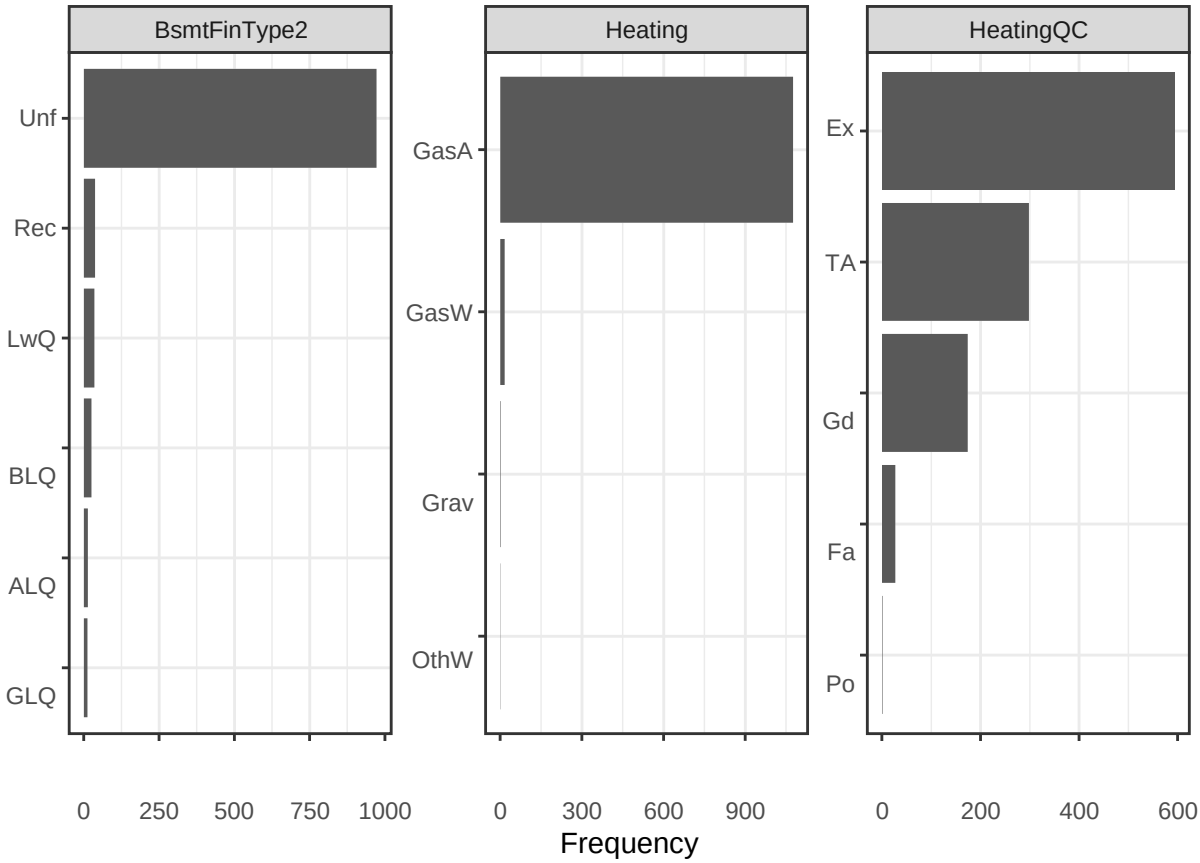

RoofStyle
Gable
Hip
Gambrel
Mansard
Flat
0
200
400
600
800
RoofMatl
CompShg
WdShngl
Tar&Grv
WdShake
Roll
Membran
ClyTile
0
300
600
900
Exterior1st
VinylSd
MetalSd
HdBoard
Wd Sdng
Plywood
CemntBd
BrkFace
WdShing
Stucco
AsbShng
Stone
ImStucc
CBlock
BrkComm
0
100
200
300
400
Frequency

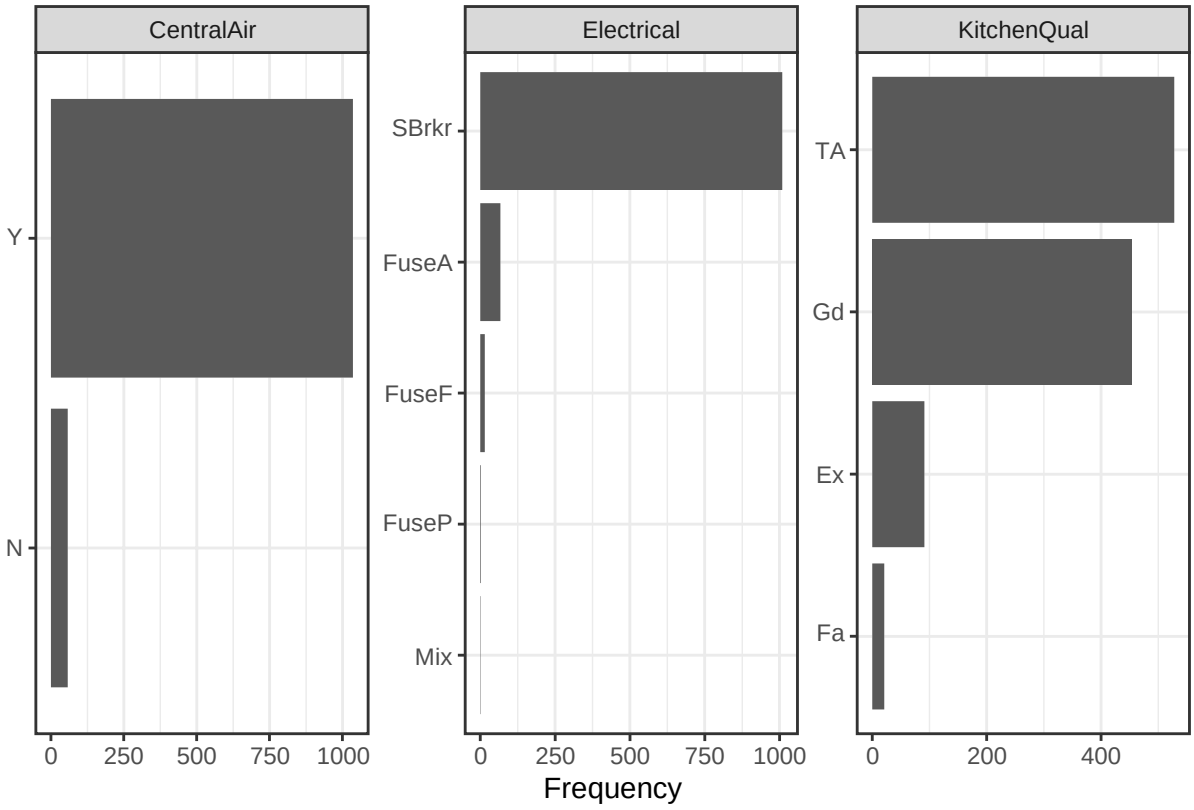
Page 5

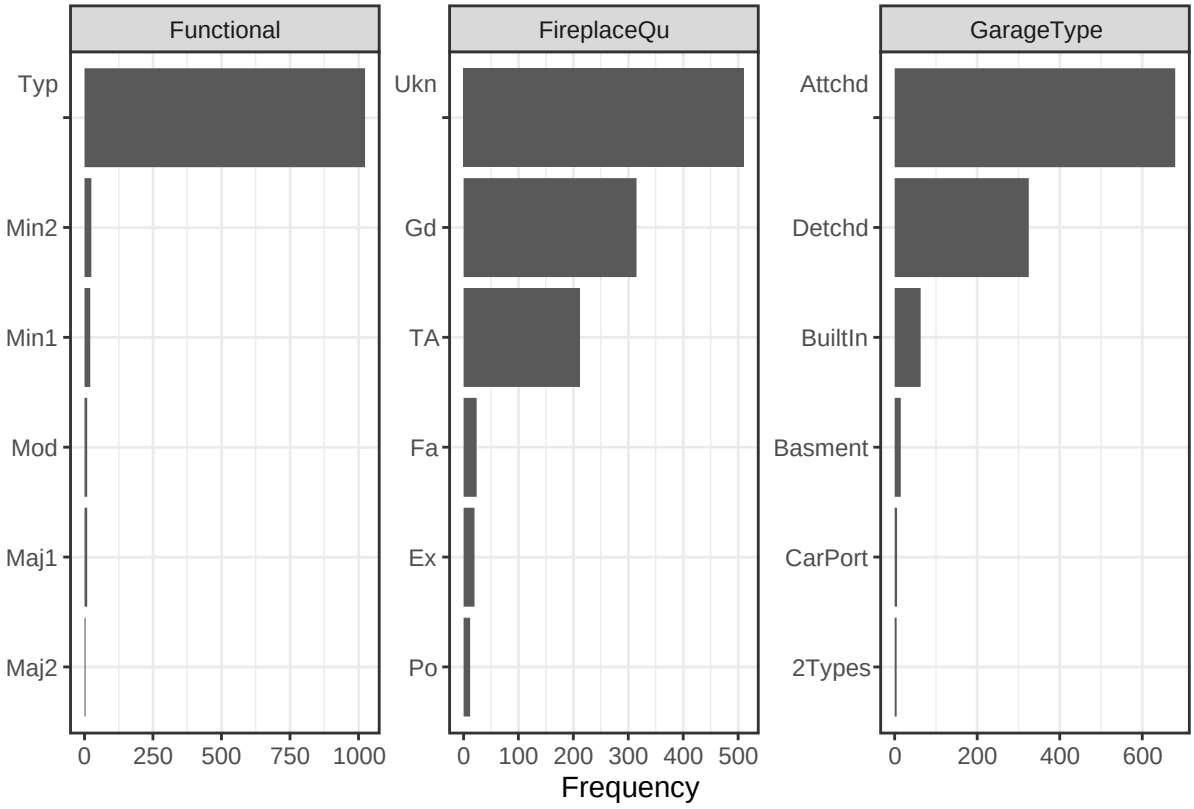

Exterior2nd
VinylSd
MetalSd
Wd Sdng
HdBoard
Plywood
CmentBd
Wd Shng
Stucco
BrkFace
AsbShng
ImStucc
Brk Cmn
Stone
AsphShn
Other
CBlock
0
100
200
300
400
MasVnrType
None
BrkFace
Stone
BrkCmn
0
200
400
600
ExterQual
TA
Gd
Ex
Fa
0
200
400
600
Frequency

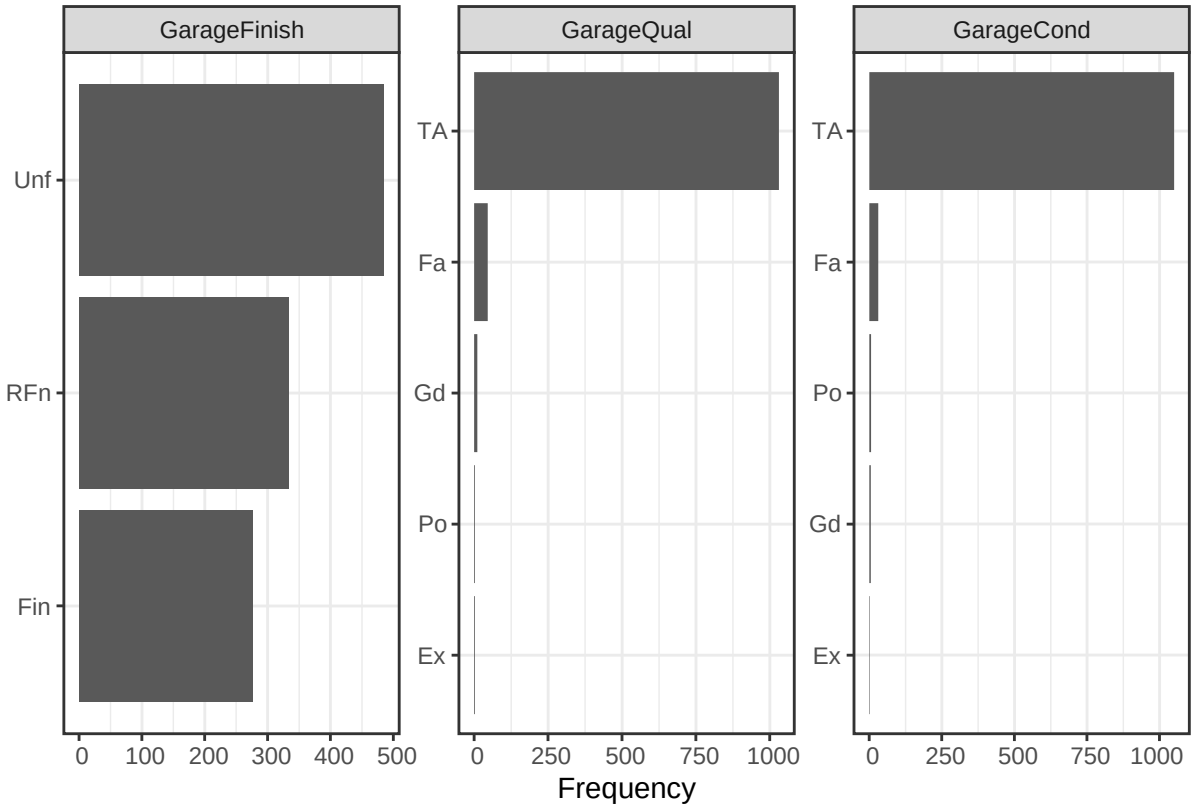

ExterCond
TA
Gd
Fa
Ex
0
250
500
750
1000
Foundation
PConc
CBlock
BrkTil
Stone
Wood
0
100
200
300
400
500
BsmtQual
TA
Gd
Ex
Fa
0
100
200
300
400
500
Frequency

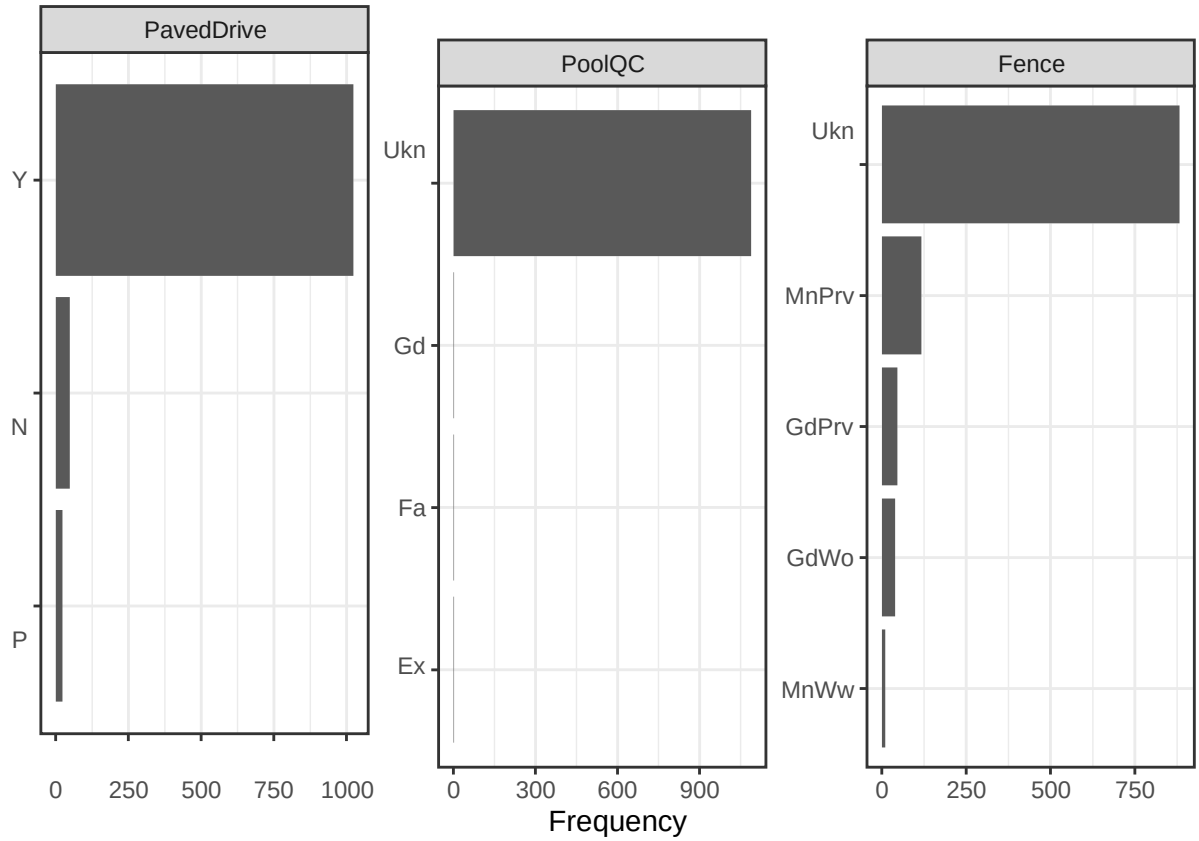

BsmtCond
TA
Gd
Fa
Po
0
250
500
750
1000
BsmtExposure
No
Av
Gd
Mn
0
200
400
600
BsmtFinType1
Unf
GLQ
ALQ
Rec
BLQ
LwQ
0
100
200
300
Frequency

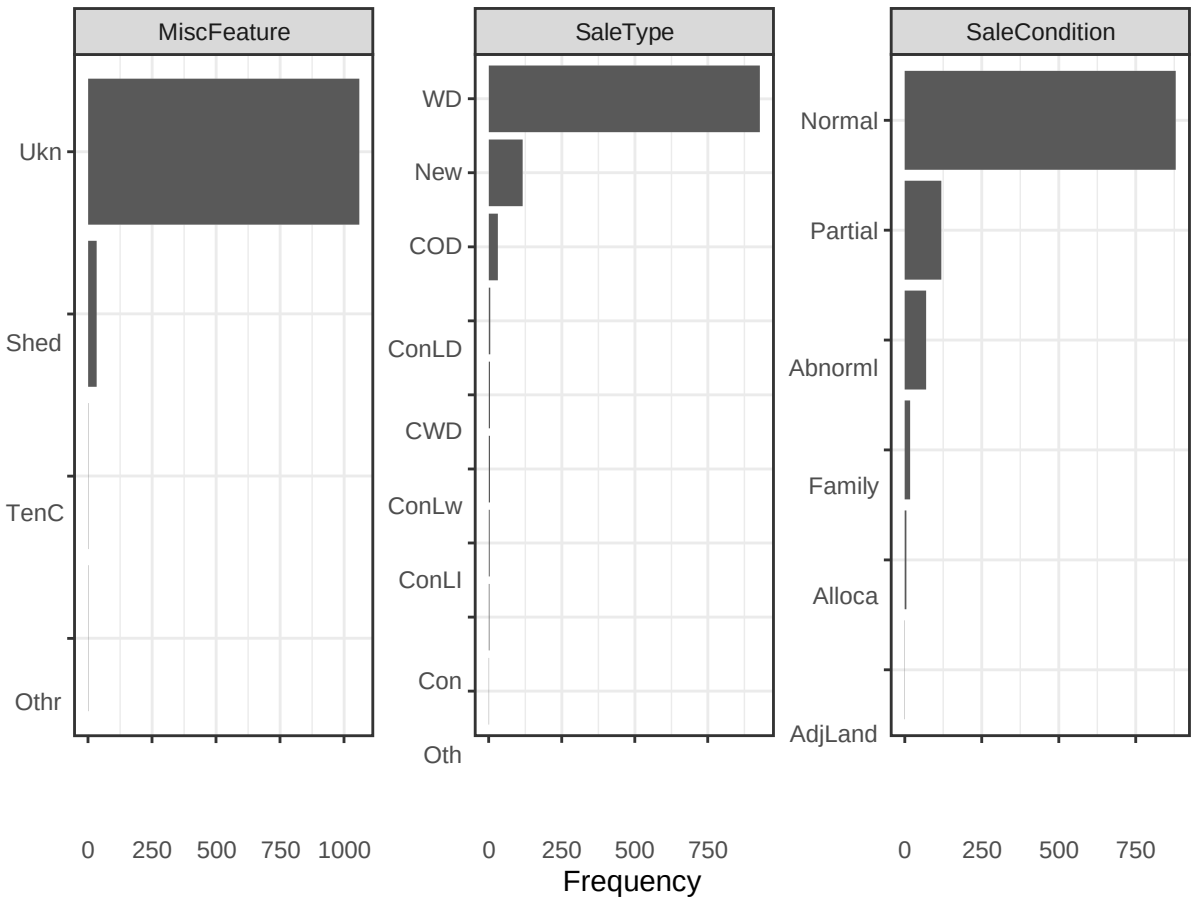

BsmtFinType2
Unf
Rec
LwQ
BLQ
ALQ
GLQ
0
250
500
750
1000
Heating
GasA
GasW
Grav
OthW
0
300
600
900
HeatingQC
Ex
TA
Gd
Fa
Po
0
200
400
600
Frequency


CentralAir
Y
N
0
250
500
750
1000
Electrical
SBrkr
FuseA
FuseF
FuseP
Mix
0
250
500
750
1000
KitchenQual
TA
Gd
Ex
Fa
0
200
400
Frequency


Functional
Typ
Min2
Min1
Mod
Maj1
Maj2
0
250
500
750
1000
FireplaceQu
Ukn
Gd
TA
Fa
Ex
Po
0
100
200
300
400
500
GarageType
Attchd
Detchd
BuiltIn
Basment
CarPort
2Types
0
200
400
600
Frequency

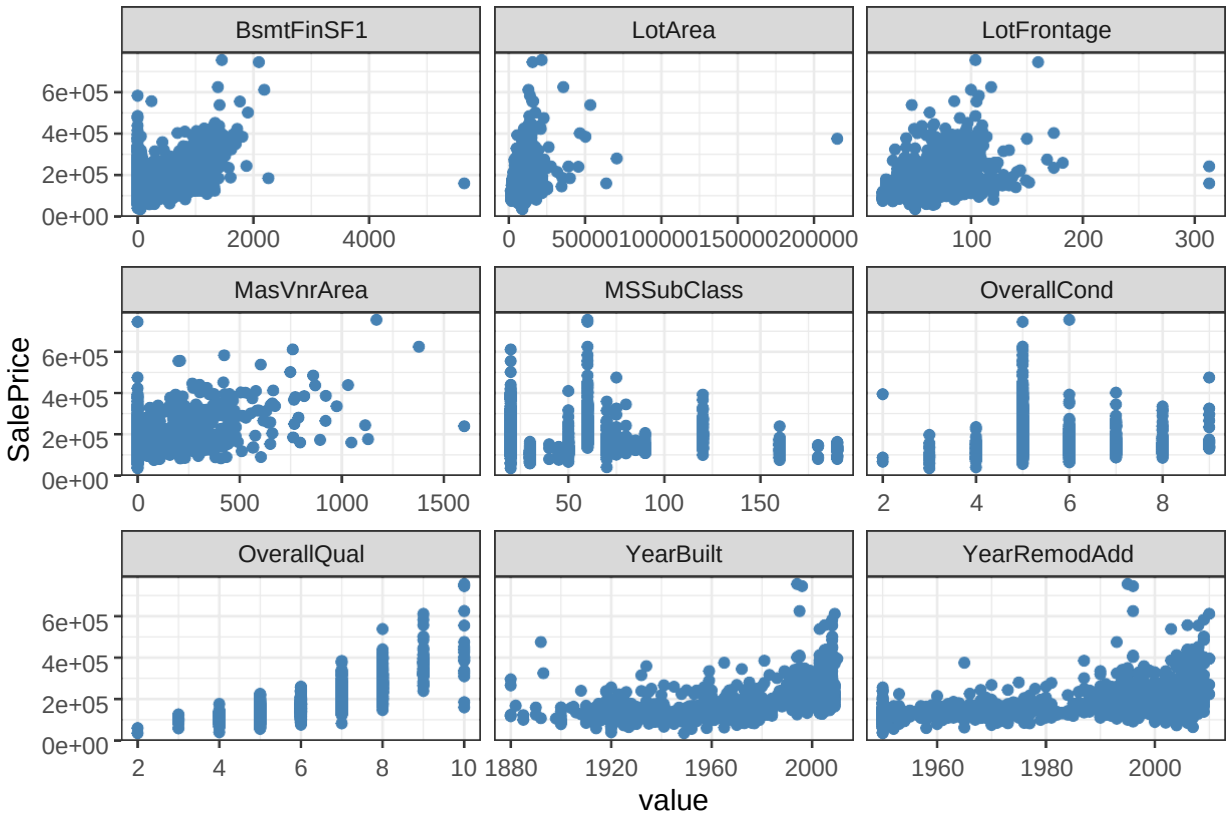

GarageFinish
GarageQual
GarageCond
Unf
RFn
Fin
TA
Fa
Gd
Po
Ex
0
100
200
300
400
500
250
750
1000
Frequency


PavedDrive
Y
N
P
0
250
500
750
1000
PoolQC
Ukn
Gd
Fa
Ex
0
300
600
900
Fence
Ukn
MnPrv
GdPrv
GdWo
MnWw
0
250
500
750
Frequency

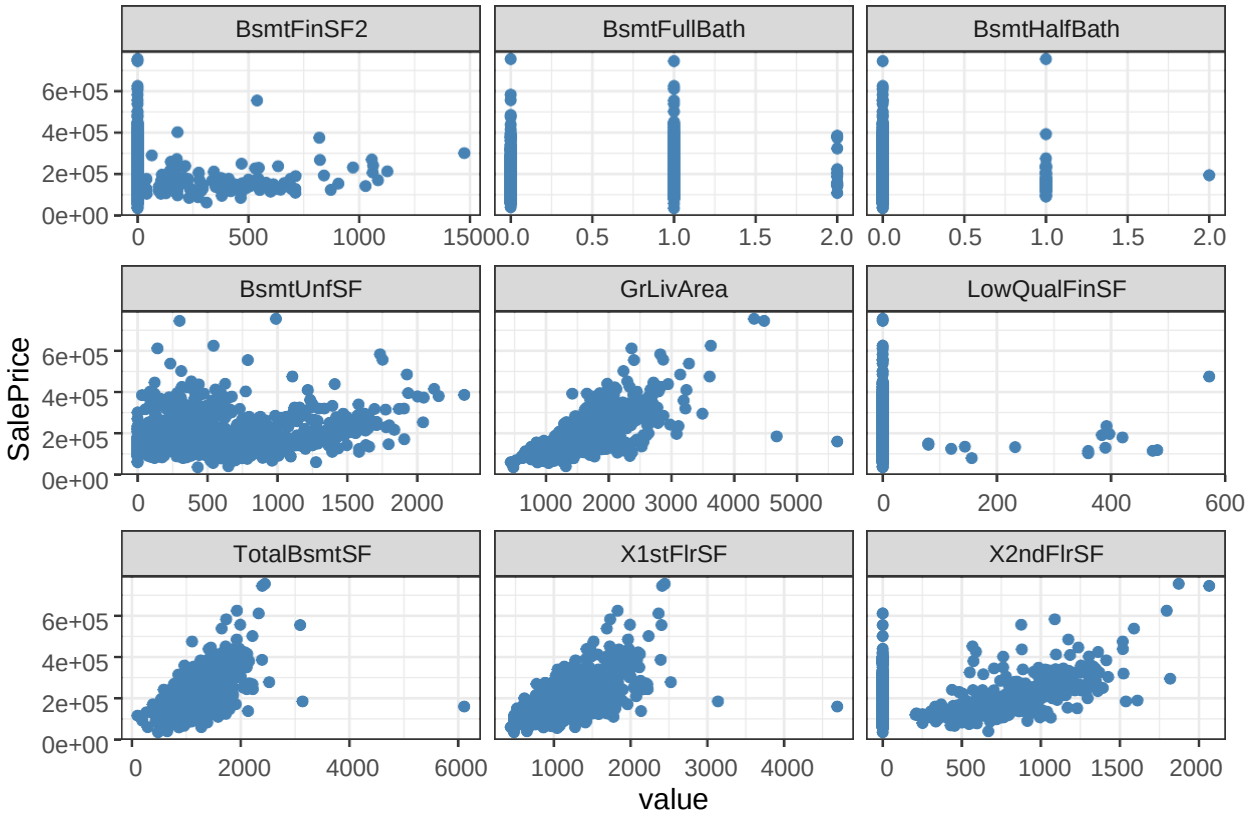


Memeriksa Korelasi Peubah

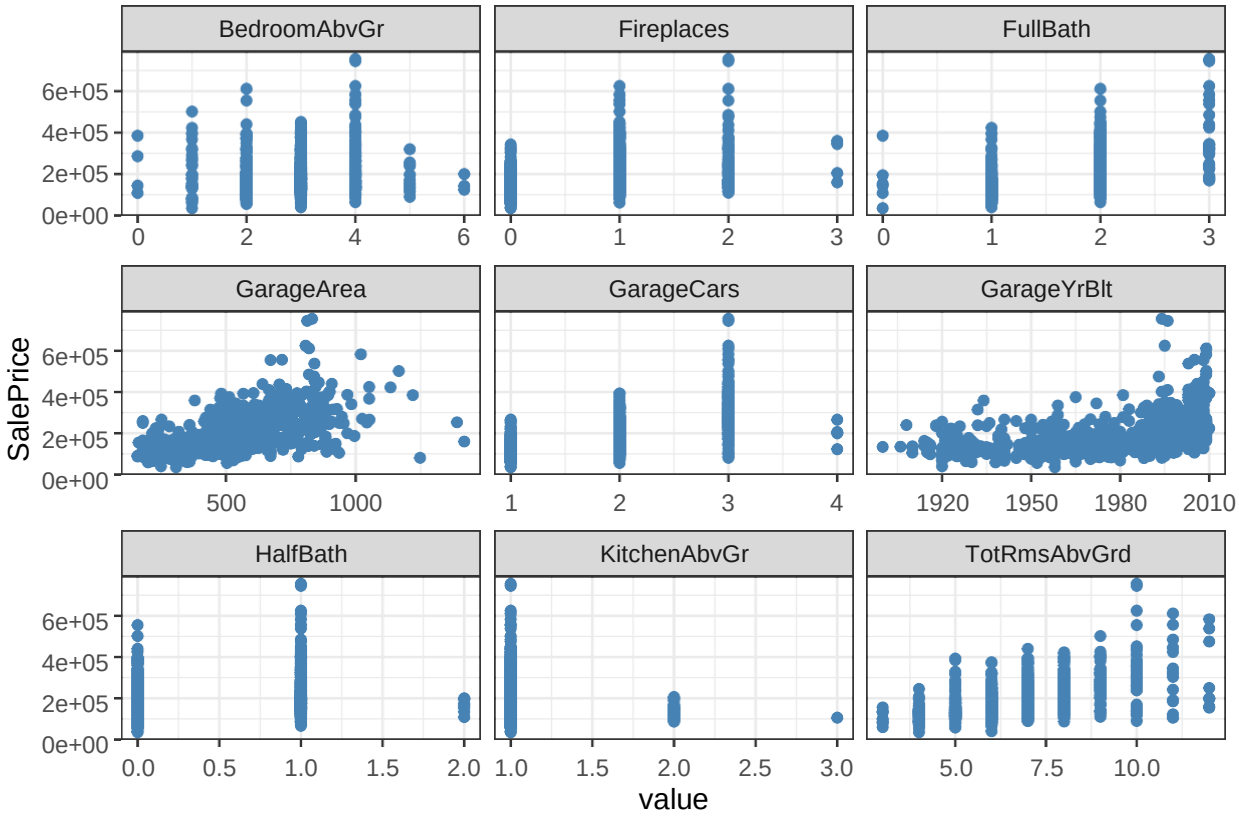
```
plot_scatterplot(data = data_house1 %>%
                   select_if(is.numeric),
                 by="SalePrice",geom_point_args = list(color="steelblue"),ggtheme = theme_bw() )
```

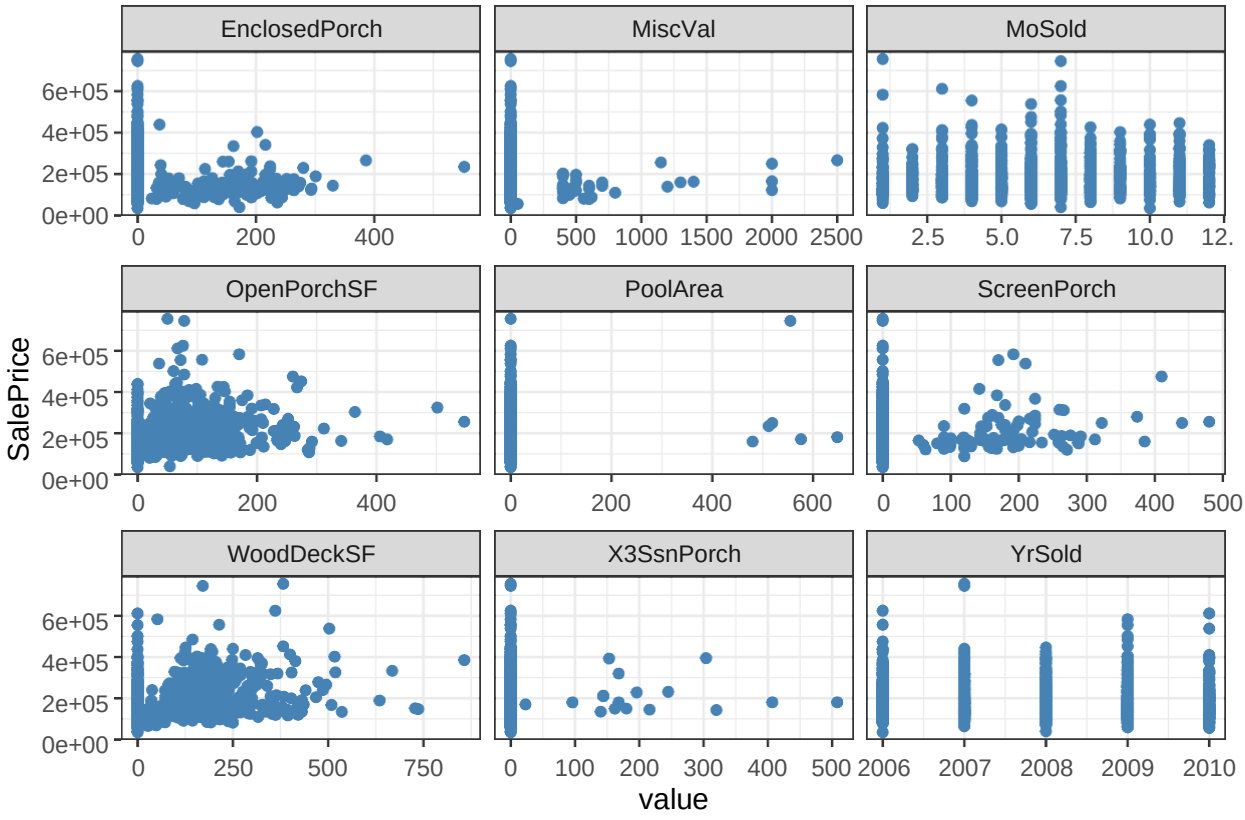

BsmtFinSF1
LotArea
LotFrontage
MasVnrArea
MSSubClass
OverallCond
OverallQual
YearBuilt
YearRemodAdd
SalePrice
value
6e+05
4e+05
2e+05
0e+00
0
2000
4000
50000
100000
150000
200000
100
200
300
500
1000
1500
50
150
2
4
6
8
10
1880
1920
1960
2000
1980

Page 1


BsmtFinSF2
BsmtFullBath
BsmtHalfBath
BsmtUnfSF
GrLivArea
LowQualFinSF
TotalBsmtSF
X1stFlrSF
X2ndFlrSF
SalePrice
value
0e+00
2e+05
4e+05
6e+05

Page 2


BedroomAbvGr
Fireplaces
FullBath
GarageArea
GarageCars
GarageYrBlt
HalfBath
KitchenAbvGr
TotRmsAbvGrd
SalePrice
value
0e+00
2e+05
4e+05
6e+05




```
cor_mat <- cor(data_house1%>%
                 select_if(is.numeric),method = "spearman")
cor_mat[upper.tri(cor_mat,diag = TRUE)] <- NA
cor_df <- cor_mat   %>%
    as.data.frame() %>%
    rownames_to_column(var = "Var1") %>%
  pivot_longer(names_to = "Var2",
               values_to = "corr",
               -Var1) %>% na.omit

cor_df %>% filter(abs(corr)>0.6) %>% arrange(desc(abs(corr)))
```

```
## # A tibble: 31 x 3
##    Var1         Var2          corr
##    <chr>        <chr>        <dbl>
##  1 GarageYrBlt  YearBuilt    0.895
##  2 X1stFlrSF    TotalBsmtSF  0.877
##  3 GarageArea   GarageCars   0.841
##  4 TotRmsAbvGrd GrLivArea    0.829
##  5 SalePrice    OverallQual  0.823
##  6 GarageYrBlt  YearRemodAdd 0.747
##  7 YearRemodAdd YearBuilt    0.738
##  8 SalePrice    GrLivArea    0.731
##  9 SalePrice    GarageCars   0.681
## 10 SalePrice    FullBath     0.671
## # ... with 21 more rows
```

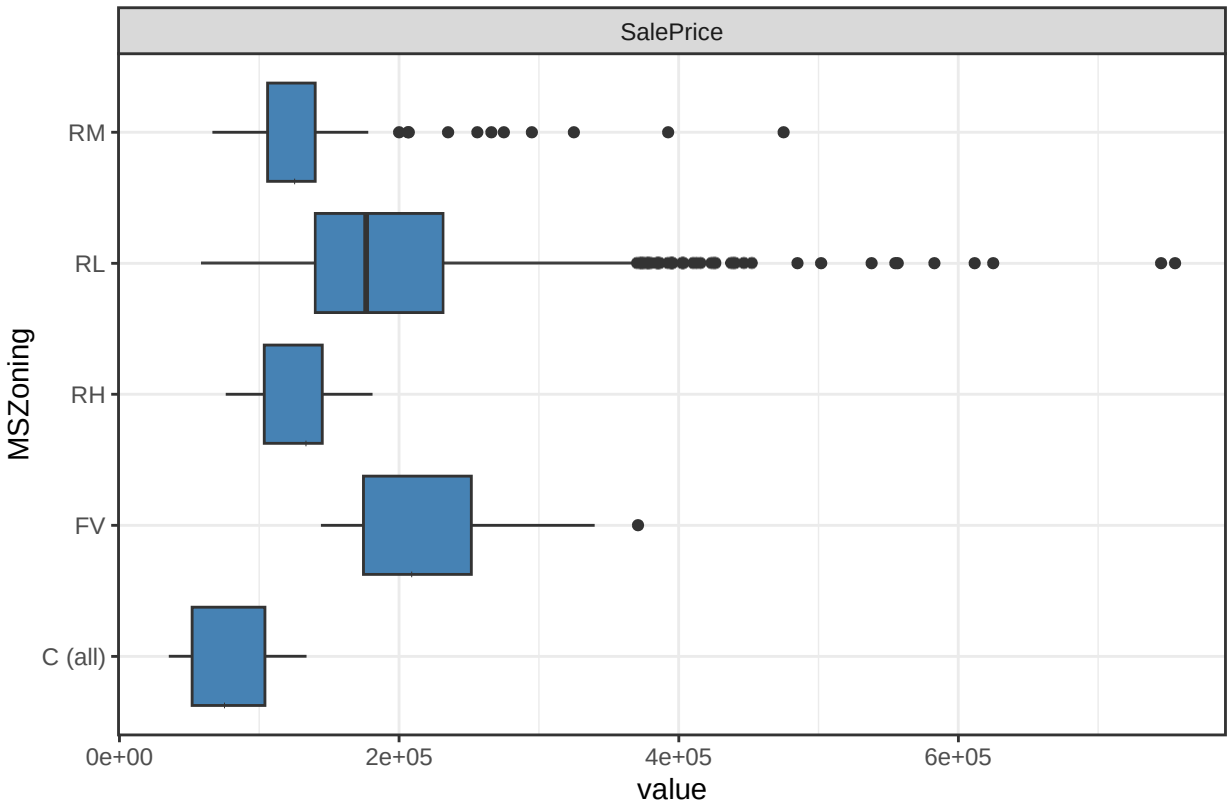
```
cor_df %>% filter(abs(corr)<=0.6)
```

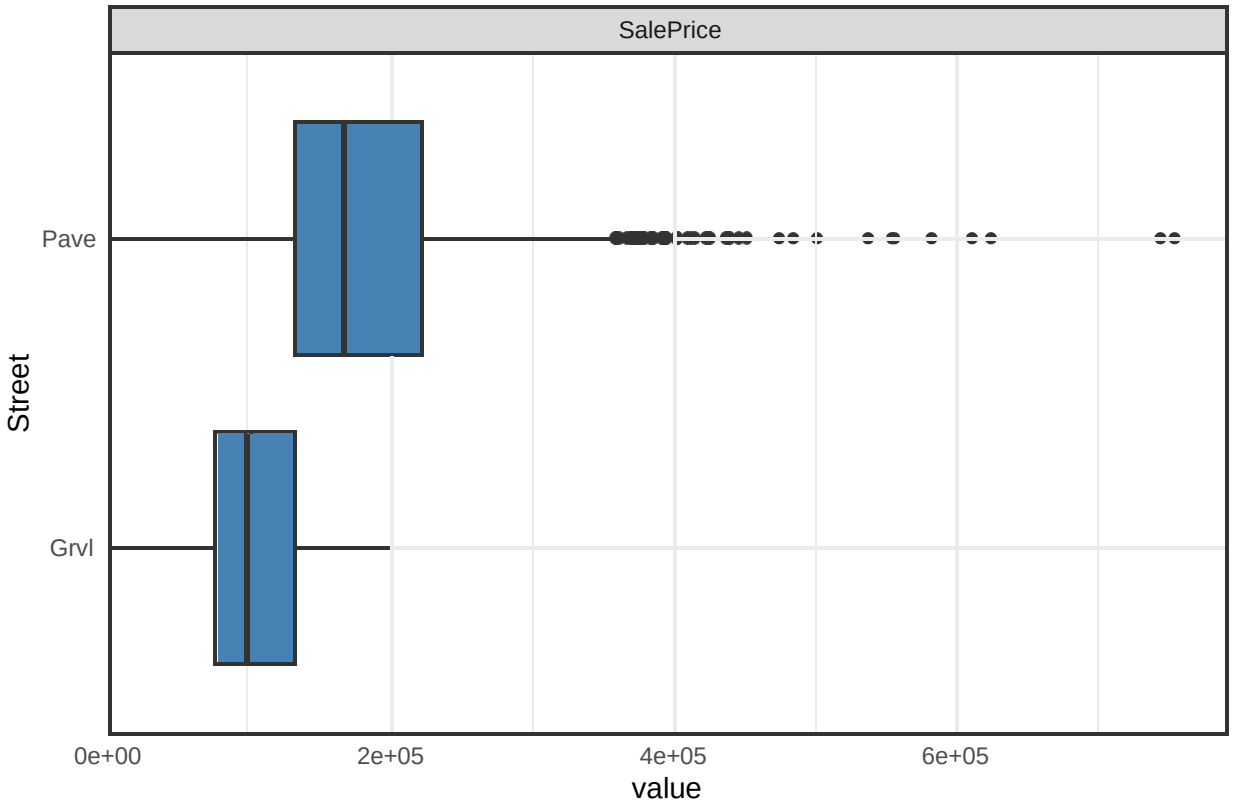
```
## # A tibble: 635 x 3
##    Var1        Var2            corr
##    <chr>       <chr>          <dbl>
##  1 LotFrontage MSSubClass  -0.313
##  2 LotArea     MSSubClass  -0.255
##  3 OverallQual MSSubClass   0.0992
##  4 OverallQual LotFrontage  0.238
##  5 OverallQual LotArea      0.283
##  6 OverallCond MSSubClass  -0.0763
##  7 OverallCond LotFrontage -0.0693
##  8 OverallCond LotArea     -0.0873
##  9 OverallCond OverallQual -0.264
## 10 YearBuilt   MSSubClass  -0.00468
## # ... with 625 more rows
```

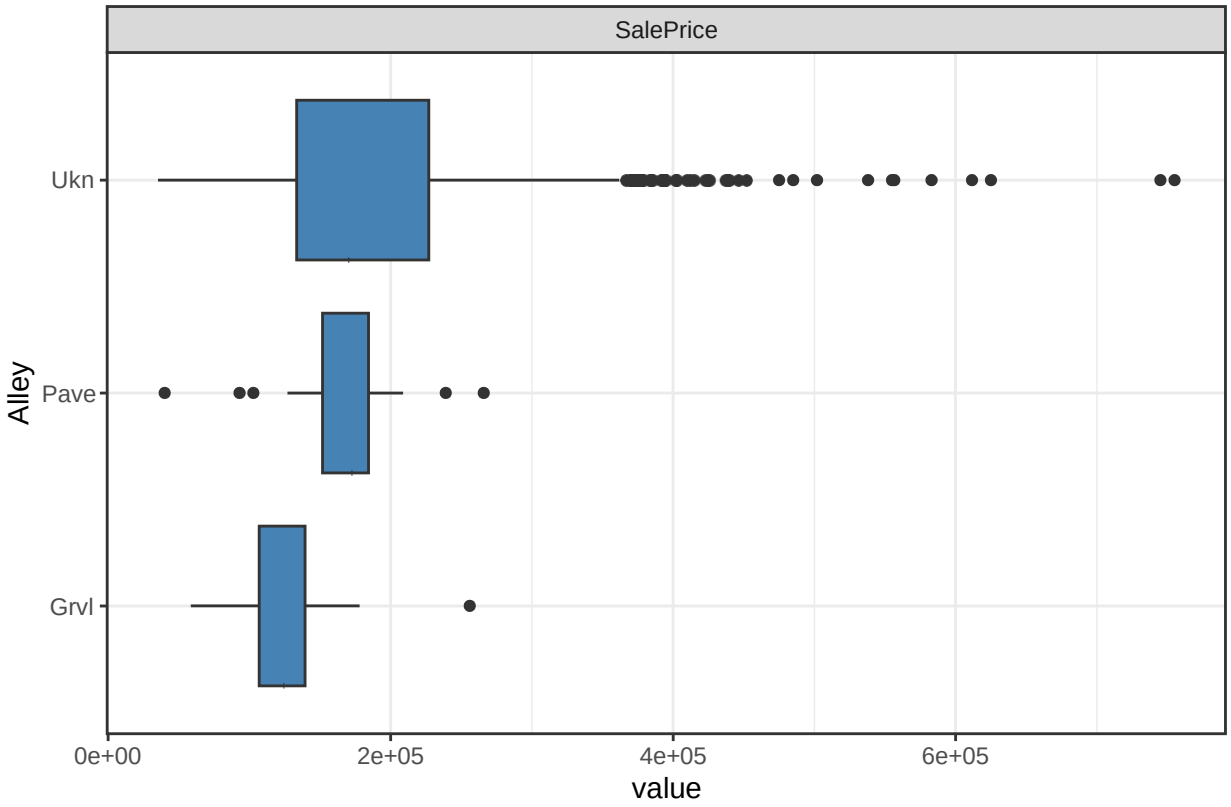
```
cat_var_names <- data_house1 %>%
  select(where(is.factor),SalePrice) %>%
  names
cat_var_names
```

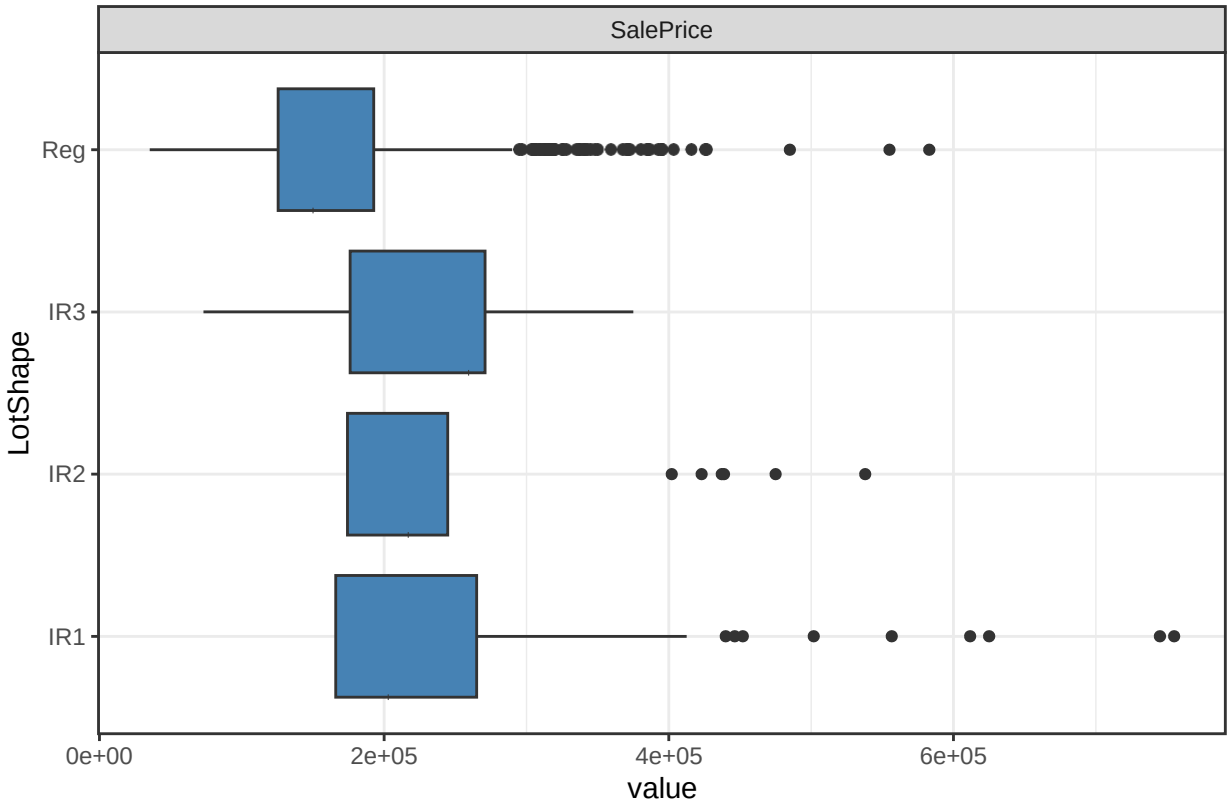
```
##  [1] "MSZoning"      "Street"        "Alley"         "LotShape"
##  [5] "LandContour"   "LotConfig"     "LandSlope"     "Neighborhood"
##  [9] "Condition1"    "Condition2"    "BldgType"      "HouseStyle"
## [13] "RoofStyle"     "RoofMatl"      "Exterior1st"   "Exterior2nd"
## [17] "MasVnrType"    "ExterQual"     "ExterCond"     "Foundation"
## [21] "BsmtQual"      "BsmtCond"      "BsmtExposure"  "BsmtFinType1"
## [25] "BsmtFinType2"  "Heating"       "HeatingQC"     "CentralAir"
## [29] "Electrical"    "KitchenQual"   "Functional"    "FireplaceQu"
## [33] "GarageType"    "GarageFinish"  "GarageQual"    "GarageCond"
## [37] "PavedDrive"    "PoolQC"        "Fence"         "MiscFeature"
## [41] "SaleType"      "SaleCondition" "SalePrice"
```

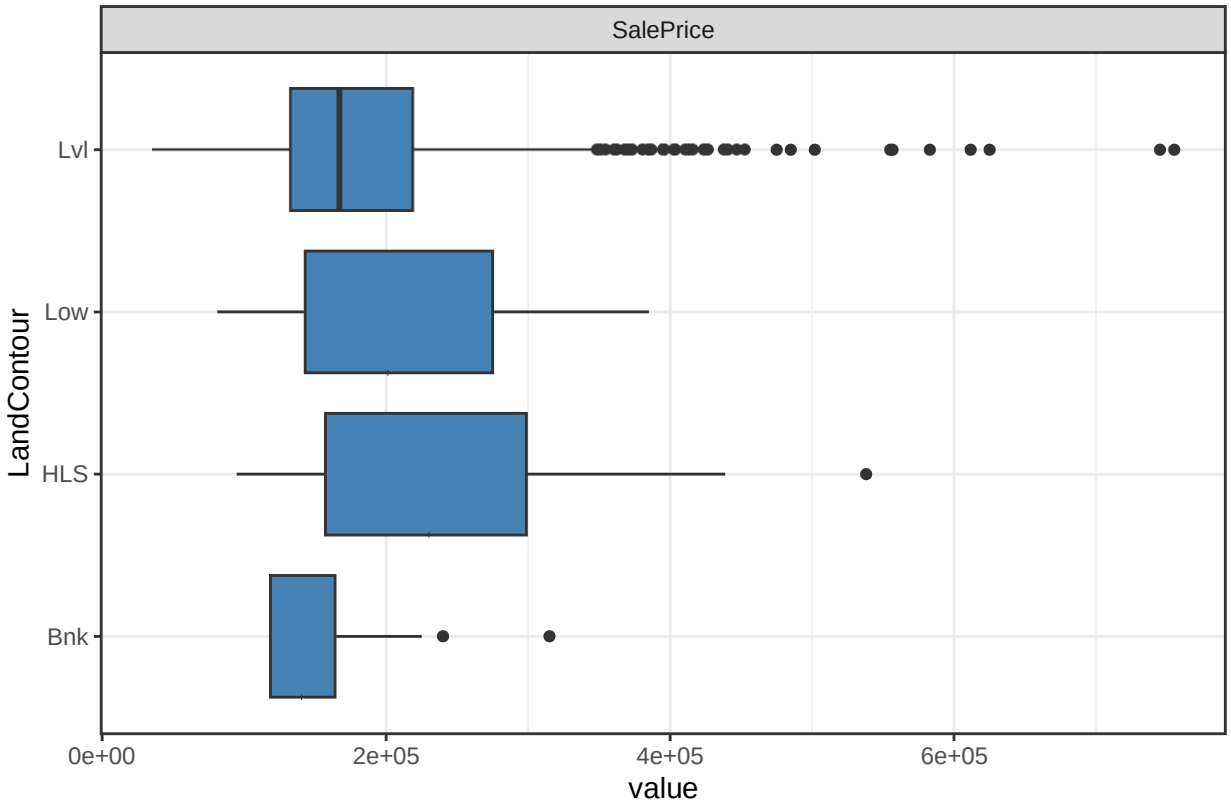
```
for(i in cat_var_names[-43]){
plot_boxplot(data = data_house1 %>%
               select(where(is.factor),SalePrice),
             geom_boxplot_args=list(fill="steelblue"),
             by=i,ggtheme = theme_bw())
}
```

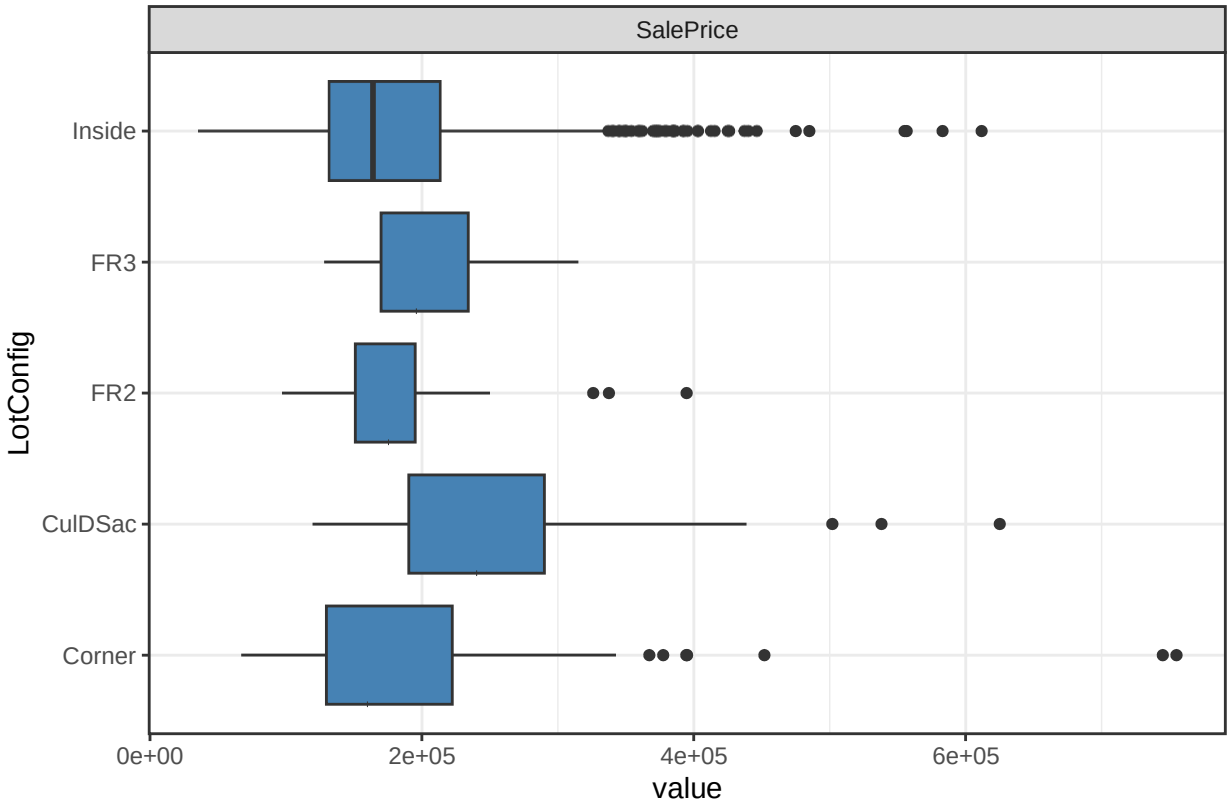

SalePrice
MSZoning
RM
RL
RH
FV
C (all)
0e+00
2e+05
4e+05
6e+05
value

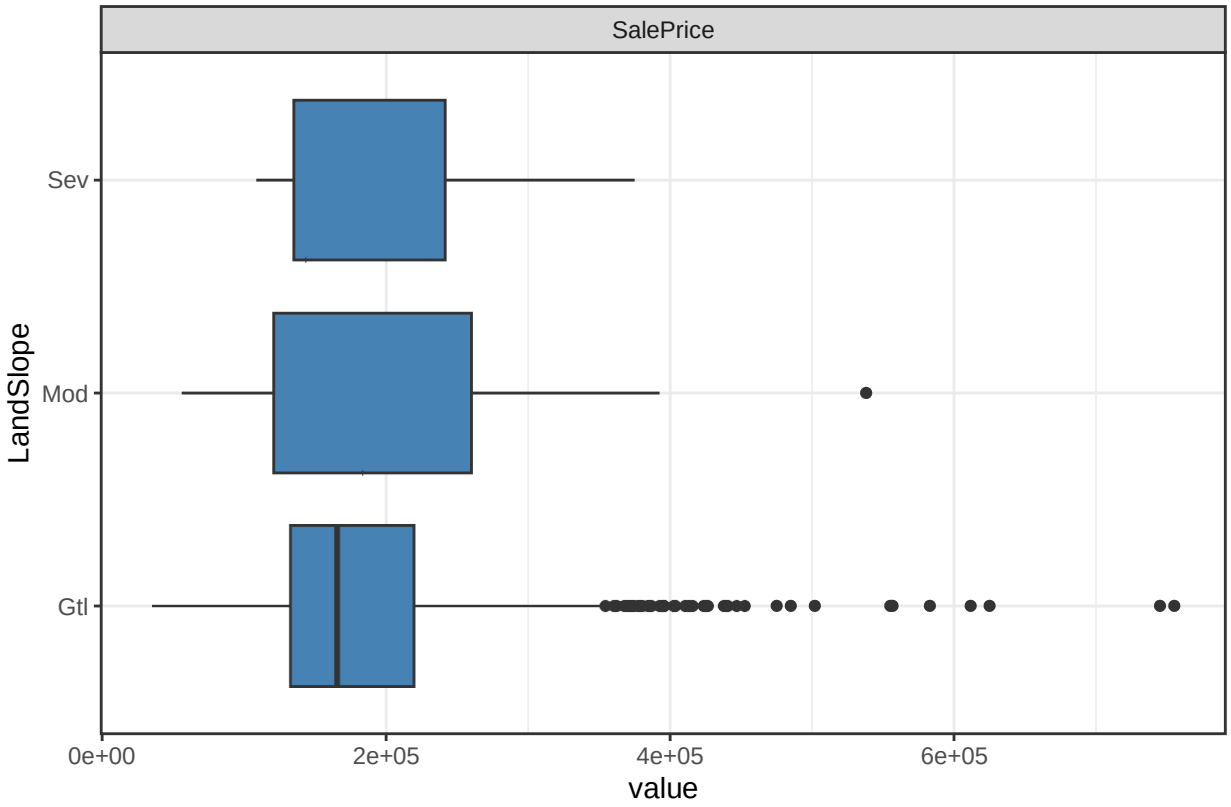

SalePrice
Pave
Grvl
Street
0e+00
2e+05
4e+05
6e+05
value

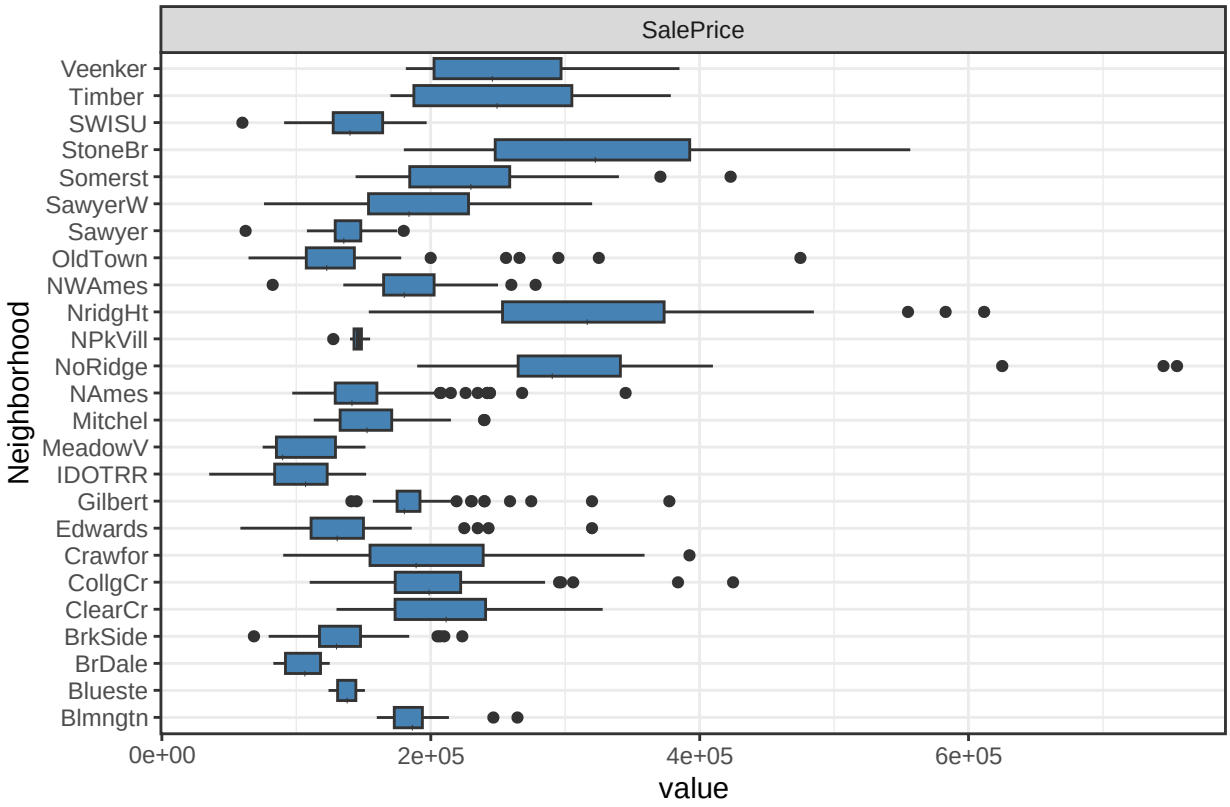

SalePrice
Alley
Ukn
Pave
Grvl
0e+00
2e+05
4e+05
6e+05
value

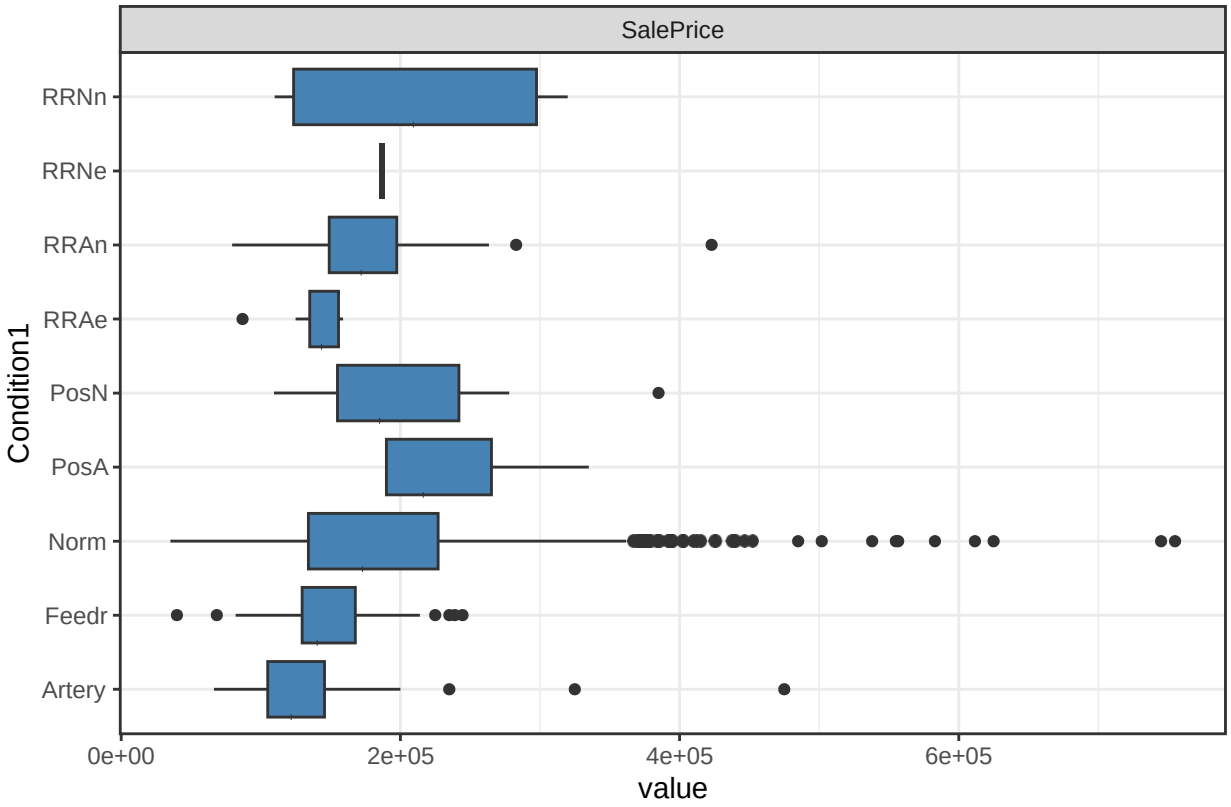

SalePrice
Reg
IR3
IR2
IR1
LotShape
0e+00
2e+05
4e+05
6e+05
value

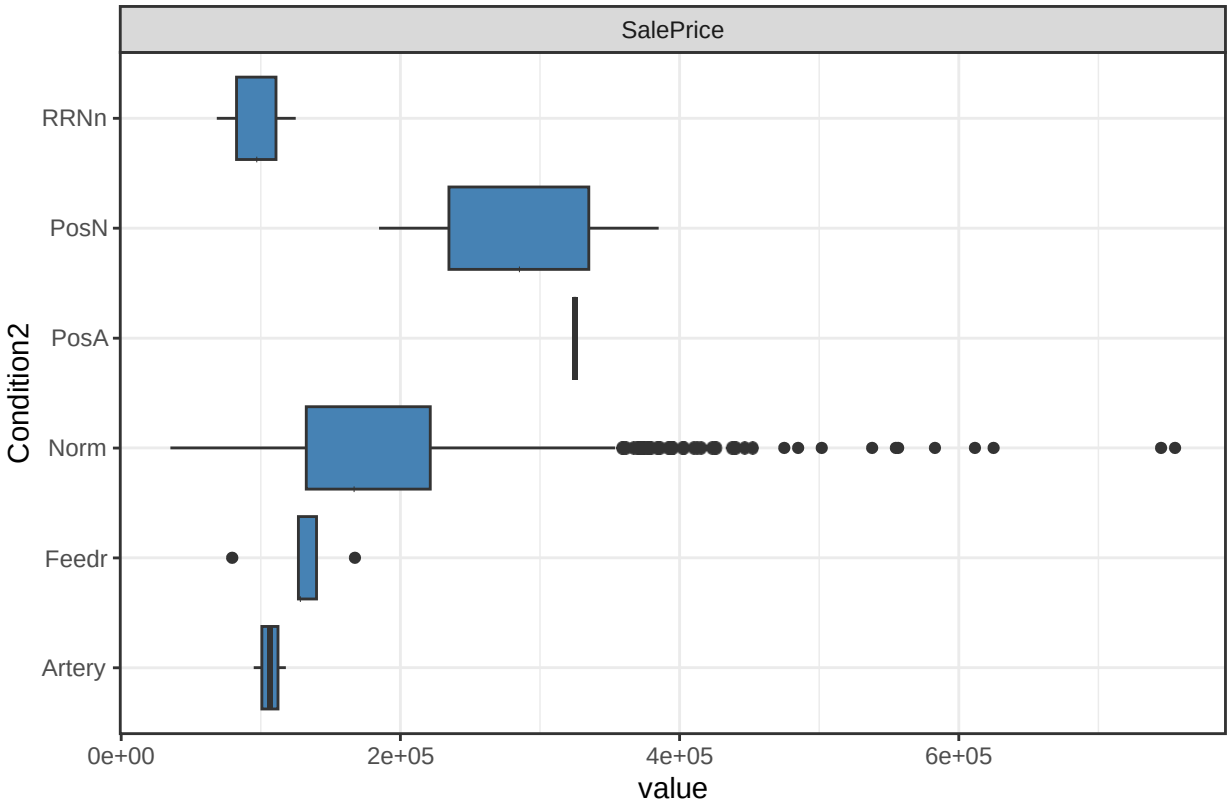

SalePrice
LandContour
Lvl
Low
HLS
Bnk
0e+00
2e+05
4e+05
6e+05
value

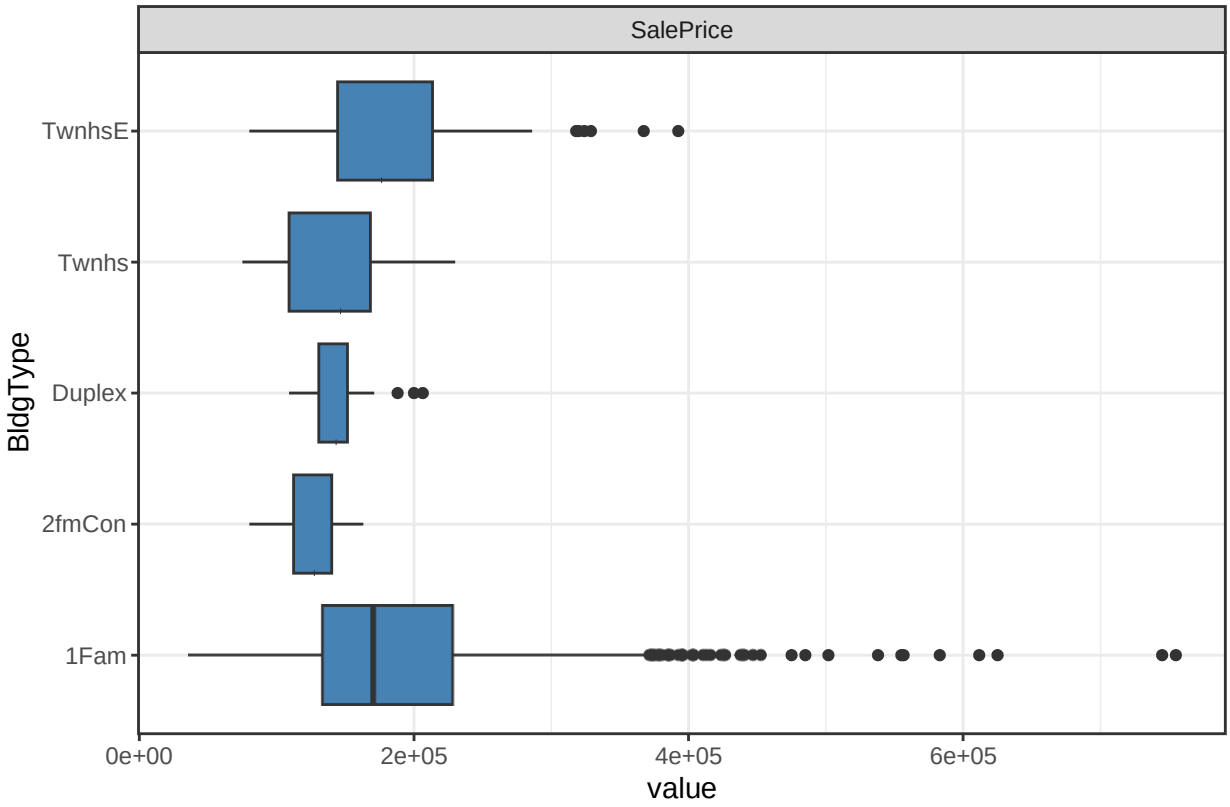

SalePrice
Inside
FR3
FR2
CulDSac
Corner
LotConfig
0e+00
2e+05
4e+05
6e+05
value

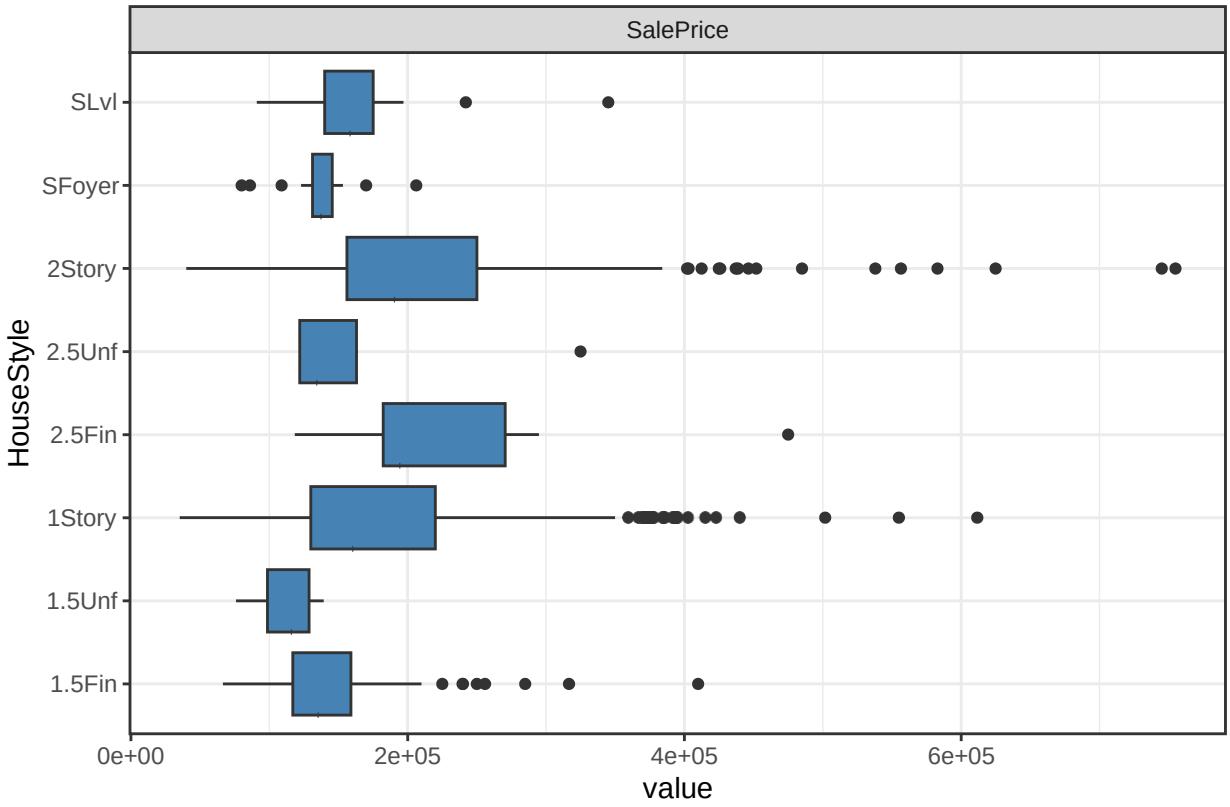

SalePrice
LandSlope
Sev
Mod
Gtl
0e+00
2e+05
4e+05
6e+05
value

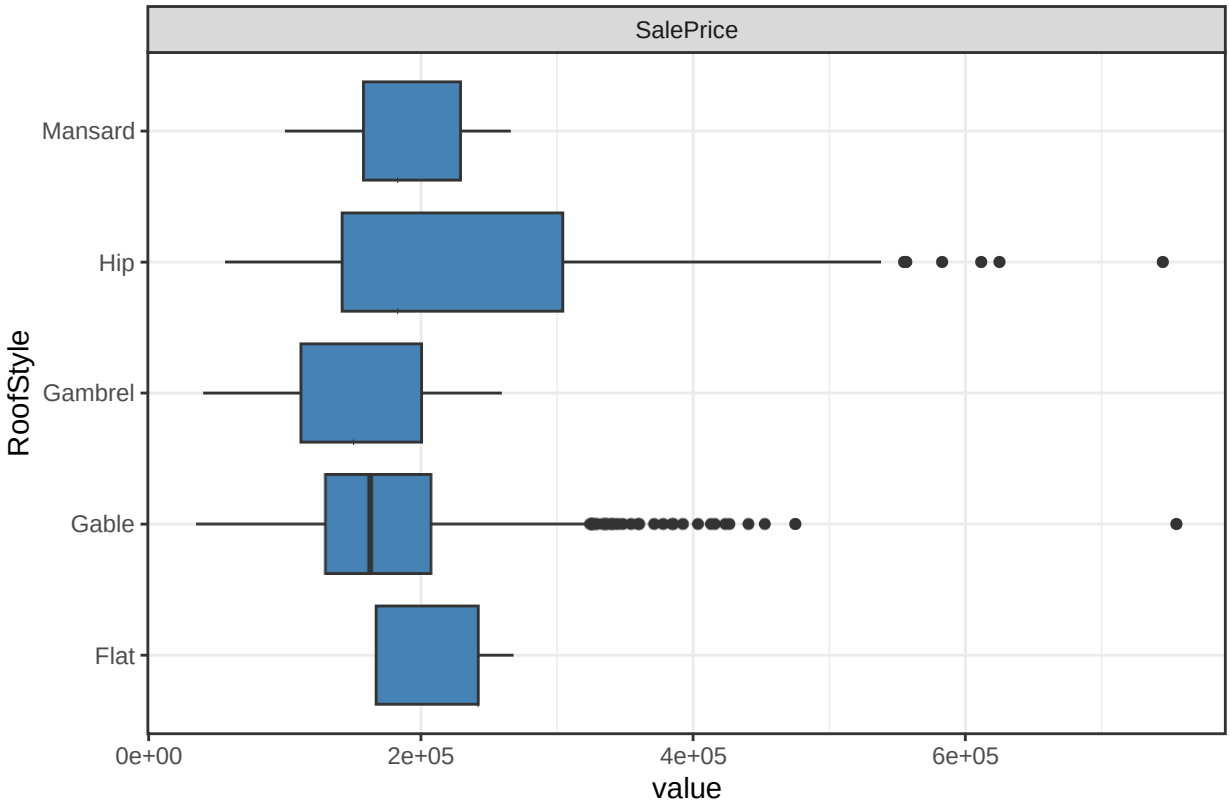

SalePrice
Neighborhood
Veenker
Timber
SWISU
StoneBr
Somerst
SawyerW
Sawyer
OldTown
NWAmes
NridgHt
NPkVill
NoRidge
NAmes
Mitchel
MeadowV
IDOTRR
Gilbert
Edwards
Crawfor
CollgCr
ClearCr
BrkSide
BrDale
Blueste
Blmngtn
0e+00
2e+05
4e+05
6e+05
value

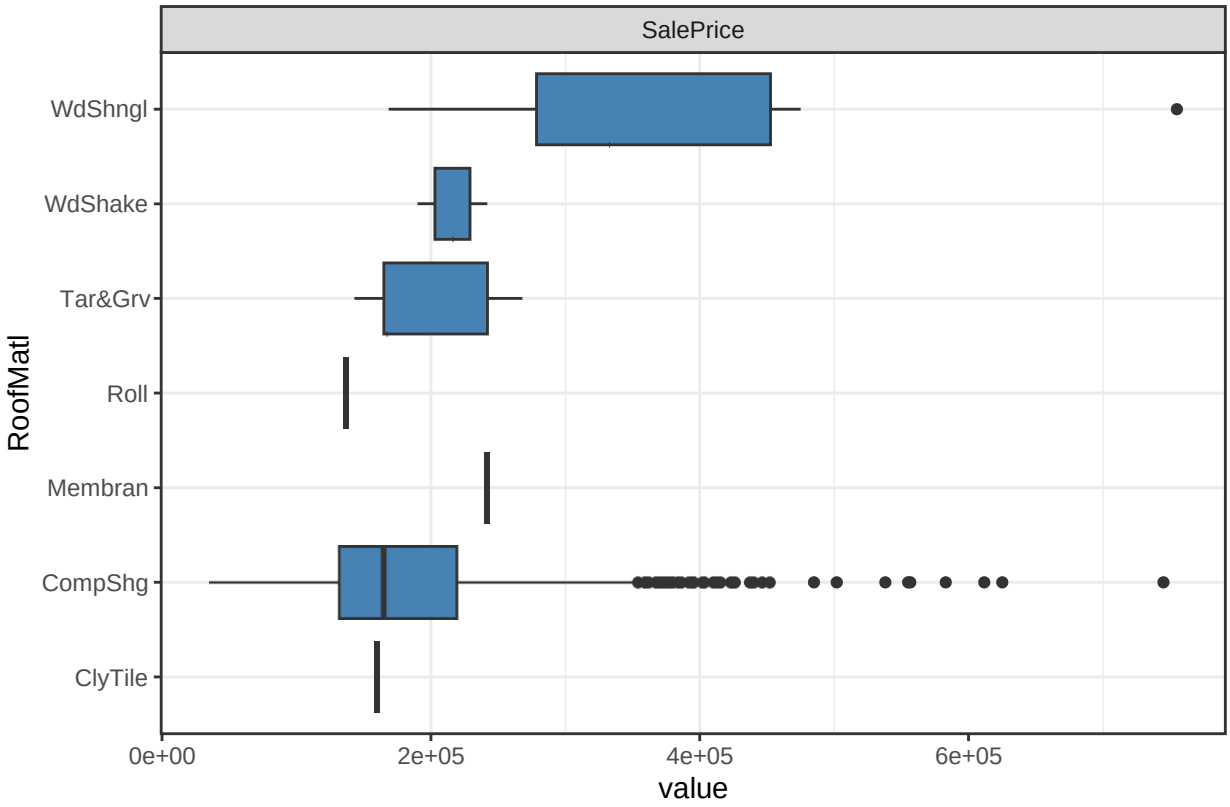

SalePrice
Condition1
RRNn
RRNe
RRAn
RRAe
PosN
PosA
Norm
Feedr
Artery
0e+00
2e+05
4e+05
6e+05
value

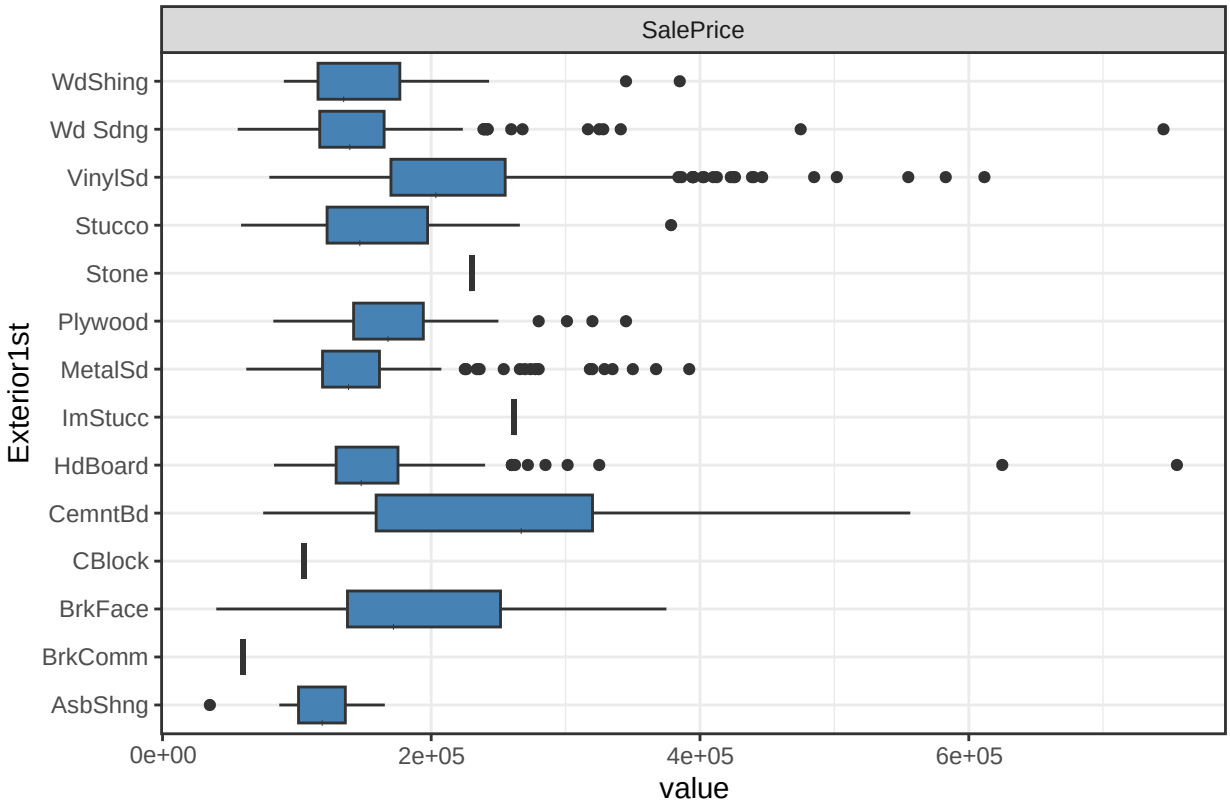

SalePrice
Condition2
RRNn
PosN
PosA
Norm
Feedr
Artery
0e+00
2e+05
4e+05
6e+05
value

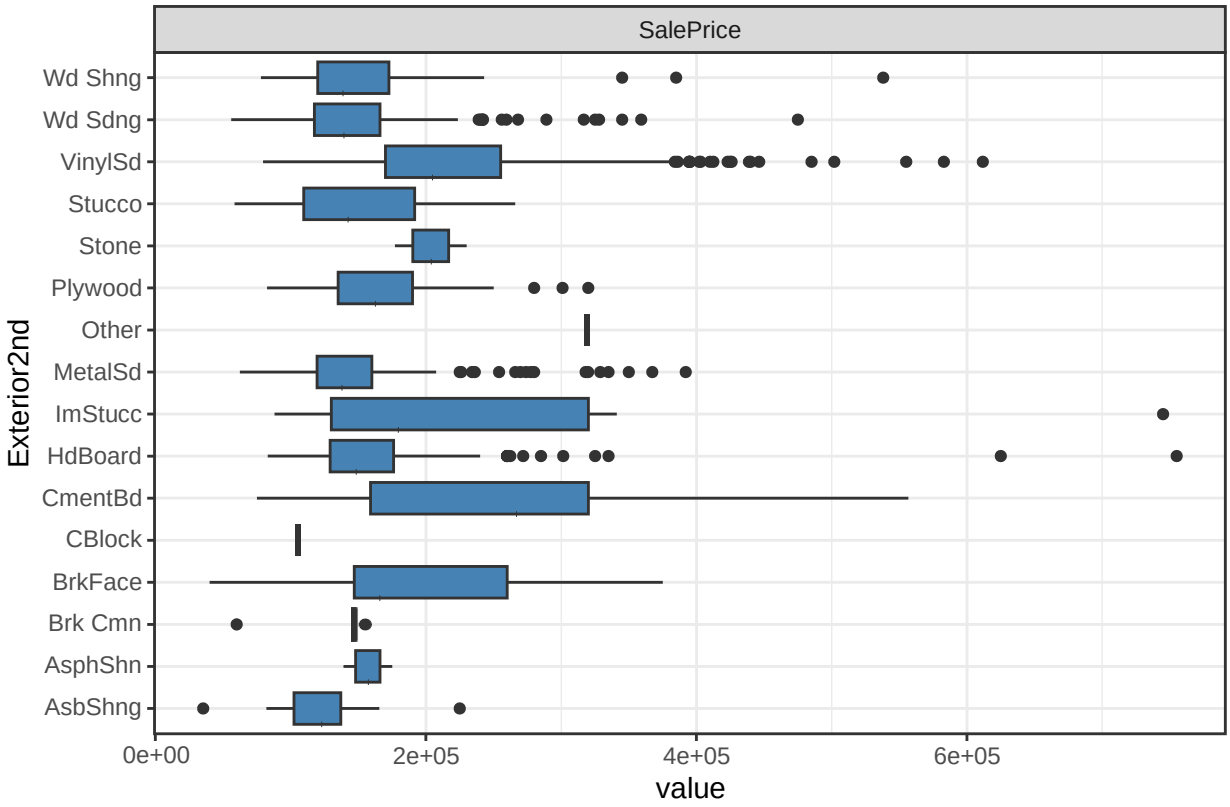

SalePrice
TwnhsE
Twnhs
Duplex
2fmCon
1Fam
BldgType
0e+00
2e+05
4e+05
6e+05
value

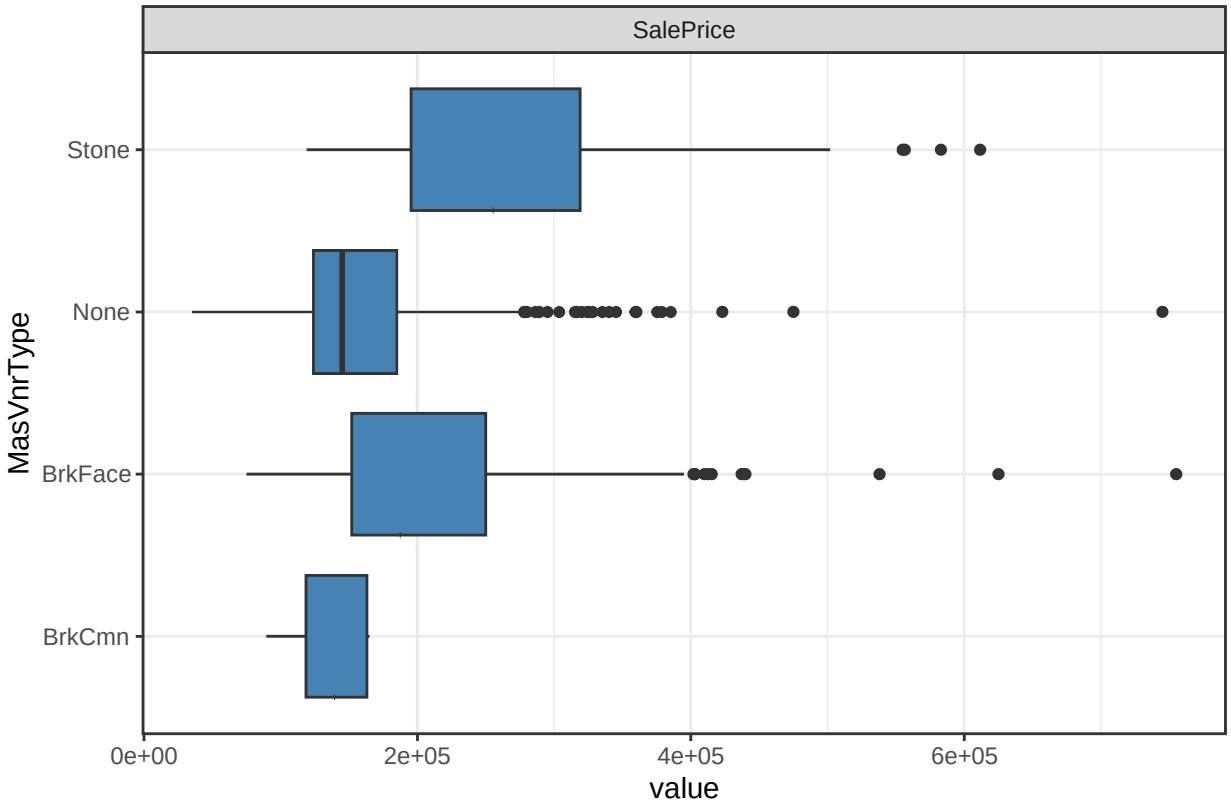

SalePrice
HouseStyle
SLvl
SFoyer
2Story
2.5Unf
2.5Fin
1Story
1.5Unf
1.5Fin
0e+00
2e+05
4e+05
6e+05
value

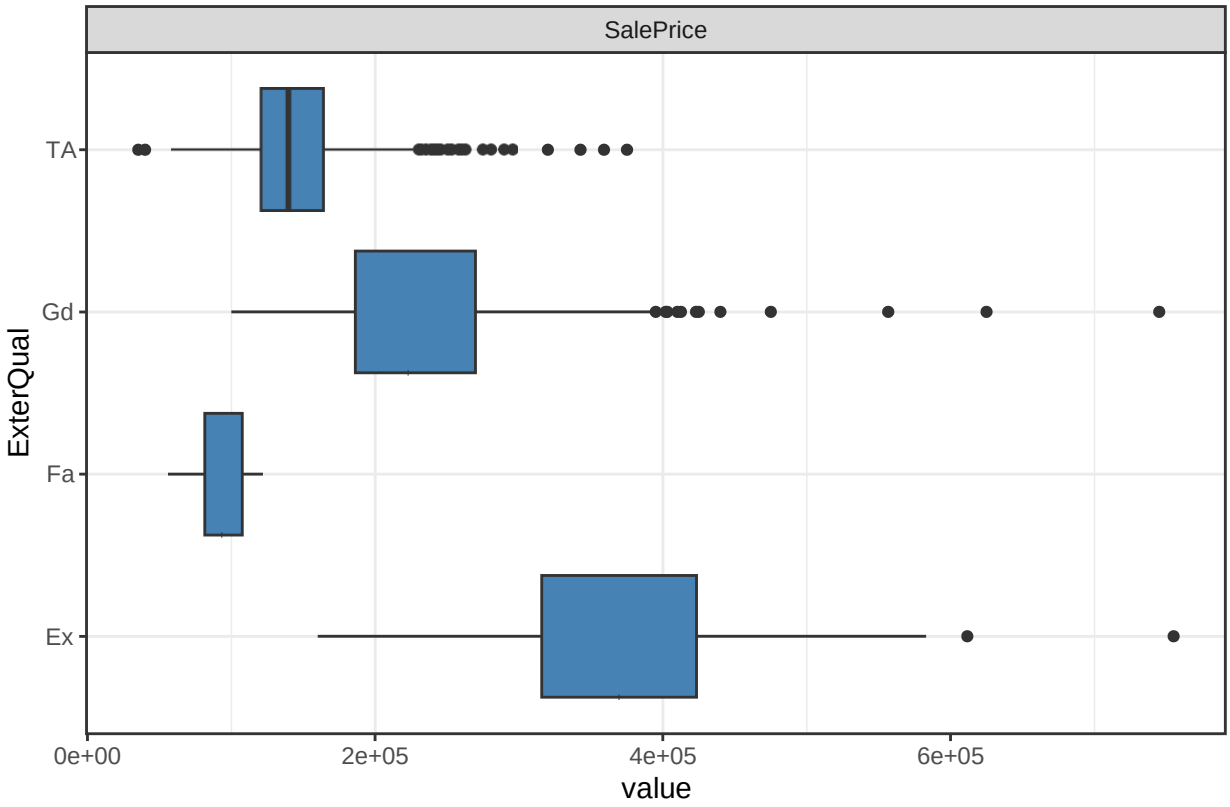

SalePrice
Mansard
Hip
Gambrel
Gable
Flat
RoofStyle
0e+00
2e+05
4e+05
6e+05
value

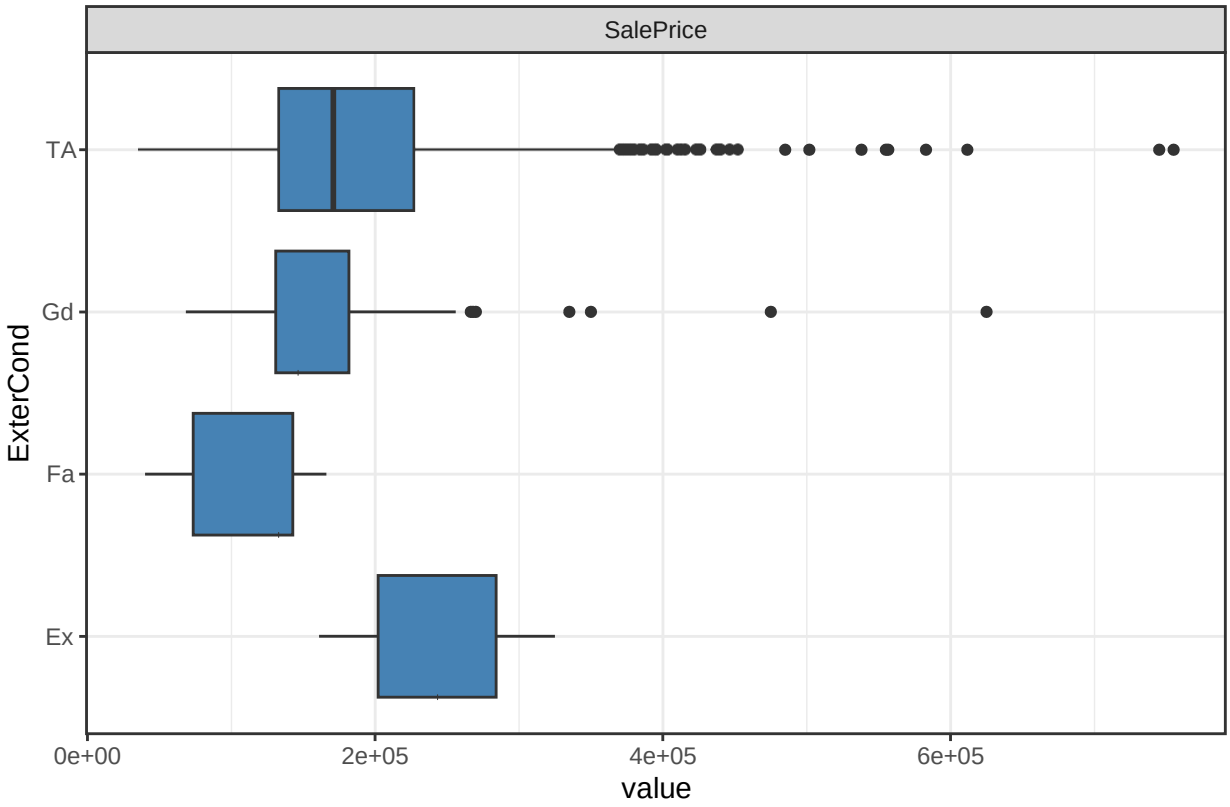

SalePrice
RoofMatl
WdShngl
WdShake
Tar&Grv
Roll
Membran
CompShg
ClyTile
0e+00
2e+05
4e+05
6e+05
value

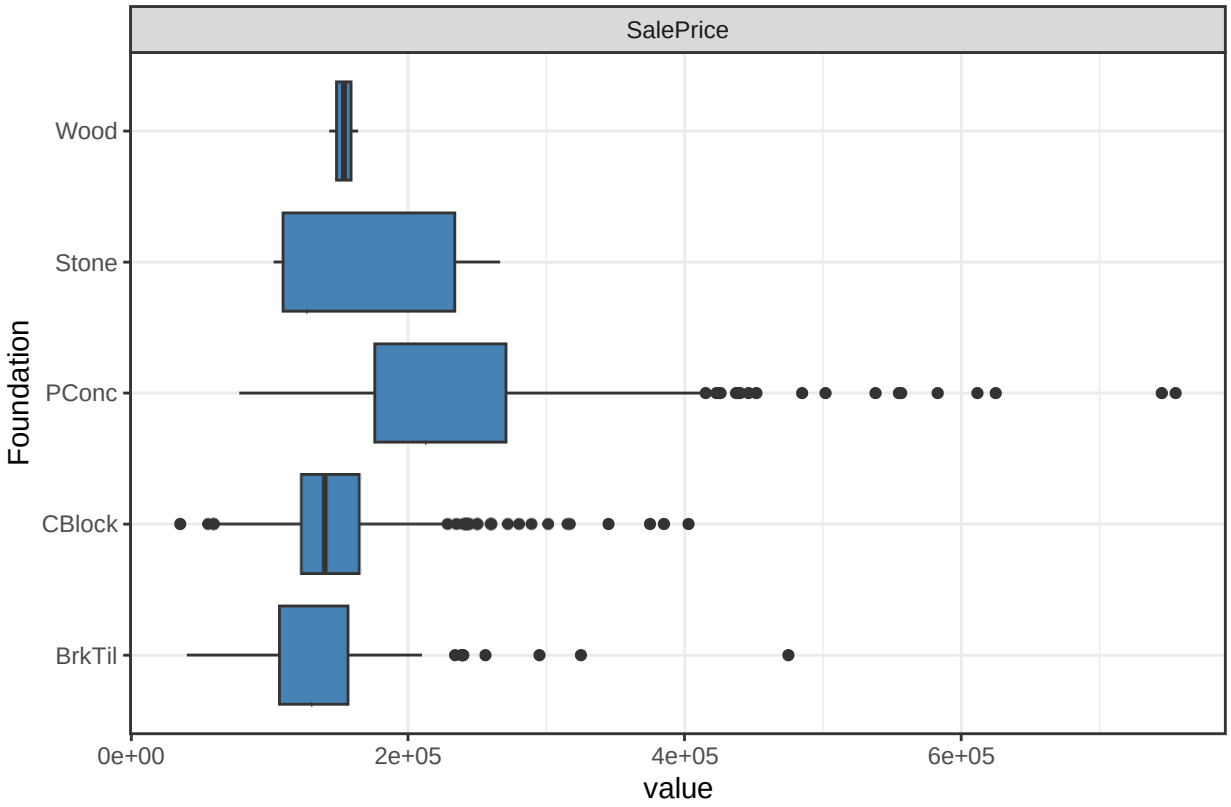

SalePrice
Exterior1st
WdShing
Wd Sdng
VinylSd
Stucco
Stone
Plywood
MetalSd
ImStucc
HdBoard
CemntBd
CBlock
BrkFace
BrkComm
AsbShng
0e+00
2e+05
4e+05
6e+05
value

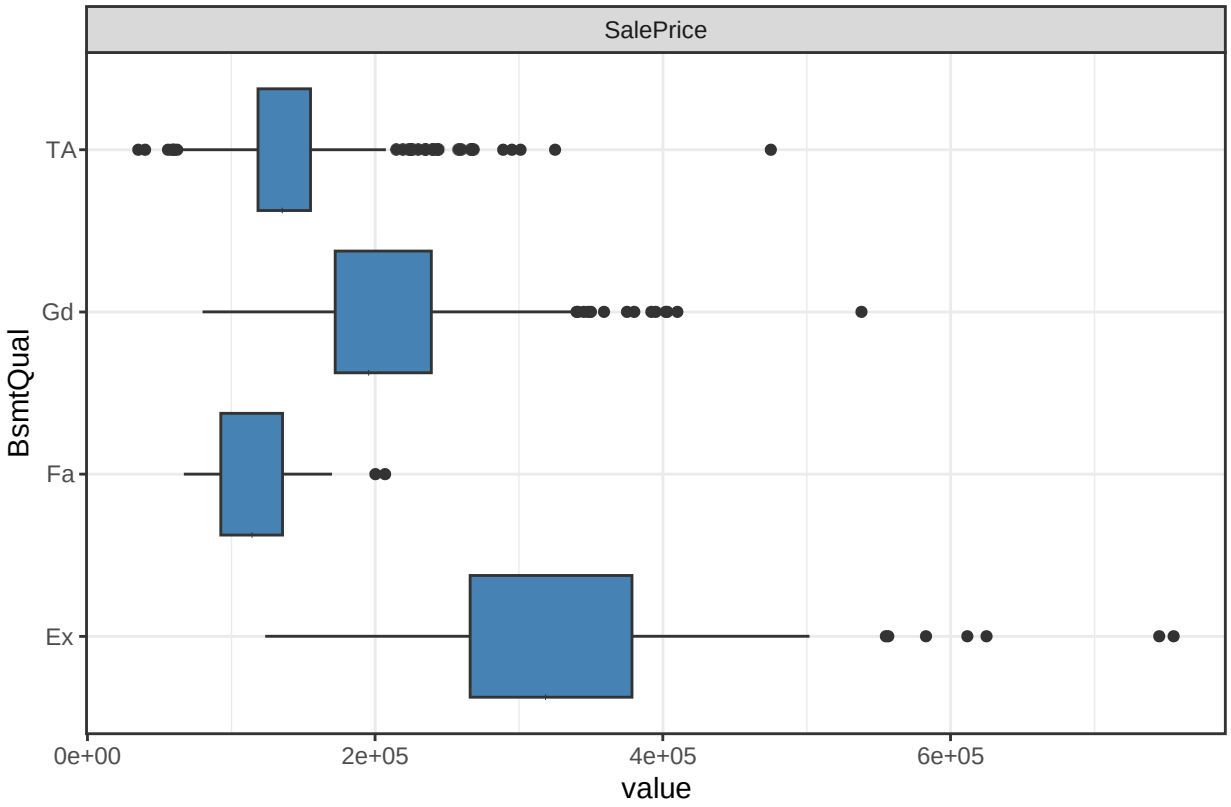

SalePrice
Exterior2nd
Wd Shng
Wd Sdng
VinylSd
Stucco
Stone
Plywood
Other
MetalSd
ImStucc
HdBoard
CmentBd
CBlock
BrkFace
Brk Cmn
AsphShn
AsbShng
0e+00
2e+05
4e+05
6e+05
value

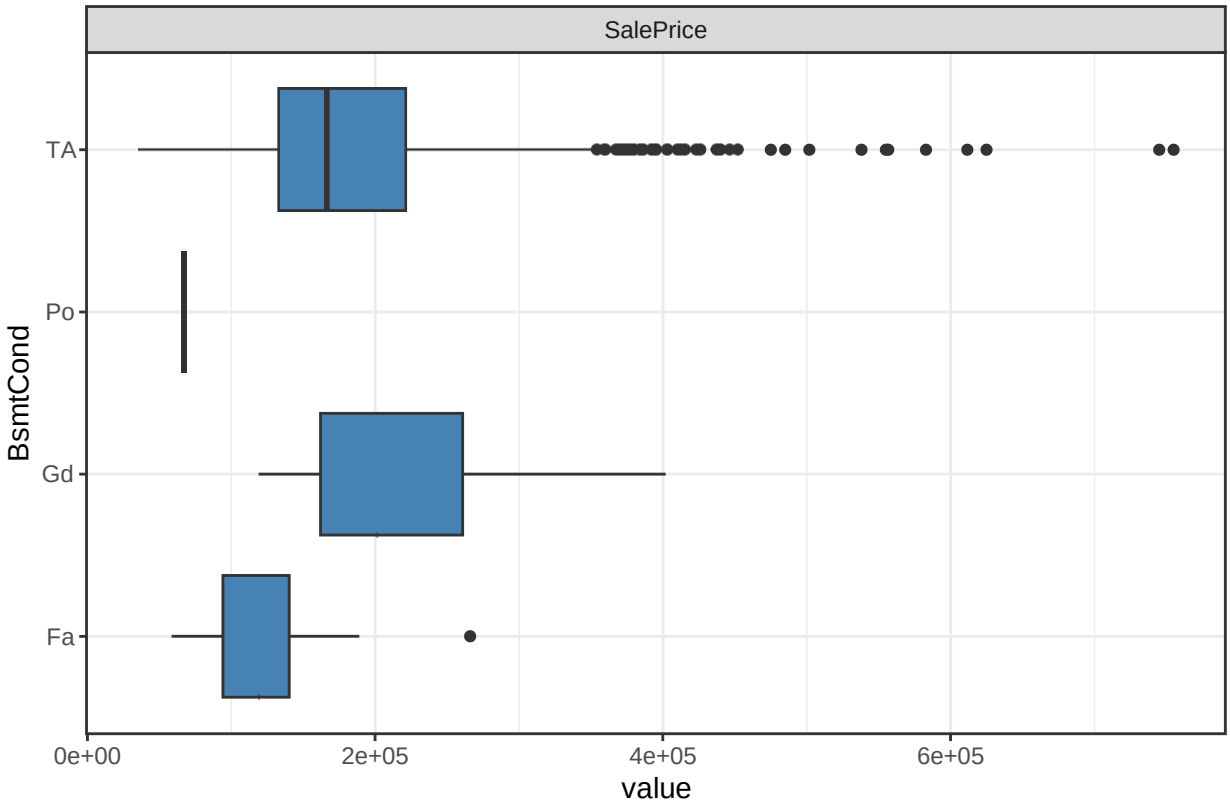

SalePrice
Stone
None
BrkFace
BrkCmn
MasVnrType
0e+00
2e+05
4e+05
6e+05
value

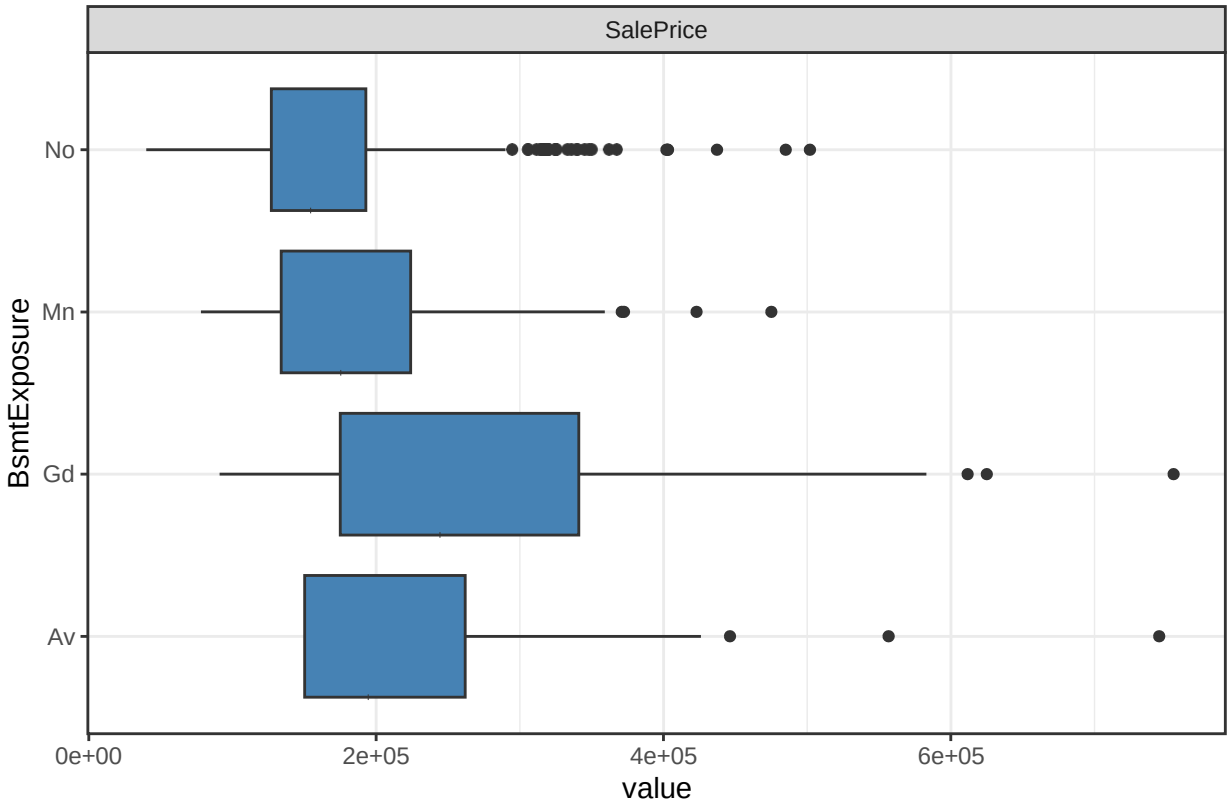

SalePrice
TA
Gd
Fa
Ex
ExterQual
0e+00
2e+05
4e+05
6e+05
value

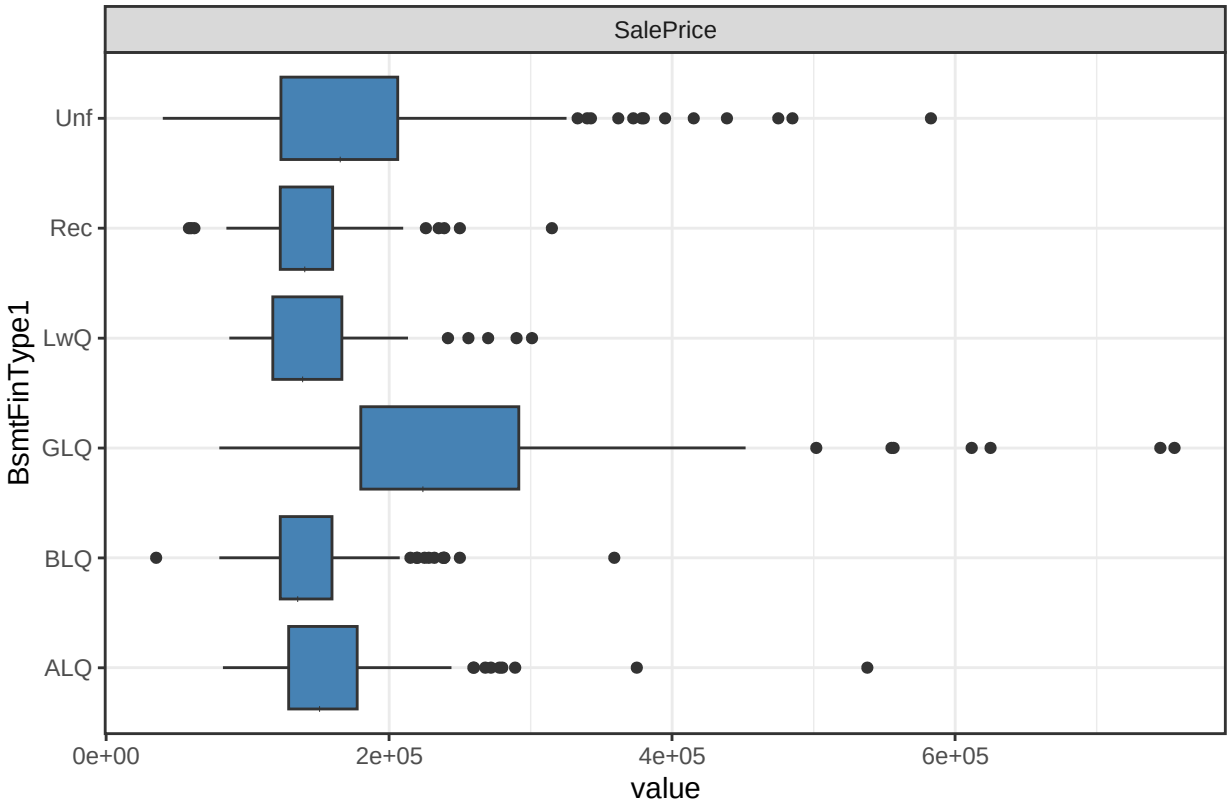

SalePrice
TA
Gd
Fa
Ex
ExterCond
0e+00
2e+05
4e+05
6e+05
value

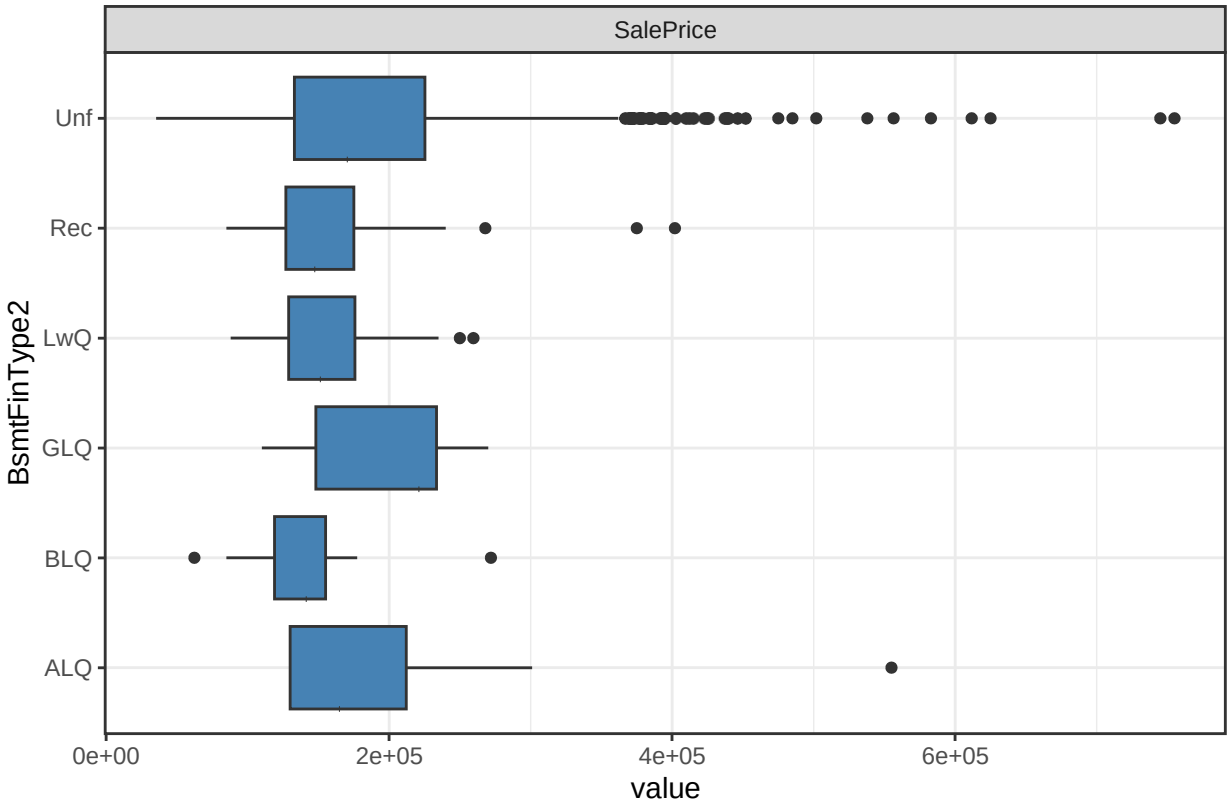

SalePrice
Foundation
Wood
Stone
PConc
CBlock
BrkTil
0e+00
2e+05
4e+05
6e+05
value

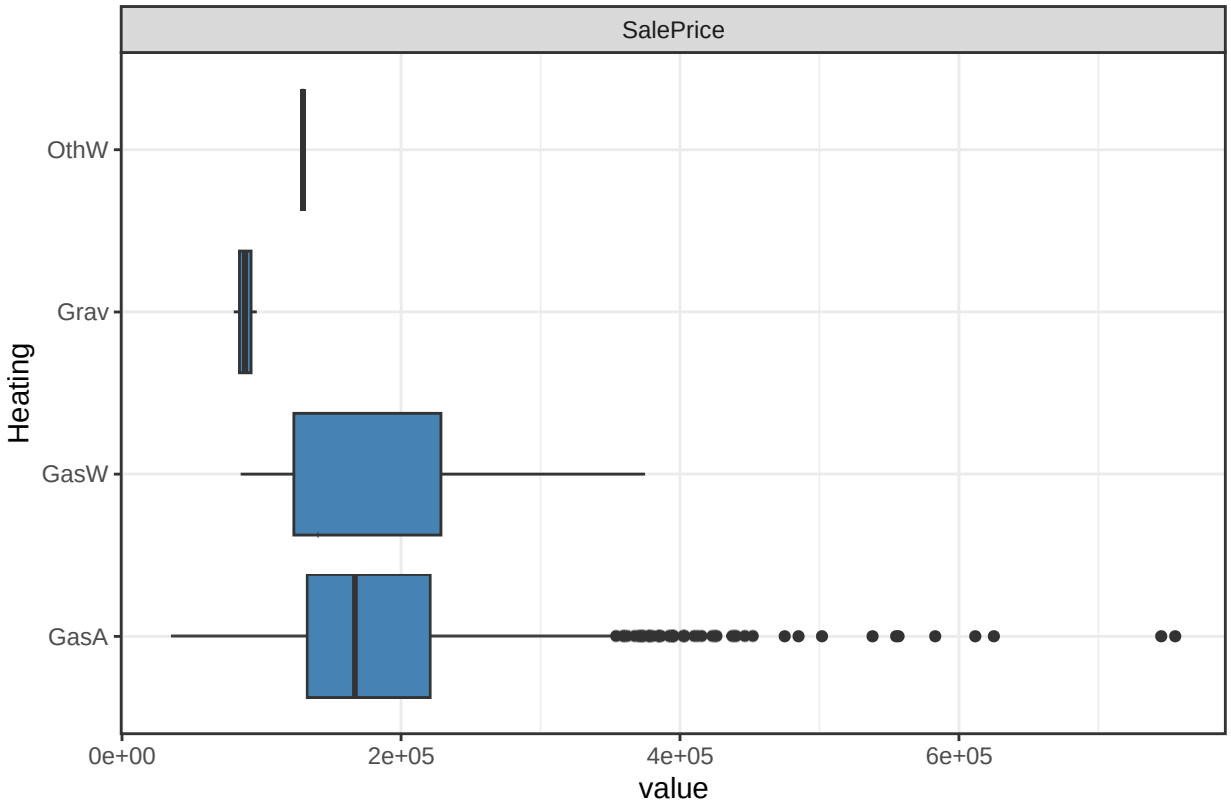

SalePrice
TA
Gd
Fa
Ex
BsmtQual
0e+00
2e+05
4e+05
6e+05
value

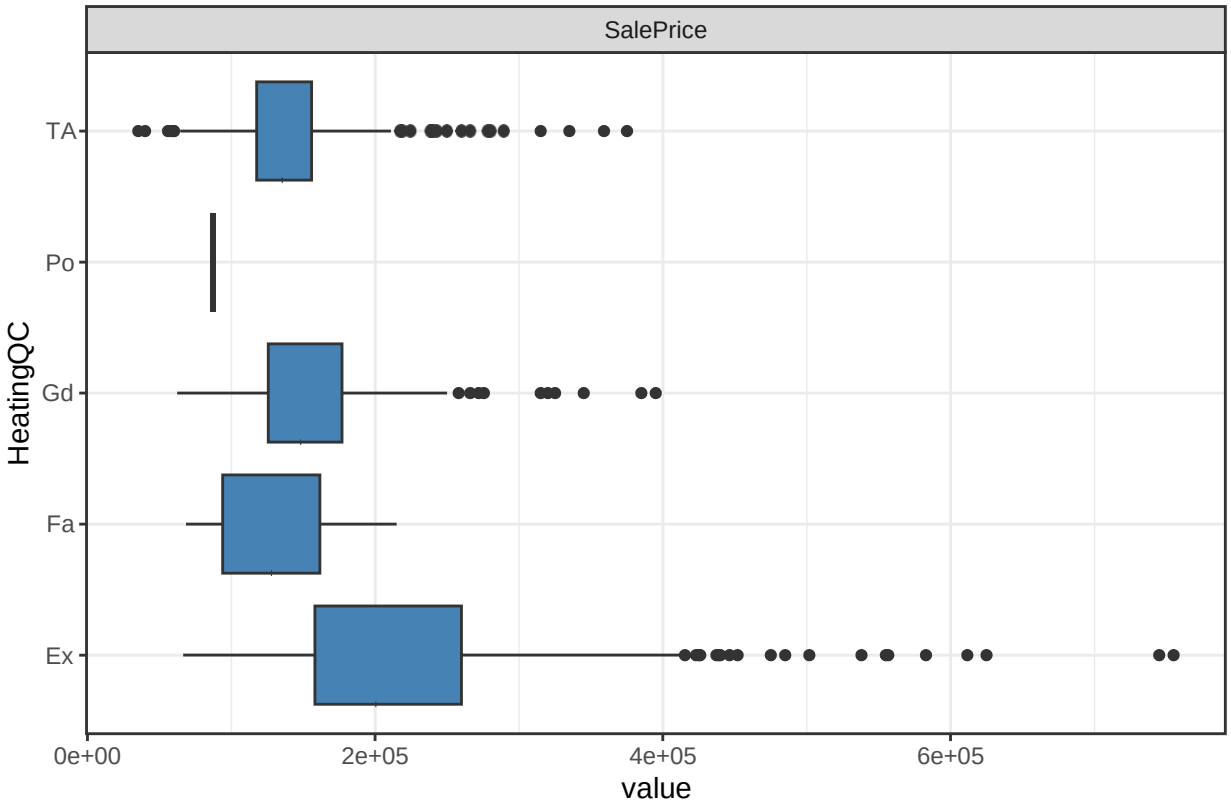

SalePrice
TA
Po
Gd
Fa
BsmtCond
0e+00
2e+05
4e+05
6e+05
value

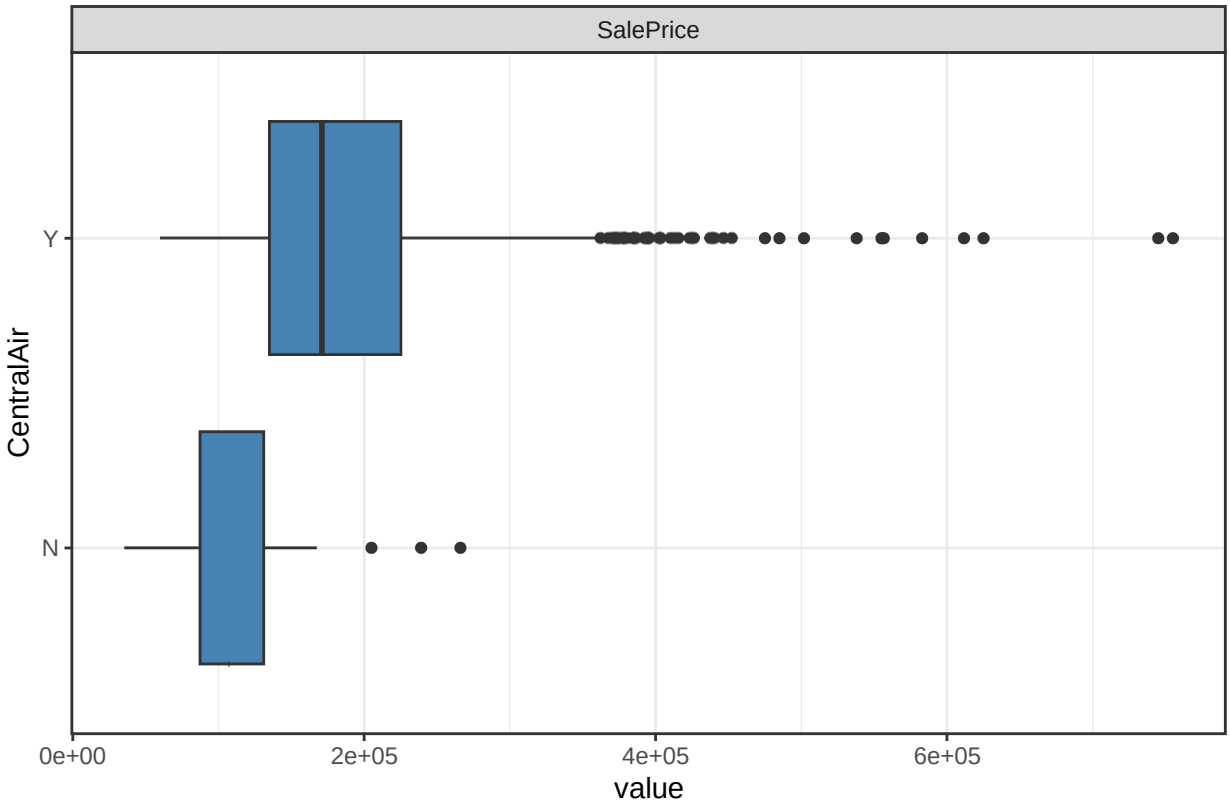

SalePrice
No
Mn
Gd
Av
BsmtExposure
0e+00
2e+05
4e+05
6e+05
value

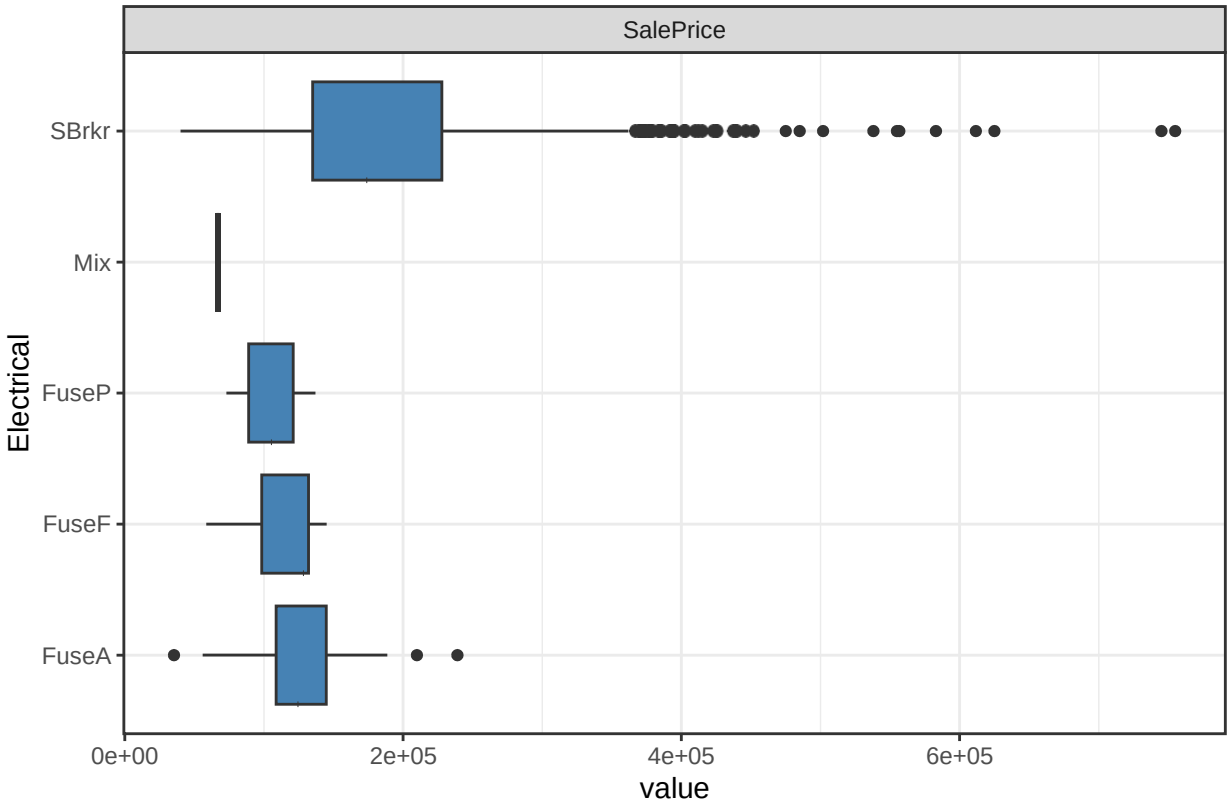

SalePrice
BsmtFinType1
Unf
Rec
LwQ
GLQ
BLQ
ALQ
0e+00
2e+05
4e+05
6e+05
value

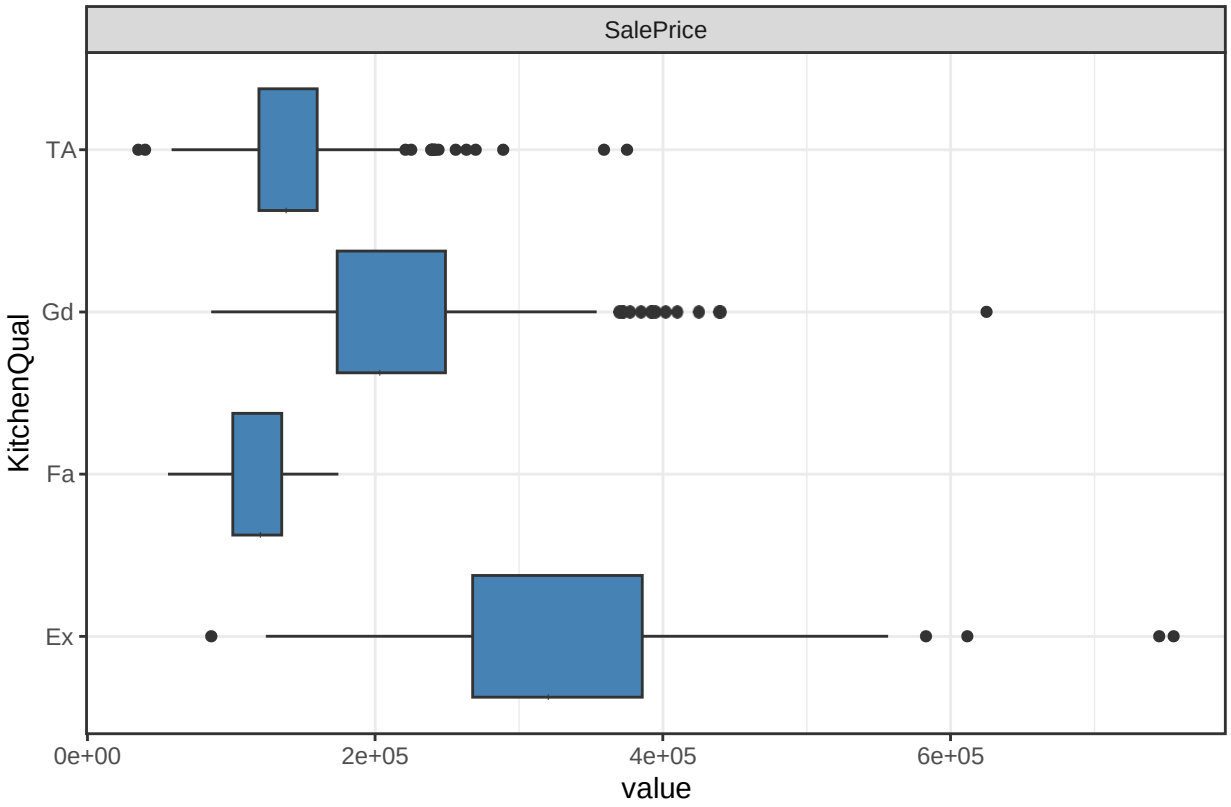

SalePrice
BsmtFinType2
Unf
Rec
LwQ
GLQ
BLQ
ALQ
0e+00
2e+05
4e+05
6e+05
value

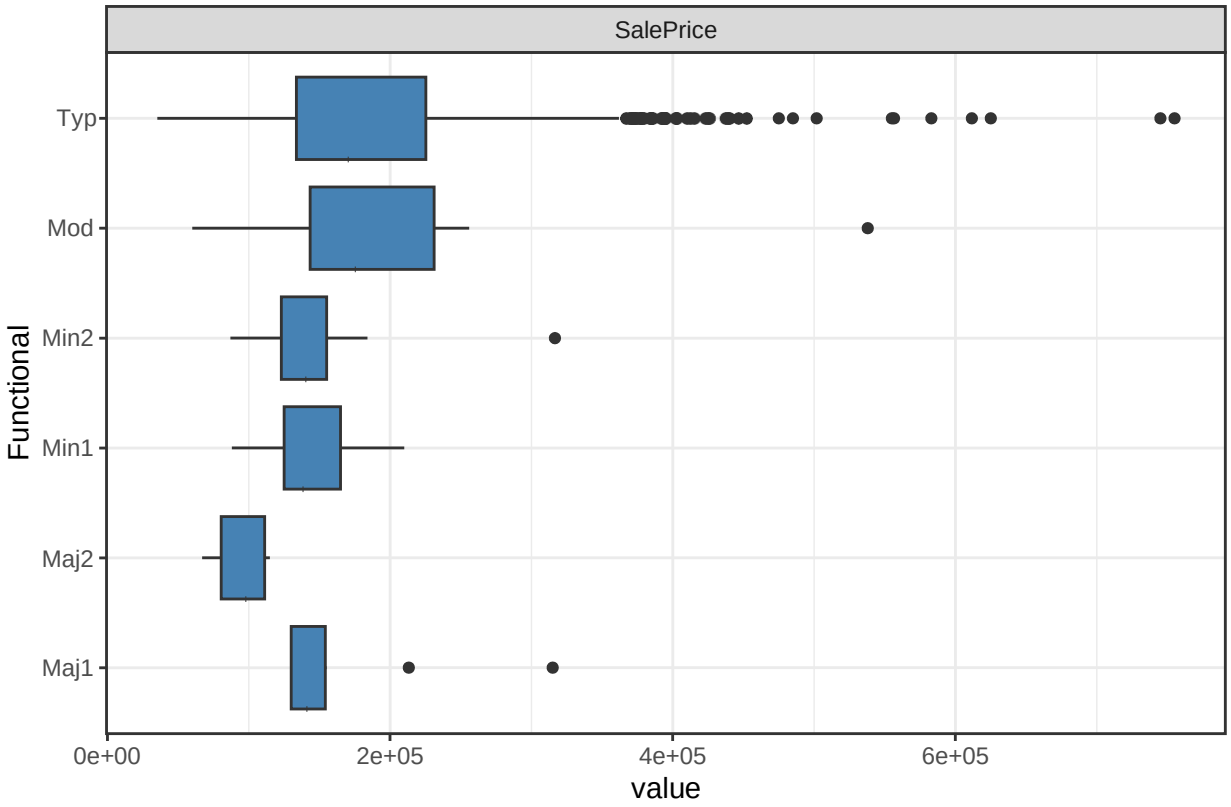

SalePrice
OthW
Grav
GasW
GasA
Heating
0e+00
2e+05
4e+05
6e+05
value

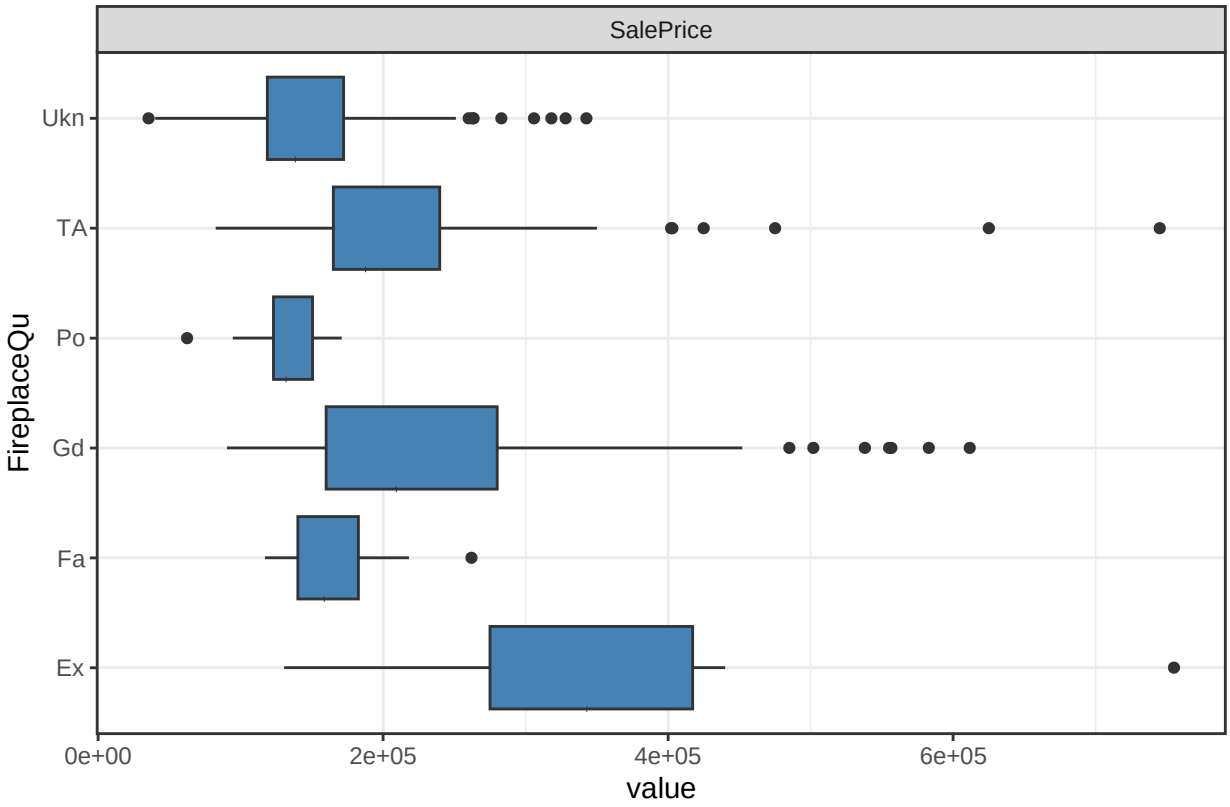

SalePrice
TA
Po
Gd
Fa
Ex
HeatingQC
0e+00
2e+05
4e+05
6e+05
value

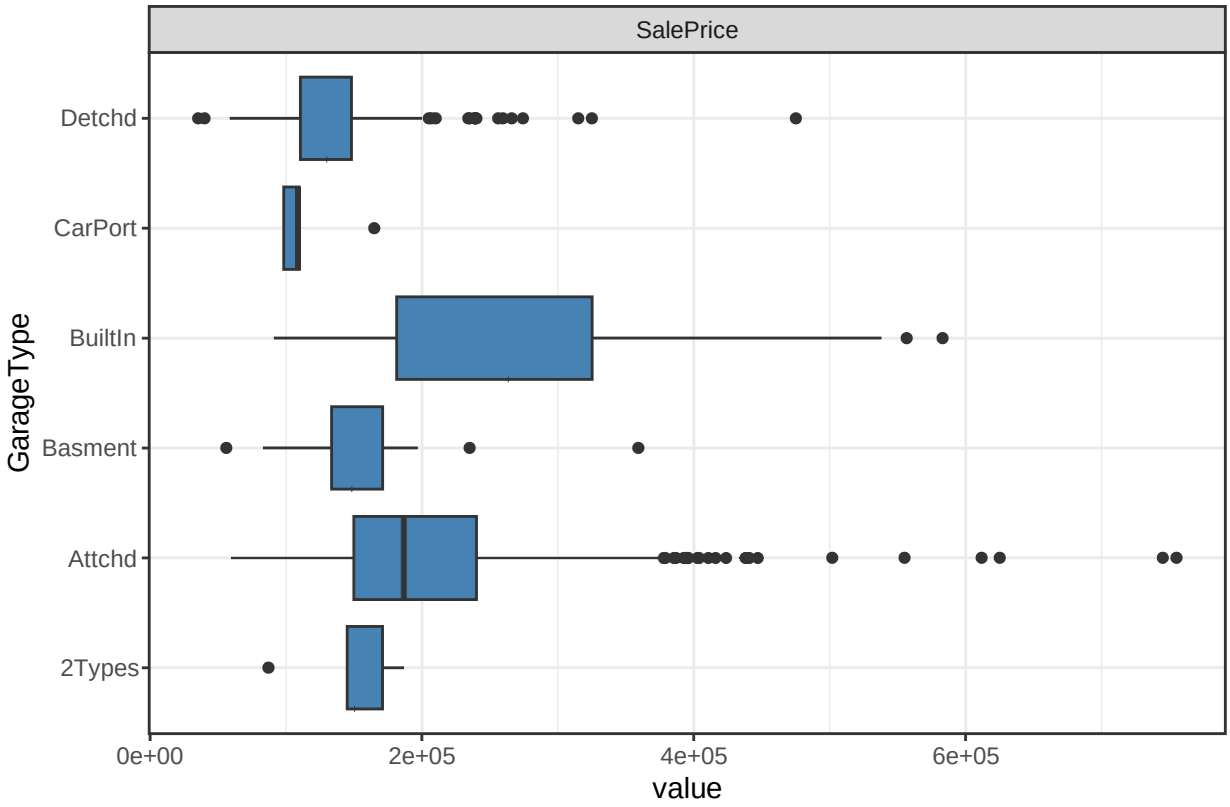

SalePrice
CentralAir
Y
N
0e+00
2e+05
4e+05
6e+05
value

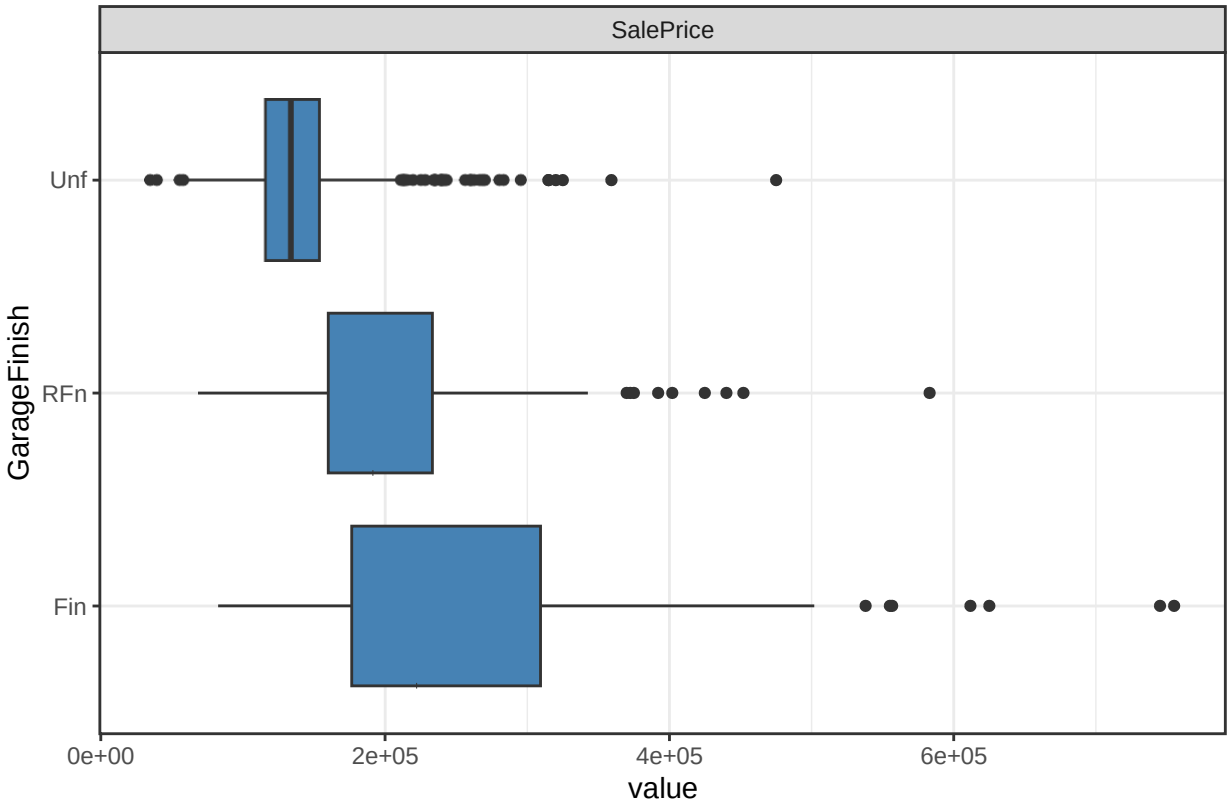

SalePrice
SBrkr
Mix
FuseP
FuseF
FuseA
Electrical
0e+00
2e+05
4e+05
6e+05
value

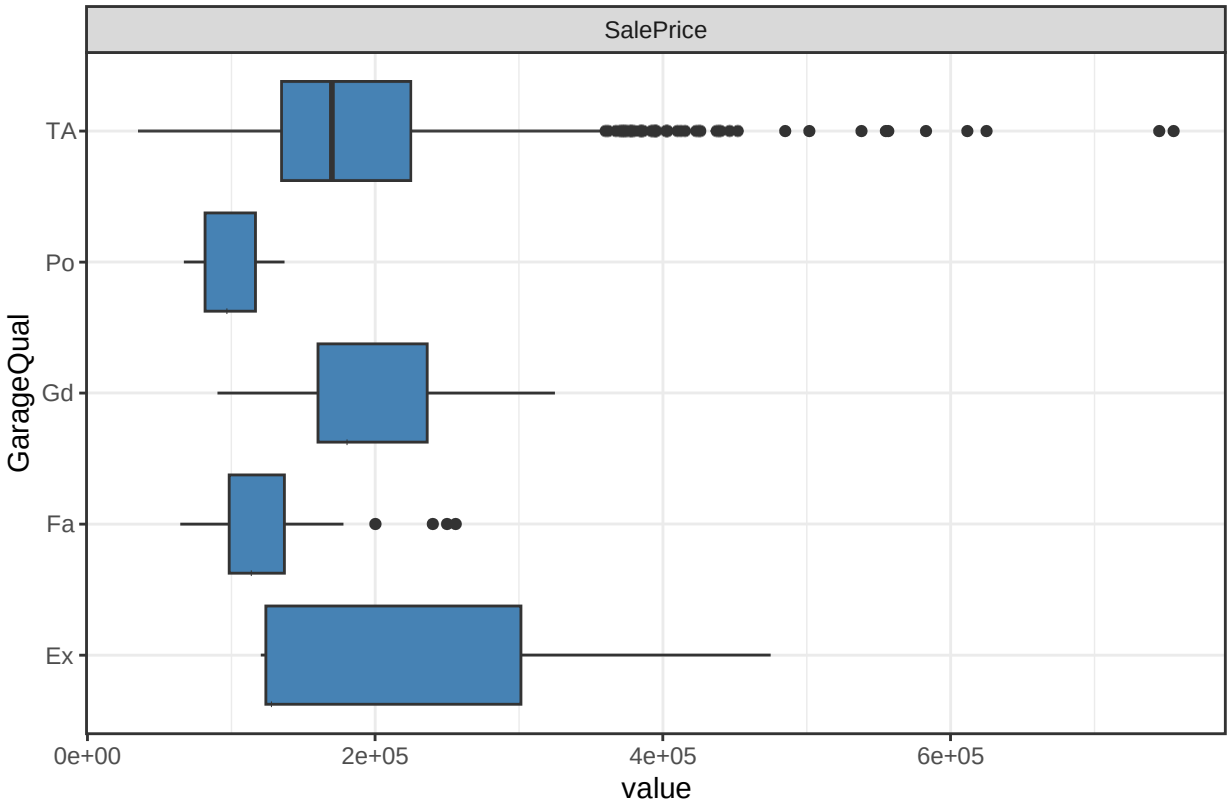

SalePrice
KitchenQual
TA
Gd
Fa
Ex
0e+00
2e+05
4e+05
6e+05
value

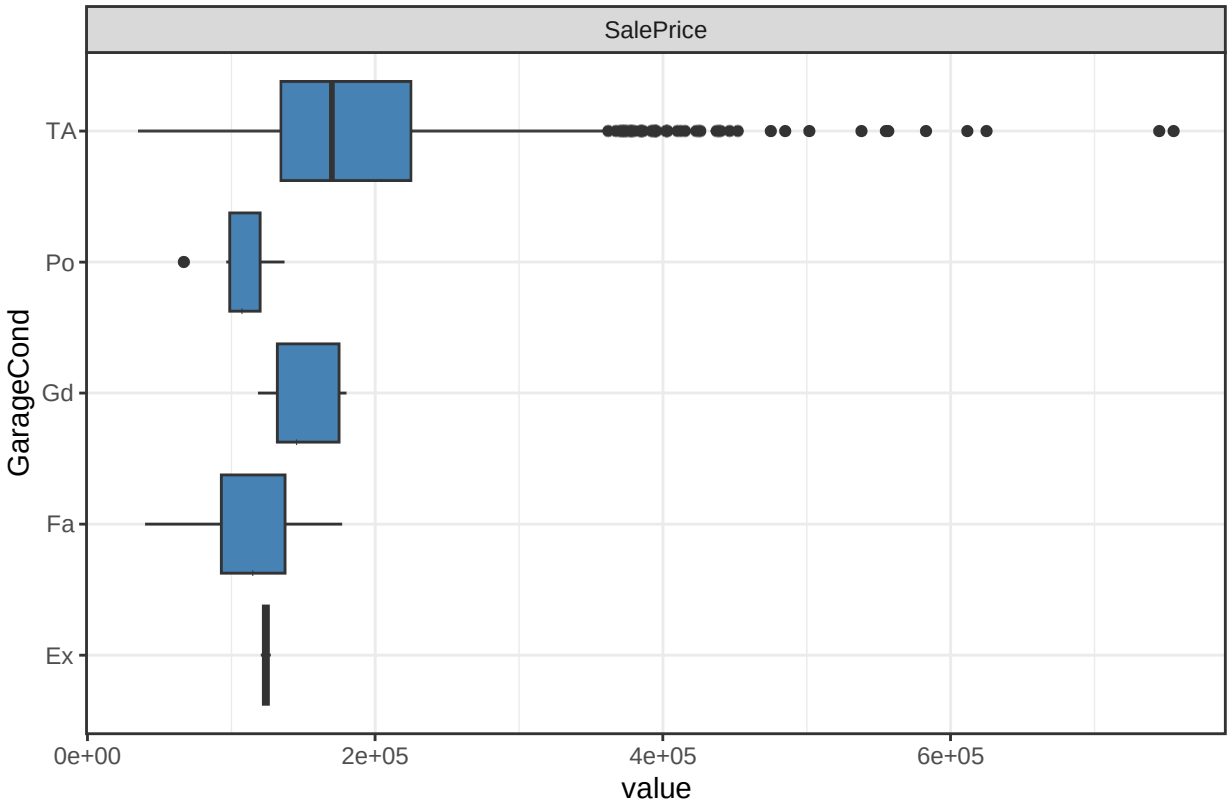

SalePrice
Functional
Typ
Mod
Min2
Min1
Maj2
Maj1
0e+00
2e+05
4e+05
6e+05
value

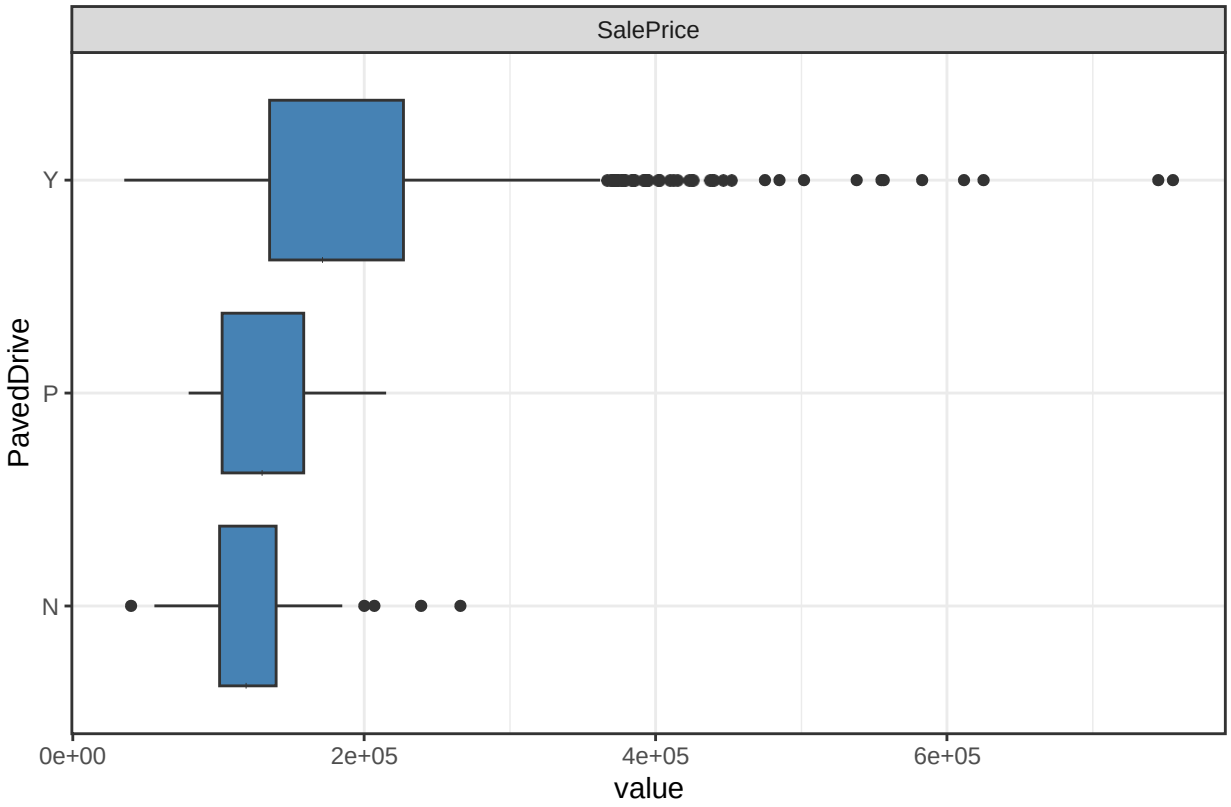

SalePrice
FireplaceQu
Ukn
TA
Po
Gd
Fa
Ex
0e+00
2e+05
4e+05
6e+05
value

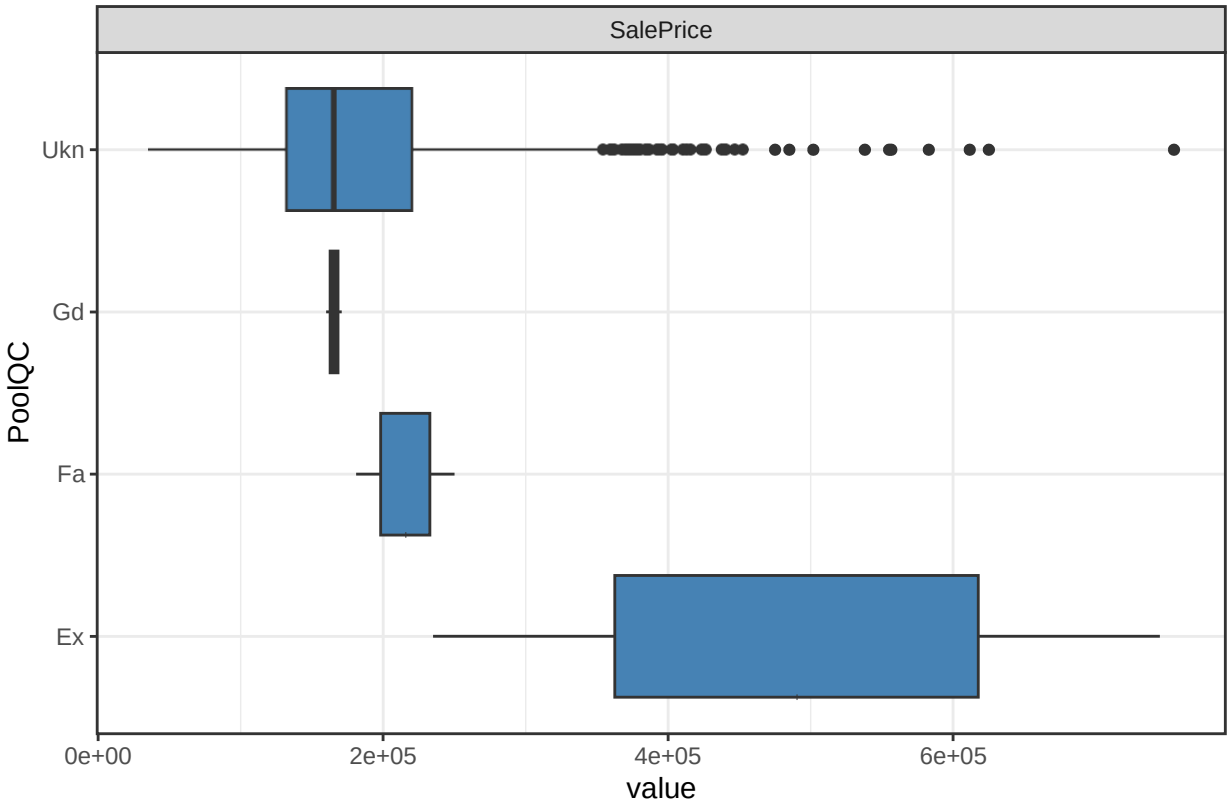

SalePrice
Detchd
CarPort
BuiltIn
Basment
Attchd
2Types
GarageType
0e+00
2e+05
4e+05
6e+05
value

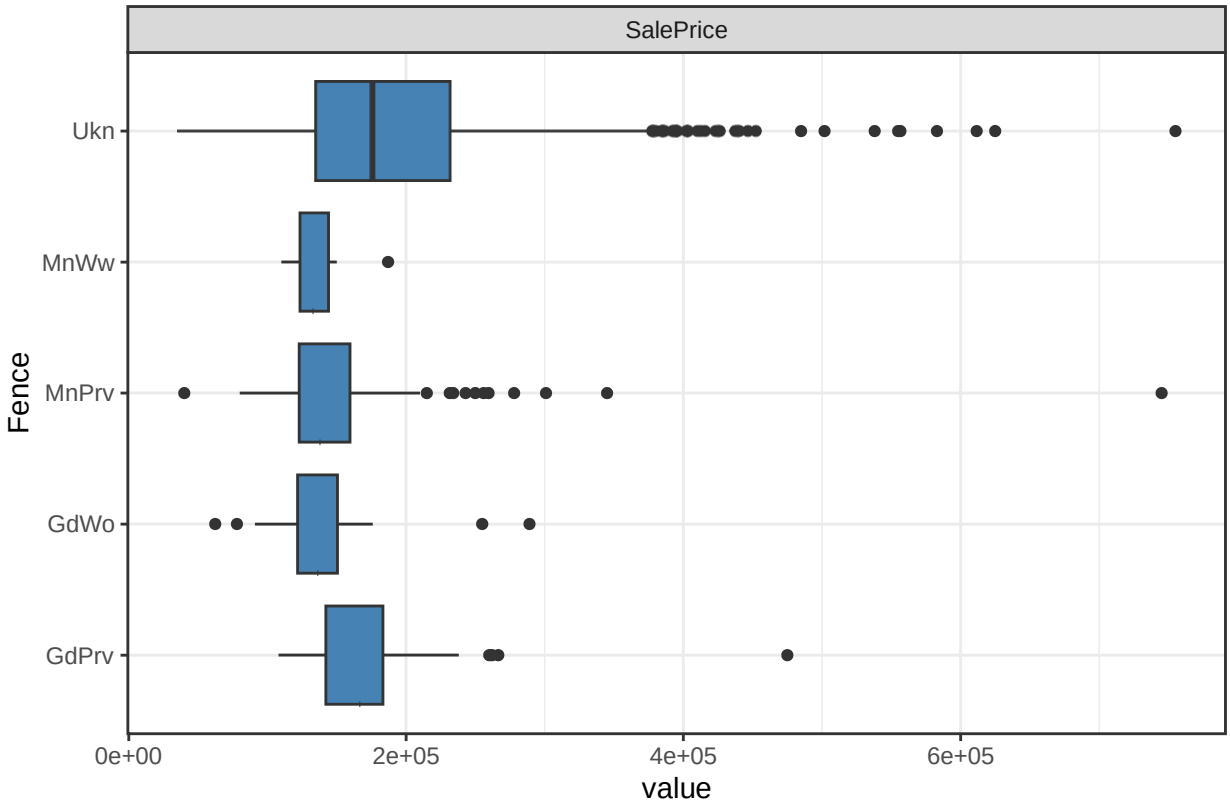

SalePrice
GarageFinish
Unf
RFn
Fin
0e+00
2e+05
4e+05
6e+05
value

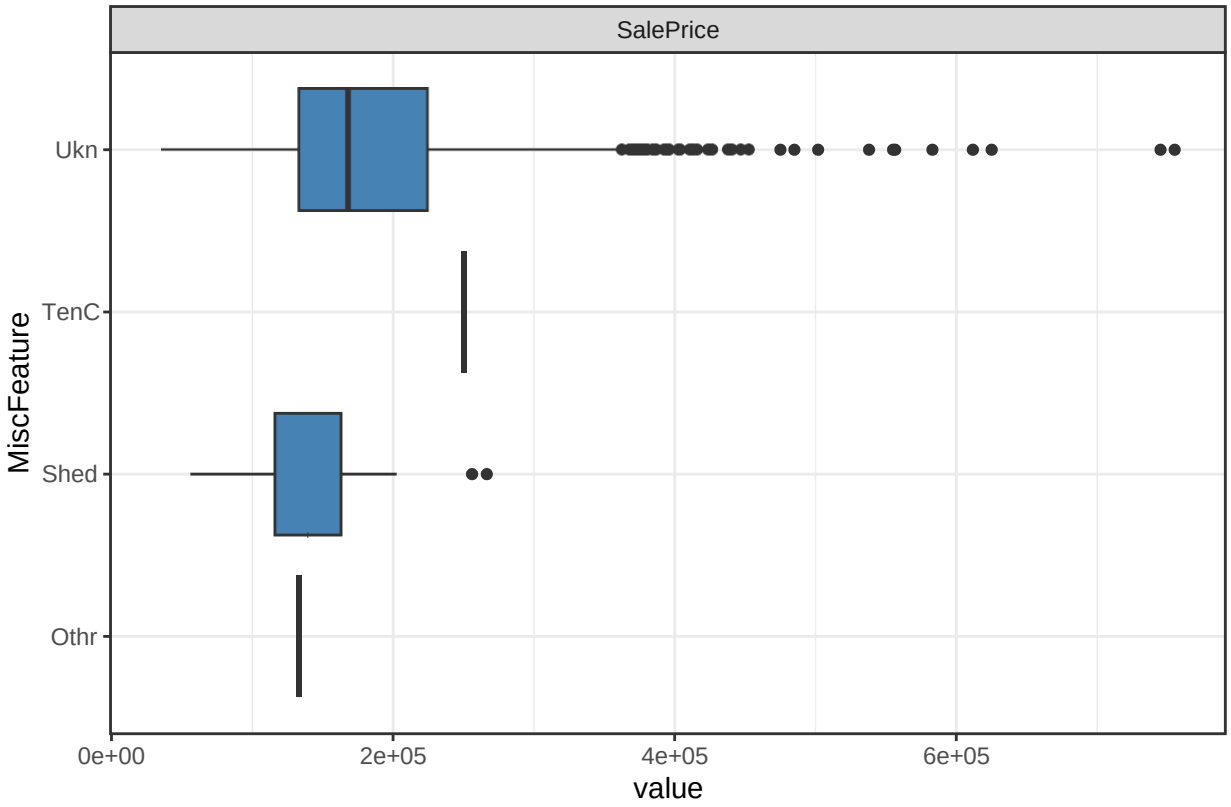

SalePrice
TA
Po
Gd
Fa
Ex
GarageQual
0e+00
2e+05
4e+05
6e+05
value

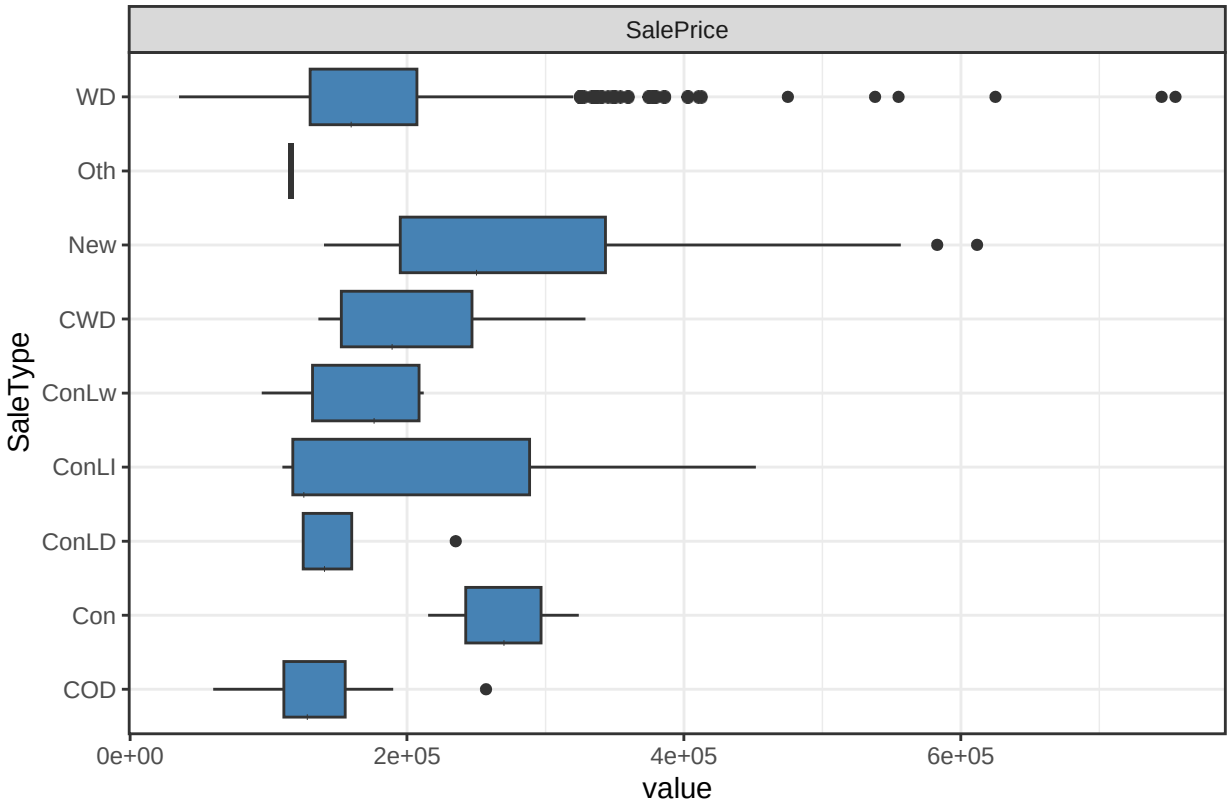

SalePrice
TA
Po
Gd
Fa
Ex
GarageCond
0e+00
2e+05
4e+05
6e+05
value

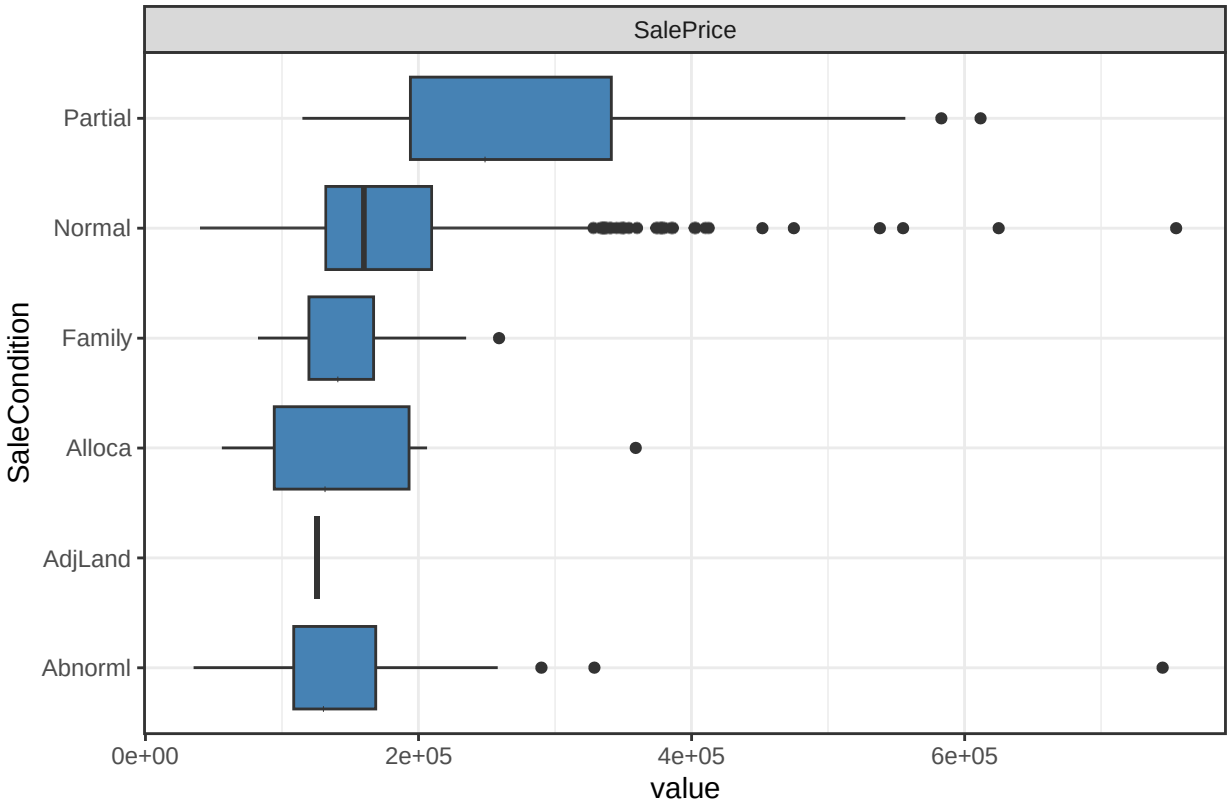

SalePrice
PavedDrive
Y
P
N
0e+00
2e+05
4e+05
6e+05
value


SalePrice
PoolQC
Ukn
Gd
Fa
Ex
0e+00
2e+05
4e+05
6e+05
value


SalePrice
Ukn
MnWw
MnPrv
GdWo
GdPrv
Fence
0e+00
2e+05
4e+05
6e+05
value


SalePrice
Ukn
TenC
Shed
Othr
MiscFeature
0e+00
2e+05
4e+05
6e+05
value


SalePrice
SaleType
WD
Oth
New
CWD
ConLw
ConLI
ConLD
Con
COD
0e+00
2e+05
4e+05
6e+05
value



#Credit: “Gerry Alfa Dito”