Modelagem

Davi Guerra

Separando as variáveis numéricas e categóricas

```
## [1] 76
```

```
## [1] 7
```

```
## [1] 8
```

```
## [1] 37
```

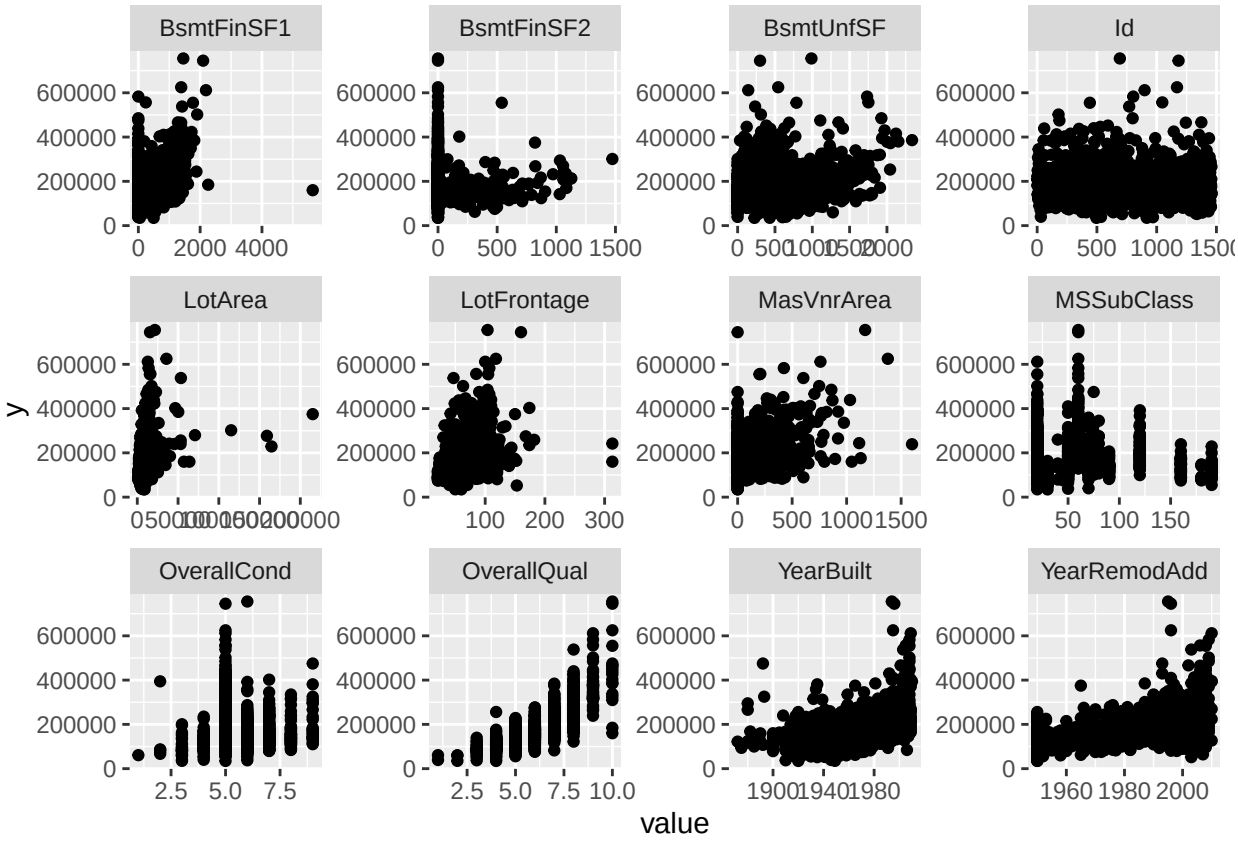
```
## [1] 32
```

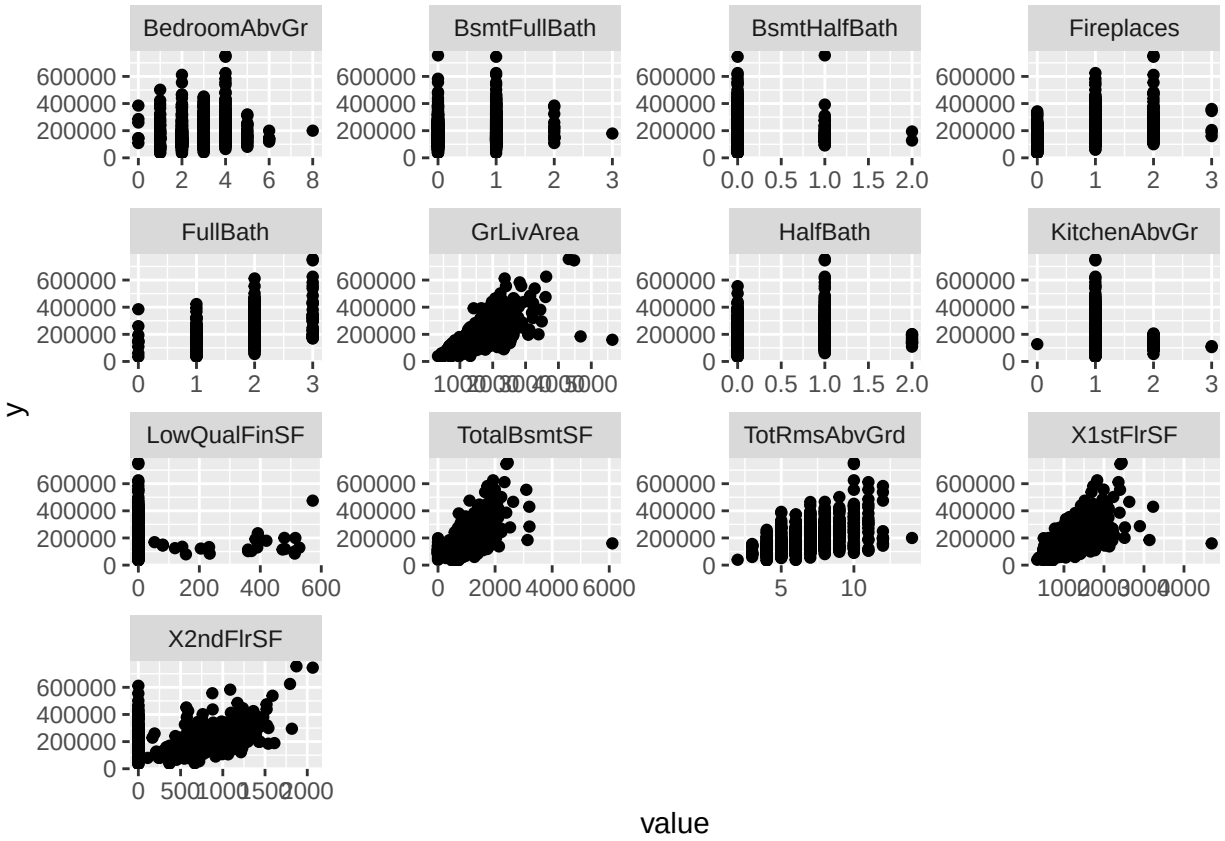
```
## [1] 32
```

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

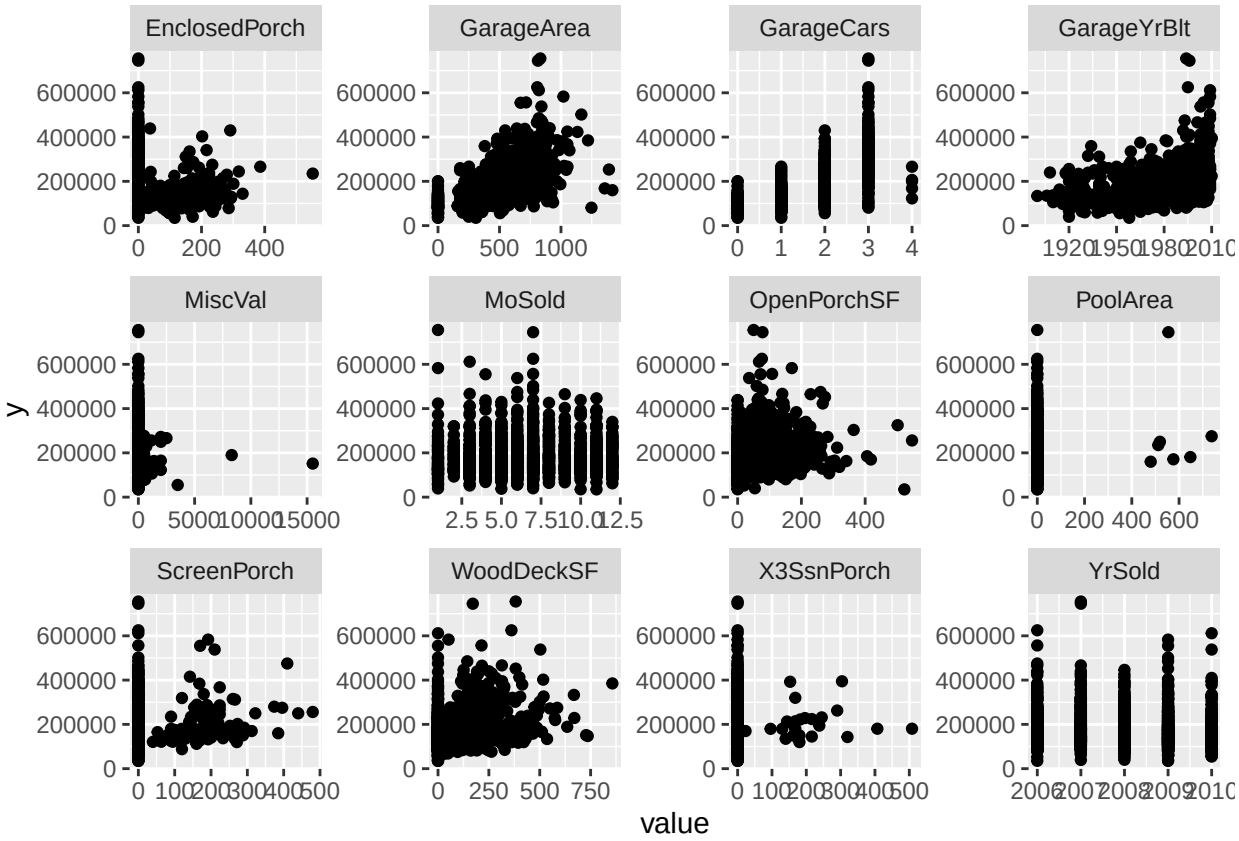
Gráfico de Dispersão das covariáveis numéricas pela variável resposta

```
## Warning: Removed 267 rows containing missing values (geom_point).
```


BsmtFinSF1
BsmtFinSF2
BsmtUnfSF
Id
LotArea
LotFrontage
MasVnrArea
MSSubClass
OverallCond
OverallQual
YearBuilt
YearRemodAdd
y
value




```
## Warning: Removed 81 rows containing missing values (geom_point).
```



Teste de correlação de pearson

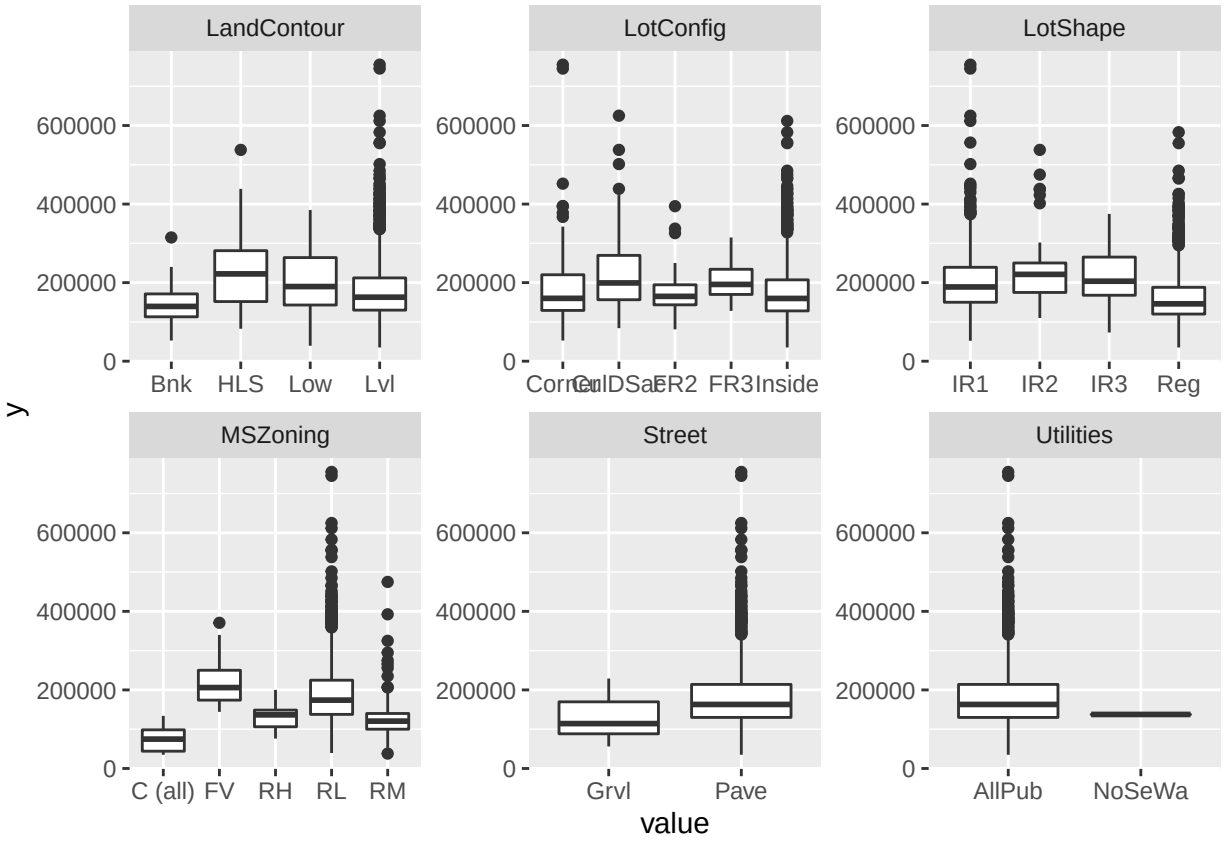
```
##            Id    MSSubClass   LotFrontage       LotArea   OverallQual
##       0.40269       0.00127       0.00000       0.00000       0.00000
##   OverallCond     YearBuilt  YearRemodAdd    MasVnrArea    BsmtFinSF1
##       0.00291       0.00000       0.00000       0.00000       0.00000
##    BsmtFinSF2     BsmtUnfSF   TotalBsmtSF     X1stFlrSF     X2ndFlrSF
##       0.66400       0.00000       0.00000       0.00000       0.00000
##  LowQualFinSF     GrLivArea  BsmtFullBath  BsmtHalfBath      FullBath
##       0.32821       0.00000       0.00000       0.52015       0.00000
##      HalfBath  BedroomAbvGr  KitchenAbvGr  TotRmsAbvGrd    Fireplaces
##       0.00000       0.00000       0.00000       0.00000       0.00000
##   GarageYrBlt    GarageCars    GarageArea    WoodDeckSF   OpenPorchSF
##       0.00000       0.00000       0.00000       0.00000       0.00000
## EnclosedPorch    X3SsnPorch   ScreenPorch      PoolArea       MiscVal
##       0.00000       0.08858       0.00002       0.00041       0.41849
##        MoSold        YrSold
##       0.07613       0.26941
```

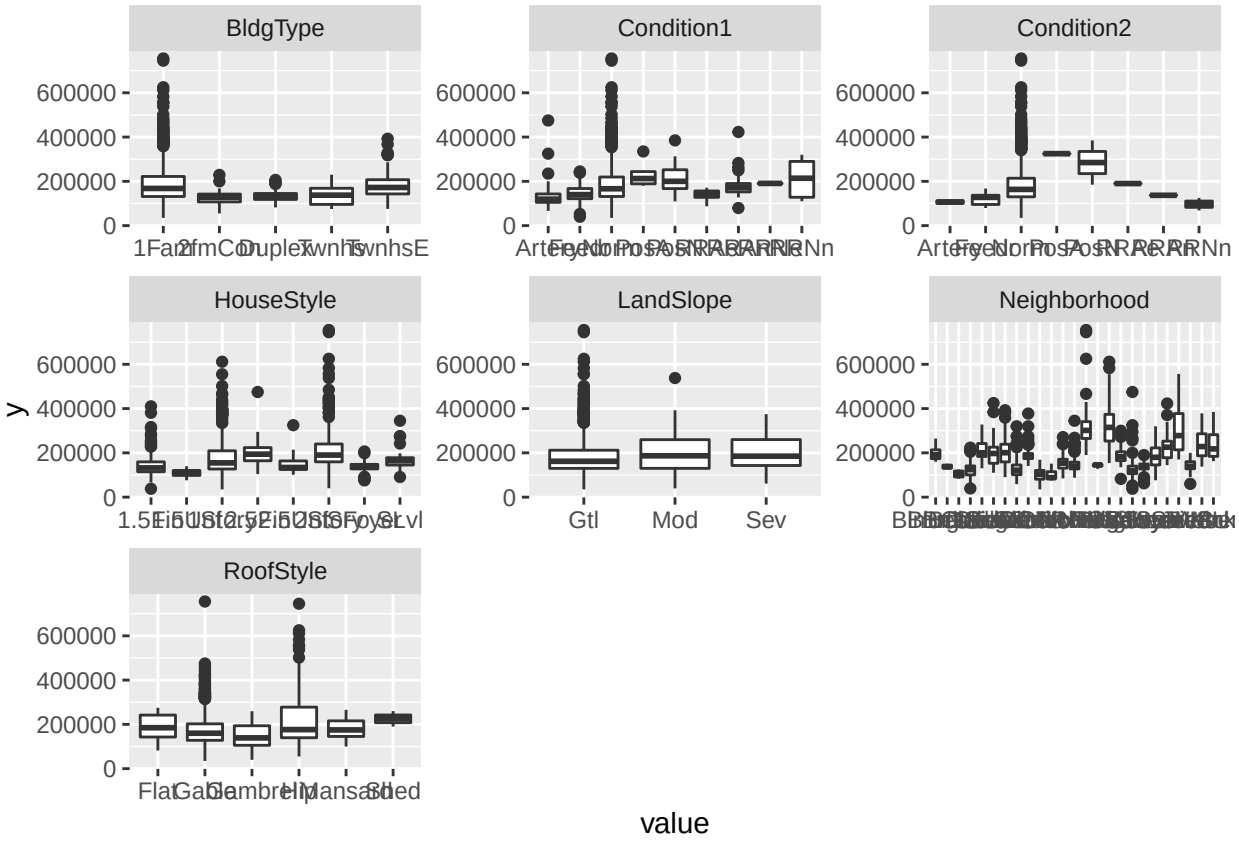
```
##   MSSubClass LotFrontage LotArea OverallQual OverallCond YearBuilt YearRemodAdd
## 1         60          65    8450           7           5      2003         2003
## 2         20          80    9600           6           8      1976         1976
## 3         60          68   11250           7           5      2001         2002
## 4         70          60    9550           7           5      1915         1970
## 5         60          84   14260           8           5      2000         2000
```

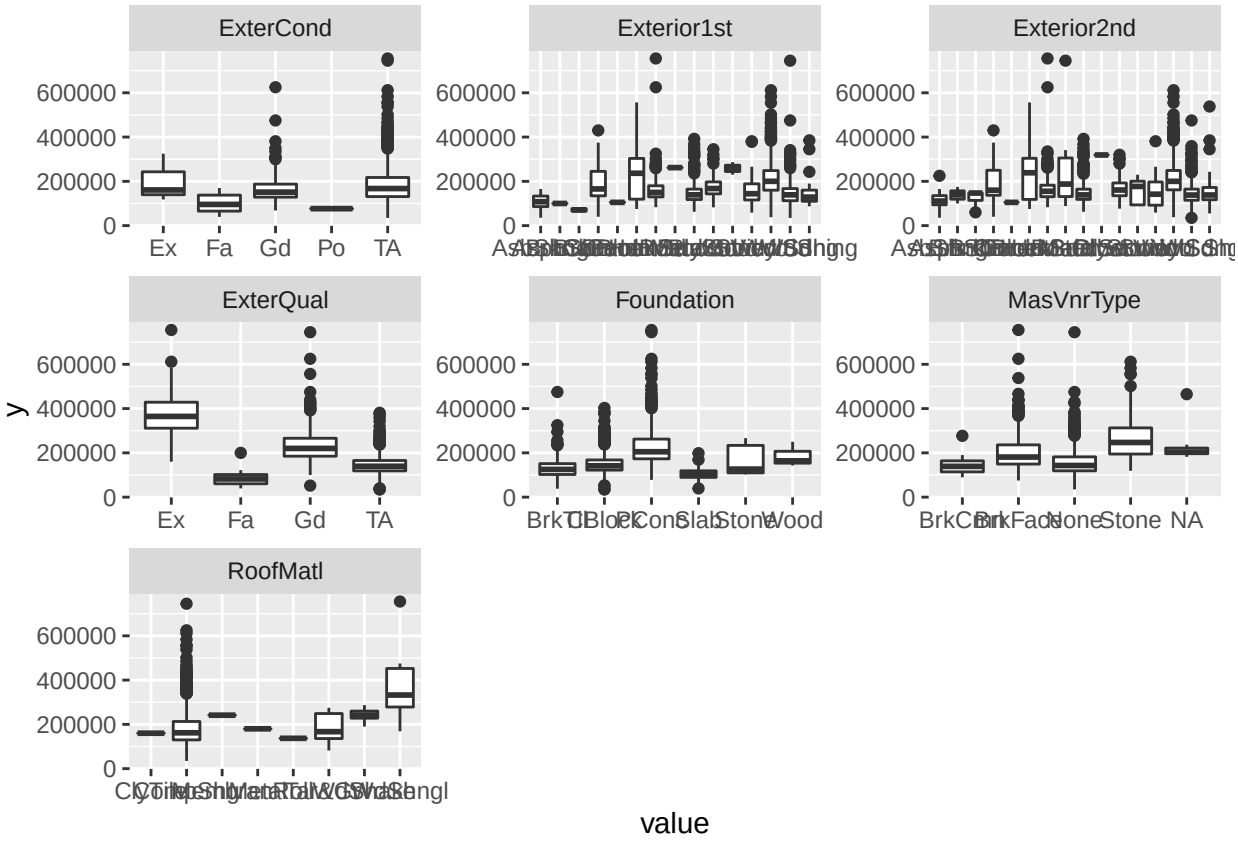
```
## 6         50          85   14115           5           5      1993         1995
##   MasVnrArea BsmtFinSF1 BsmtUnfSF TotalBsmtSF X1stFlrSF X2ndFlrSF GrLivArea
## 1        196        706       150         856       856       854      1710
## 2          0        978       284        1262      1262         0      1262
## 3        162        486       434         920       920       866      1786
## 4          0        216       540         756       961       756      1717
## 5        350        655       490        1145      1145      1053      2198
## 6          0        732        64         796       796       566      1362
##   BsmtFullBath FullBath HalfBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd
## 1            1        2        1            3            1            8
## 2            0        2        0            3            1            6
## 3            1        2        1            3            1            6
## 4            1        1        0            3            1            7
## 5            1        2        1            4            1            9
## 6            1        1        1            1            1            5
##   Fireplaces GarageYrBlt GarageCars GarageArea WoodDeckSF OpenPorchSF
## 1          0        2003          2        548          0          61
## 2          1        1976          2        460        298           0
## 3          1        2001          2        608          0          42
## 4          1        1998          3        642          0          35
## 5          1        2000          3        836        192          84
## 6          0        1993          2        480         40          30
##   EnclosedPorch ScreenPorch PoolArea
## 1             0           0        0
## 2             0           0        0
## 3             0           0        0
## 4           272           0        0
## 5             0           0        0
## 6             0           0        0
```

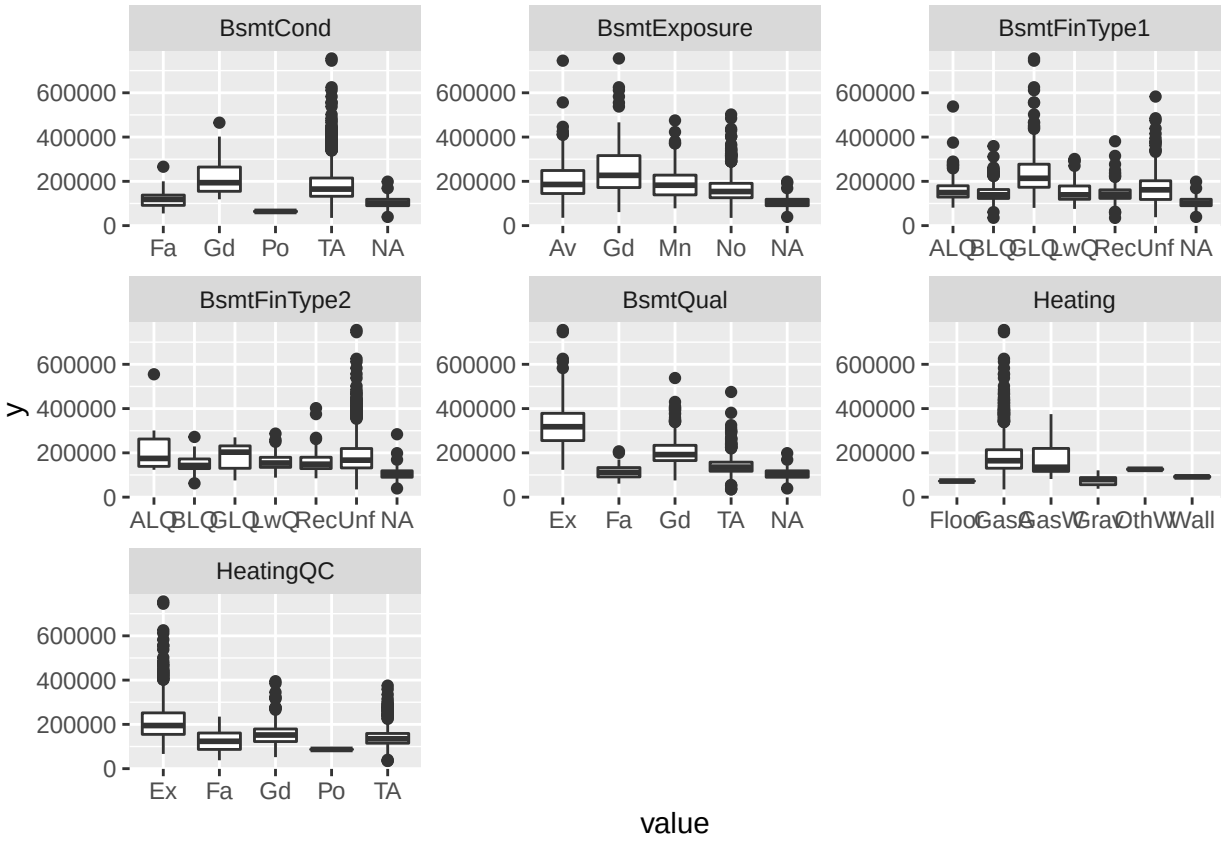
Boxplot das covariáveis categóricas em relação à variável resposta

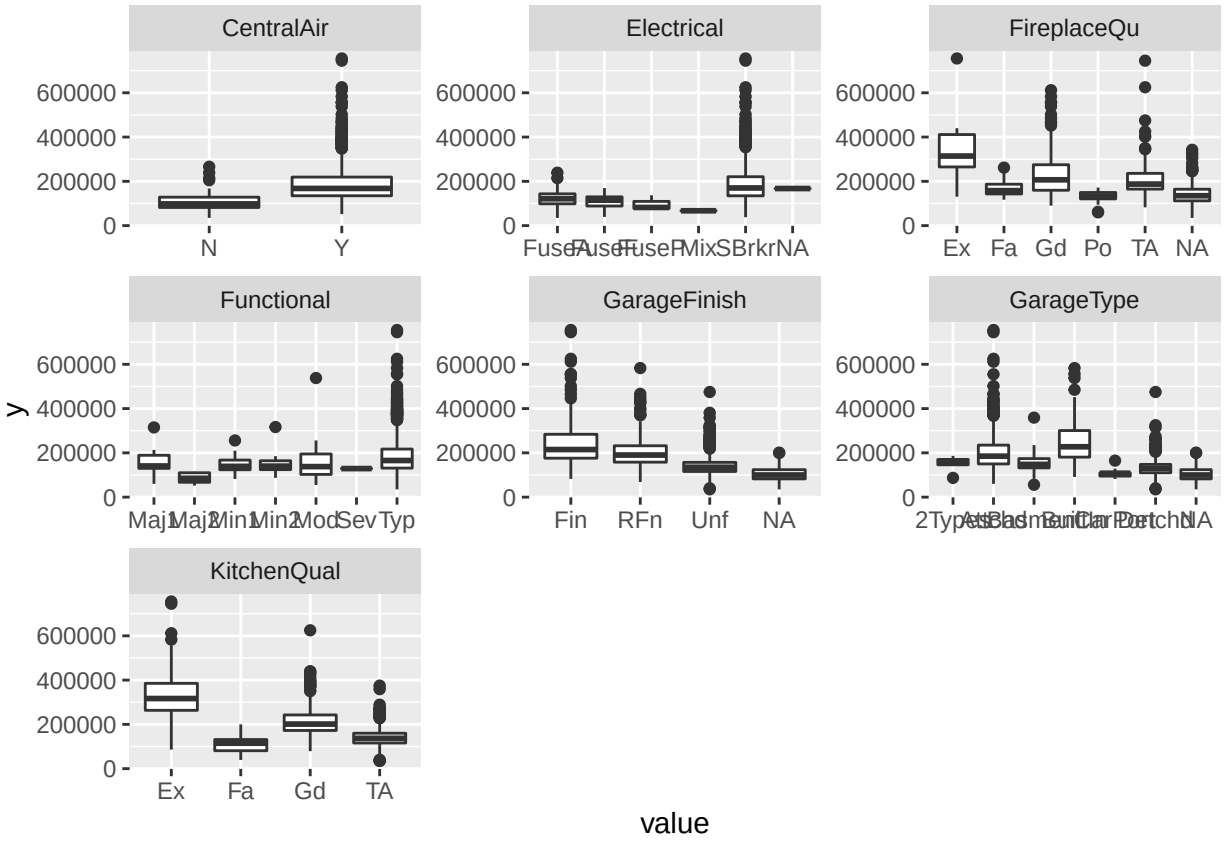
Como o intuito é verificar mais se dentro das covariáveis alguma varíavel apresenta maior influência que as outras, os nomes dentro das variáveis ficou corrompido, por isso, caso haja necessidade de ver algum covariável com mais detalhe posso criar um gráfico só pra ela.

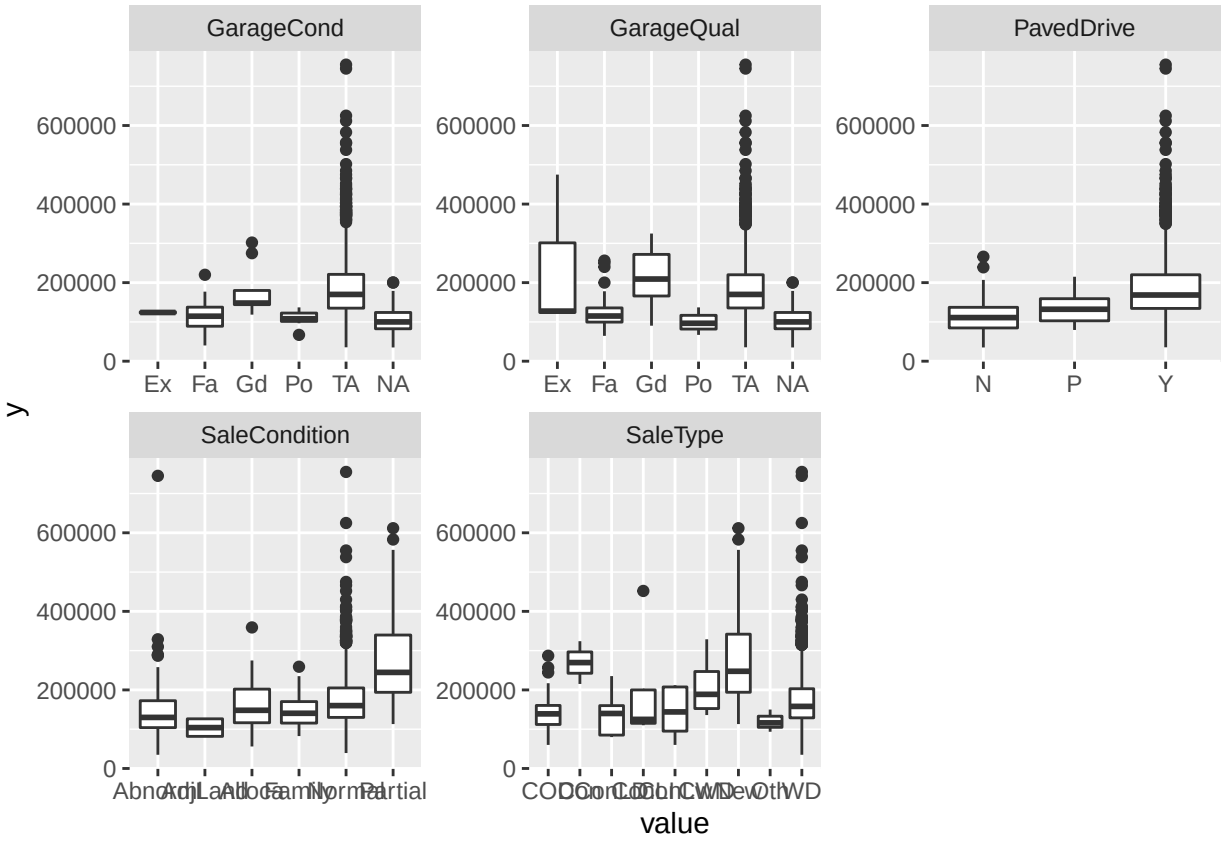

LandContour
LotConfig
LotShape
MSZoning
Street
Utilities
y
value
0
200000
400000
600000
Bnk
HLS
Low
Lvl
Corner
CulDSac
FR2
FR3
Inside
IR1
IR2
IR3
Reg
C (all)
FV
RH
RL
RM
Grvl
Pave
AllPub
NoSeWa


BldgType
Condition1
Condition2
HouseStyle
LandSlope
Neighborhood
RoofStyle
y
value
0
200000
400000
600000
Gtl
Mod
Sev


ExterCond
Exterior1st
Exterior2nd
ExterQual
Foundation
MasVnrType
RoofMatl
y
value
600000
400000
200000
0
Ex
Fa
Gd
Po
TA
BrkCmn
BrkFace
None
Stone
NA


BsmtCond
BsmtExposure
BsmtFinType1
BsmtFinType2
BsmtQual
Heating
HeatingQC
y
value
0
200000
400000
600000
Fa Gd Po TA NA
Av Gd Mn No NA
ALQ BLQ GLQ LwQ Rec Unf NA
Ex Fa Gd TA NA
Floor Gas GasW Grav OthW Wall
Ex Fa Gd Po TA


CentralAir
Electrical
FireplaceQu
Functional
GarageFinish
GarageType
KitchenQual
y
value
600000
400000
200000
0
N
Y
Ex
Fa
Gd
Po
TA
NA
Fin
RFn
Unf
Typ
Sev
Mod
Min1
Min2
Maj1
Maj2
2Types
Attchd
Basment
BuiltIn
CarPort
Detchd



Fazendo o teste da ANOVA

```
##                Df        Sum Sq     Mean Sq F value                 Pr(>F)
## MSZoning        3  247982781161 82660927054 51.1904 < 0.00000000000000022 ***
## LotShape        3  209563189418 69854396473 43.2596 < 0.00000000000000022 ***
## LandContour     3  113643260011 37881086670 23.4591  0.00000000000025110 ***
## Utilities       1    8695058870  8695058870  5.3847              0.0206675 *
## LotConfig       4   35566747349  8891686837  5.5065              0.0002355 ***
## LandSlope       2    9784324847  4892162423  3.0296              0.0491155 *
## Neighborhood   24 2258901058219 94120877426 58.2873 < 0.00000000000000022 ***
## Condition1      7   26356121924  3765160275  2.3317              0.0236741 *
## Condition2      4   34932009471  8733002368  5.4082              0.0002802 ***
## BldgType        4  269236101153 67309025288 41.6833 < 0.00000000000000022 ***
## HouseStyle      7   58436841231  8348120176  5.1698  0.000010097313630804 ***
## RoofStyle       5  176488094453 35297618891 21.8592 < 0.00000000000000022 ***
## RoofMatl        6  179337222622 29889537104 18.5100 < 0.00000000000000022 ***
## Exterior1st    11  188238664474 17112605861 10.5975 < 0.00000000000000022 ***
## Exterior2nd    14   79894901666  5706778690  3.5341  0.000013467699172397 ***
## MasVnrType      3  102172368685 34057456228 21.0912  0.00000000000575464 ***
## ExterQual       3  180701269991 60233756664 37.3017 < 0.00000000000000022 ***
## BsmtQual        3  118553690374 39517896791 24.4727  0.00000000000006632 ***
## BsmtExposure    3  110670296940 36890098980 22.8454  0.00000000000056377 ***
## BsmtFinType1    5   34814719546  6962943909  4.3120              0.0007372 ***
## KitchenQual     3   71540598180 23846866060 14.7679  0.00000002852314055 ***
## SaleType        7   24721413902  3531630557  2.1871              0.0338436 *
```

```
## SaleCondition  4   41007437442 10251859361  6.3488  0.000053002355155008 ***
## NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  [1] "MSZoning"      "LotShape"      "LandContour"   "Utilities"
##  [5] "LotConfig"     "LandSlope"     "Neighborhood"  "Condition1"
##  [9] "Condition2"    "BldgType"      "HouseStyle"    "RoofStyle"
## [13] "RoofMatl"      "Exterior1st"   "Exterior2nd"   "MasVnrType"
## [17] "ExterQual"     "BsmtQual"      "BsmtExposure"  "BsmtFinType1"
## [21] "KitchenQual"   "SaleType"      "SaleCondition" "MSSubClass"
## [25] "LotFrontage"   "LotArea"       "OverallQual"   "OverallCond"
## [29] "YearBuilt"     "YearRemodAdd"  "MasVnrArea"    "BsmtFinSF1"
## [33] "BsmtUnfSF"     "TotalBsmtSF"   "X1stFlrSF"     "X2ndFlrSF"
## [37] "GrLivArea"     "BsmtFullBath"  "FullBath"      "HalfBath"
## [41] "BedroomAbvGr"  "KitchenAbvGr"  "TotRmsAbvGrd"  "Fireplaces"
## [45] "GarageYrBlt"   "GarageCars"    "GarageArea"    "WoodDeckSF"
## [49] "OpenPorchSF"   "EnclosedPorch" "ScreenPorch"   "PoolArea"
```

Fazendo os testes de correlação e da anova, conseguiu-se reduzir o número de variáveis de 80 para 53.

Com isso, para o restante das análises serão utilizadas essas variáveis.

Criação das variáveis dummies

```
## [1] 1096  181
```

Transformando as variáveis categóricas em variáveis dummies aumentamos o número de variáveis do modelo de 53 para 181 variáveis

Seleção do modelo

Matriz de correlação entre as variáveis quantitativas

fazer um modelo com cada variável numérica pela variável resposta

#calcular o coeficiente de contingência

#comparação do teste da anova com cada variável categórica com a variável resposta