Laboratorio 1.

Data Science

Maria José Morales 19145

Luis Pedro García 19344

1. Haga una exploración rápida de sus datos para eso haga un resumen de su dataset.

Se hizo una breve exploración con la función “dim ()” y la función“summary ()” para ver las dimensiones del data set y un breve resumen de las variables.
En el data set train hay 47 variables cualitativas y 34 variables cuantitativas. Se encontraron algunas irregularidades en los nombres de las variables ya que algunas veces se pone “AbvGrd” para referirse a “Above Grade” y a veces se pone “AvdGr”.

2. Diga el tipo de cada una de las variables del dataset (cualitativa o categórica, cuantitativa continua, cuantitativa discreta)

Las variables junto con su categoría son las siguientes:

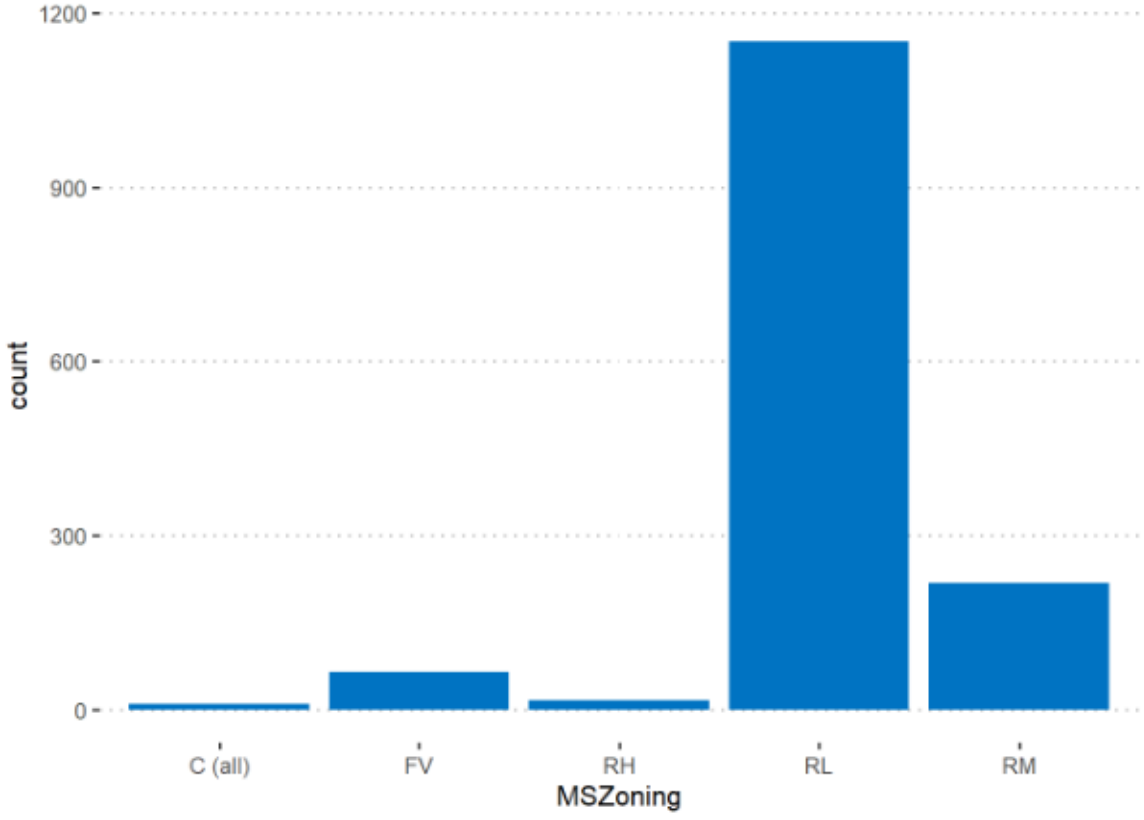
1. Id (cuantitativa discreta)
2. MSSubClass (cuantitativa discreta)
3. MSZoning (cualitativa)
4. LotFrontage(cuantitativa continua)
5. LotArea(cuantitativa continua)
6. Street(cualitativa)
7. Alley(cualitativa)
8. LotShape(cualitativa)
9. LandContour(cualitativa)
10. Utilities(cualitativa)
11. LotConfig(cualitativa)
12. LandSlope(cualitativa)
13. Neighborhood(cualitativa)
14. Condition1(cualitativa)
15. Condition2(cualitativa)
16. BldgType(cualitativa)
17. HouseStyle(cualitativa)
18. OverallQual (cuantitativa continua)
19. OverallCond(cuantitativa continua)
20. YearBuilt(cualitativa)
21. YearRemodAdd(cualitativa)
22. RoofStyle(cualitativa)
23. RoofMatl(cualitativa)

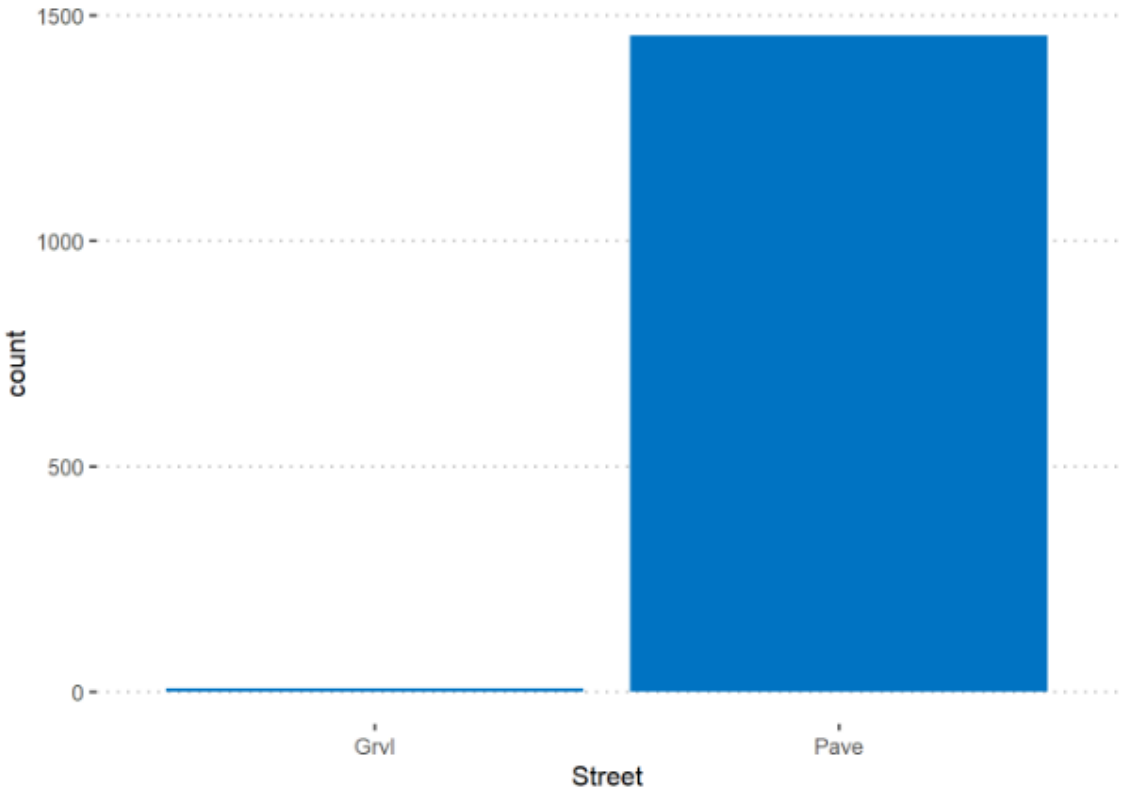
24. Exterior1st(cualitativa)
25. Exterior2nd(cualitativa)
26. MasVnrType(cualitativa)
27. MasVnrArea(cuantitativa continua)
28. ExterQual(cualitativa)
29. ExterCond(cualitativa)
30. Foundation(cualitativa)
31. BsmtQual(cualitativa)
32. BsmtCond(cualitativa)
33. BsmtExposure(cualitativa)
34. BsmtFinType1(cualitativa)
35. BsmtFinSF1(cuantitativa continua)
36. BsmtFinType2(cualitativa)
37. BsmtFinSF2(cuantitativa continua)
38. BsmtUnfSF(cuantitativa continua)
39. TotalBsmtSF(cuantitativa continua)
40. Heating(cualitativa)
41. HeatingQC(cualitativa)
42. CentralAir(cualitativa)
43. Electrical(cualitativa)
44. 1stFlrSF(cuantitativa continua)
45. 2ndFlrSF(cuantitativa continua)
46. LowQualFinSF(cuantitativa continua)
47. GrLivArea(cuantitativa continua)
48. BsmtFullBath(cuantitativa discreta)
49. BsmtHalfBath(cuantitativa discreta)
50. FullBath(cuantitativa discreta)
51. HalfBath(cuantitativa discreta)
52. BedroomAbvGr(cuantitativa discreta)
53. KitchenAbvGr(cuantitativa discreta)
54. KitchenQual(cualitativa)
55. TotRmsAbvGrd(cuantitativa discreta)
56. Functional(cualitativa)
57. Fireplaces(cuantitativa discreta)
58. FireplaceQu(cualitativa)
59. GarageType(cualitativa)
60. GarageYrBlt(cualitativa)
61. GarageFinish(cualitativa)
62. GarageCars(cuantitativa continua)
63. GarageArea(cuantitativa continua)
64. GarageQual(cualitativa)
65. GarageCond(cualitativa)
66. PavedDrive(cualitativa)
67. WoodDeckSF(cuantitativa continua)
68. OpenPorchSF(cuantitativa continua)
69. EnclosedPorch(cuantitativa continua)
70. 3SsnPorch(cuantitativa continua)

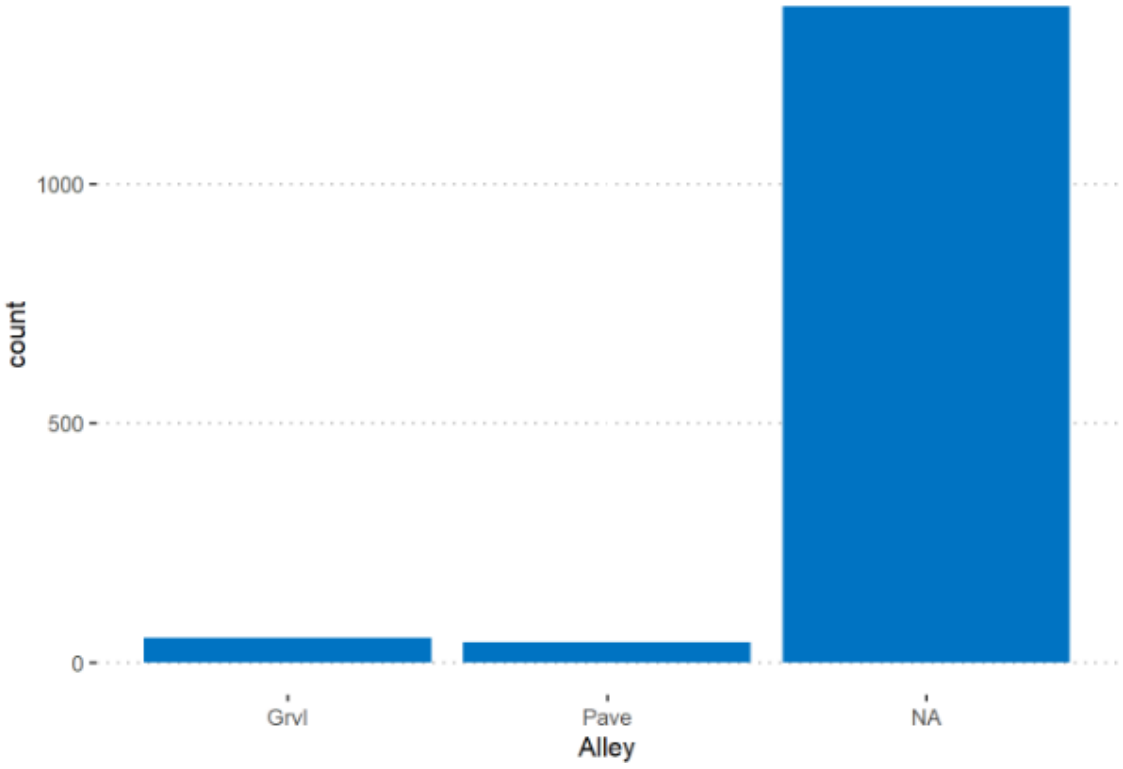
71. ScreenPorch(cuantitativa continua)
72. PoolArea(cuantitativa continua)
73. PoolQC(cualitativa)
74. Fence(cualitativa)
75. MiscFeature(cualitativa)
76. MiscVal(cuantitativa continua)
77. MoSold(cuantitativa discreta)
78. YrSold(cualitativa)
79. SaleType(cualitativa)
80. SaleCondition(cualitativa)
81. SalePrice(cuantitativa continua)

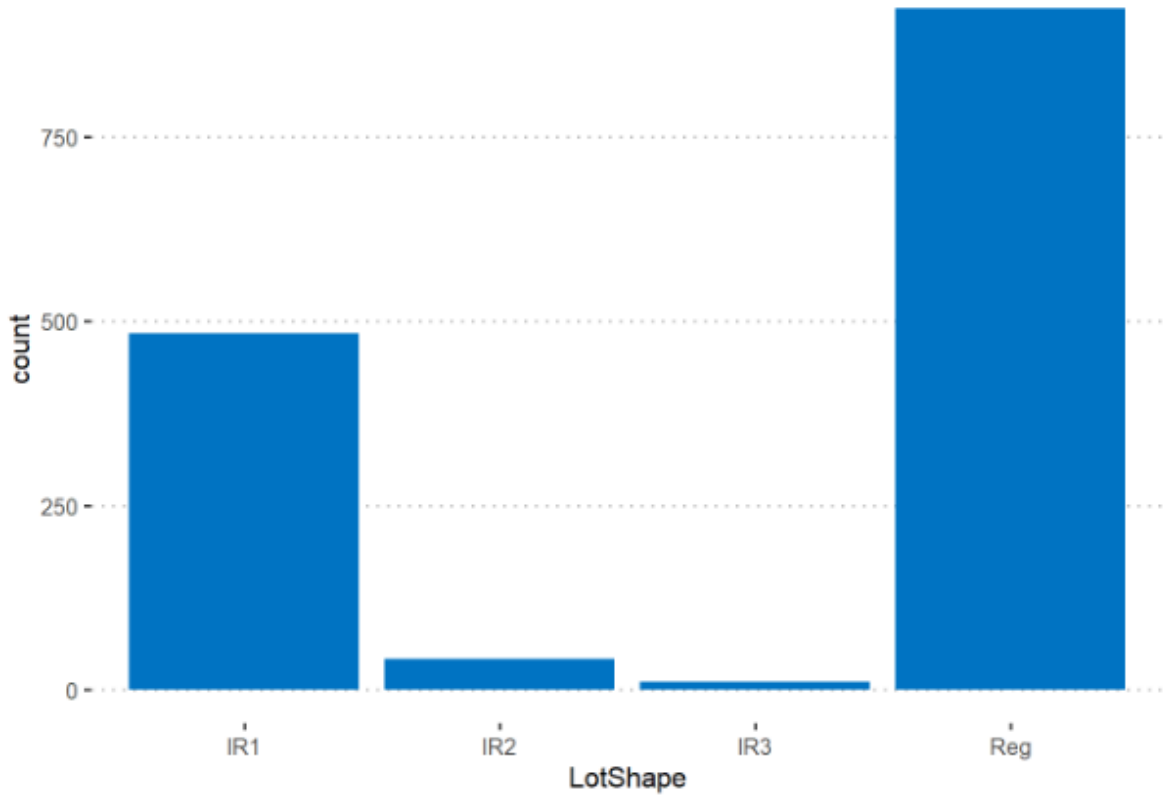
3. Incluya los gráficos exploratorios siendo consecuentes con el tipo de variable que están representando

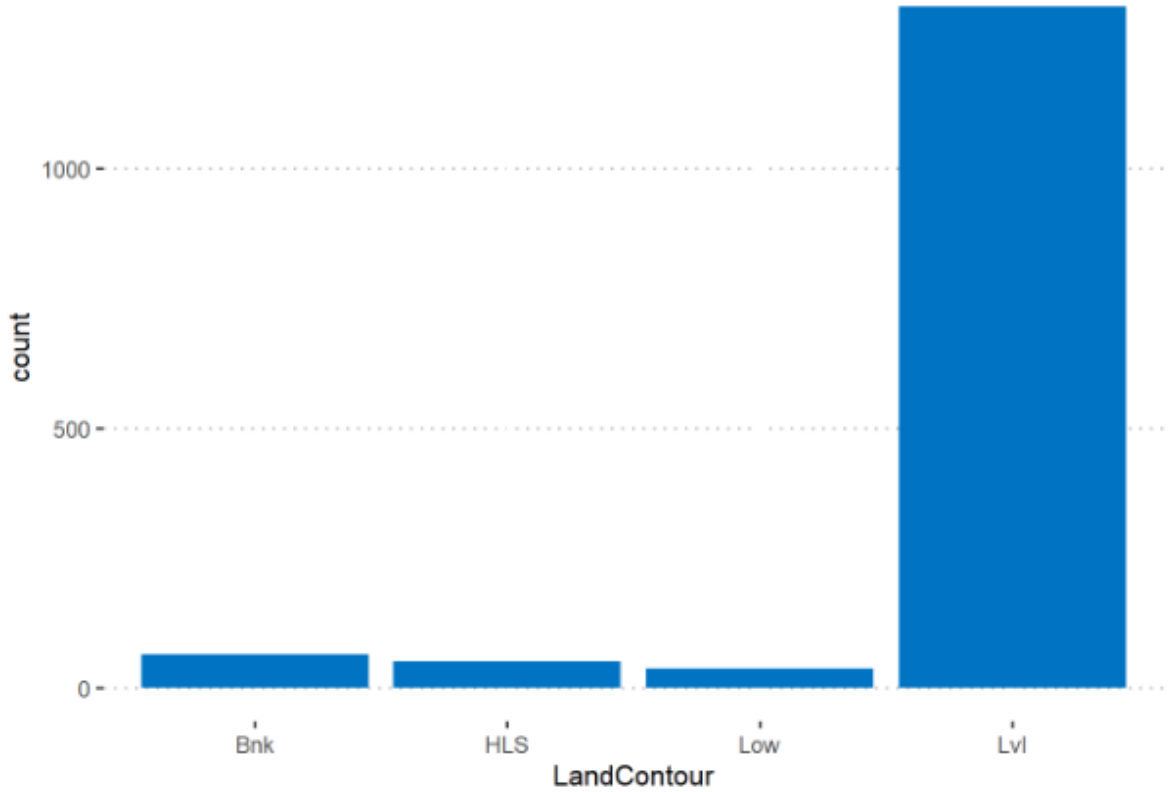
Para las variables cualitativas se hicieron las siguientes tablas de frecuencia:

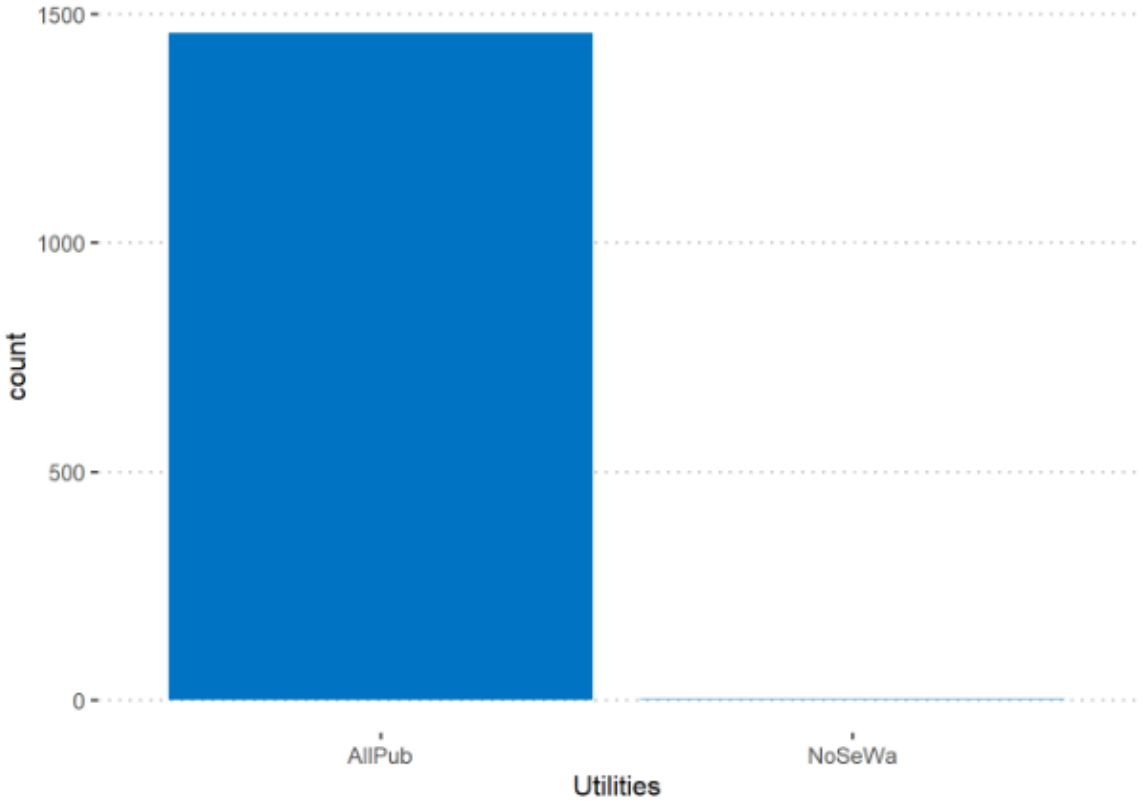


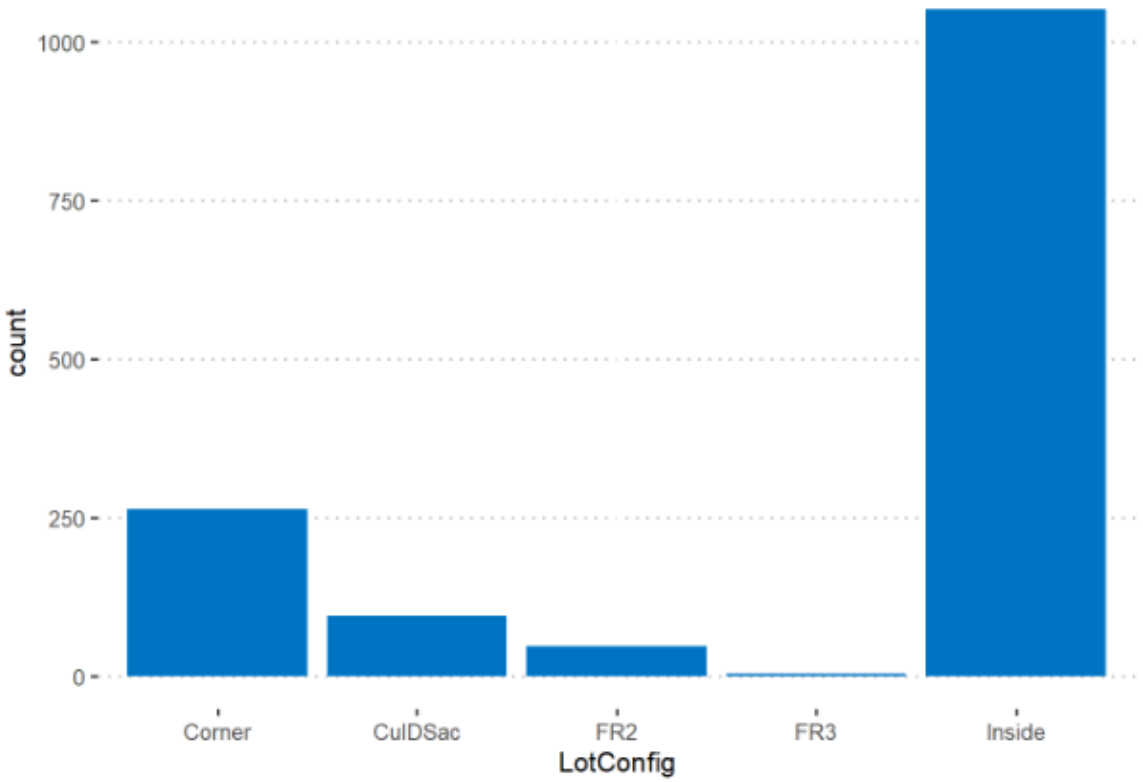

1500
1000
500
0
count
Grvl
Pave
Street

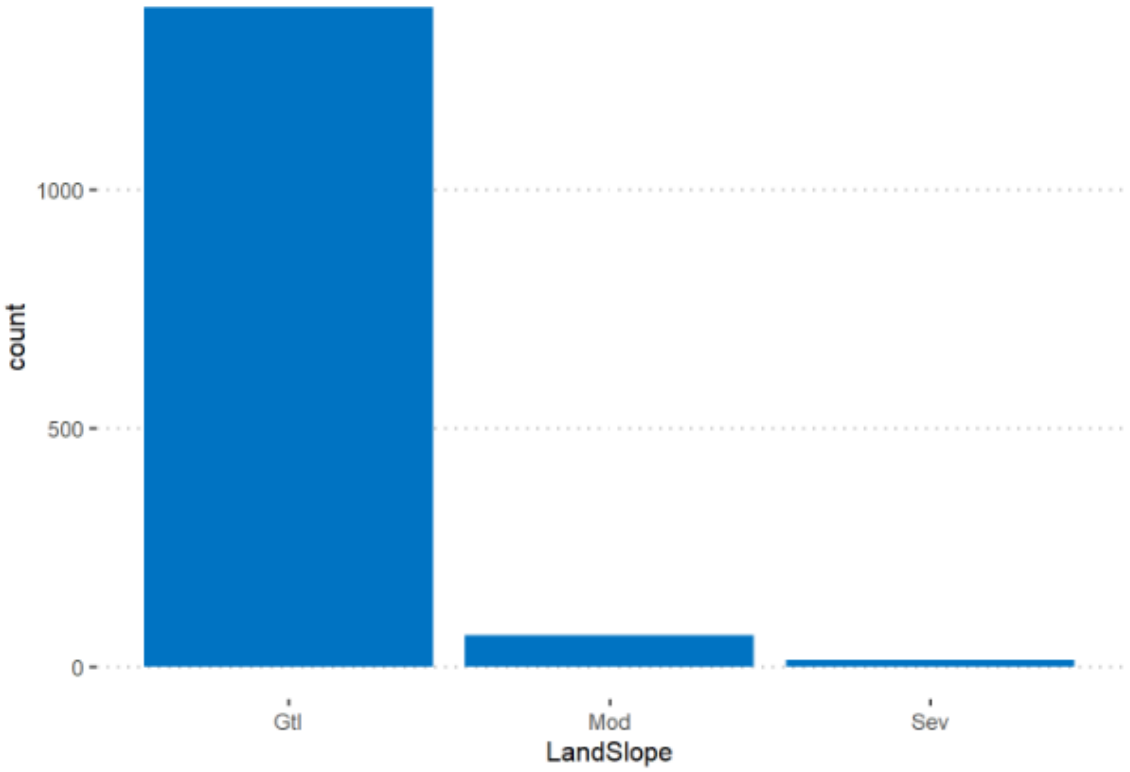

1000
500
0
count
Grvl
Pave
NA
Alley

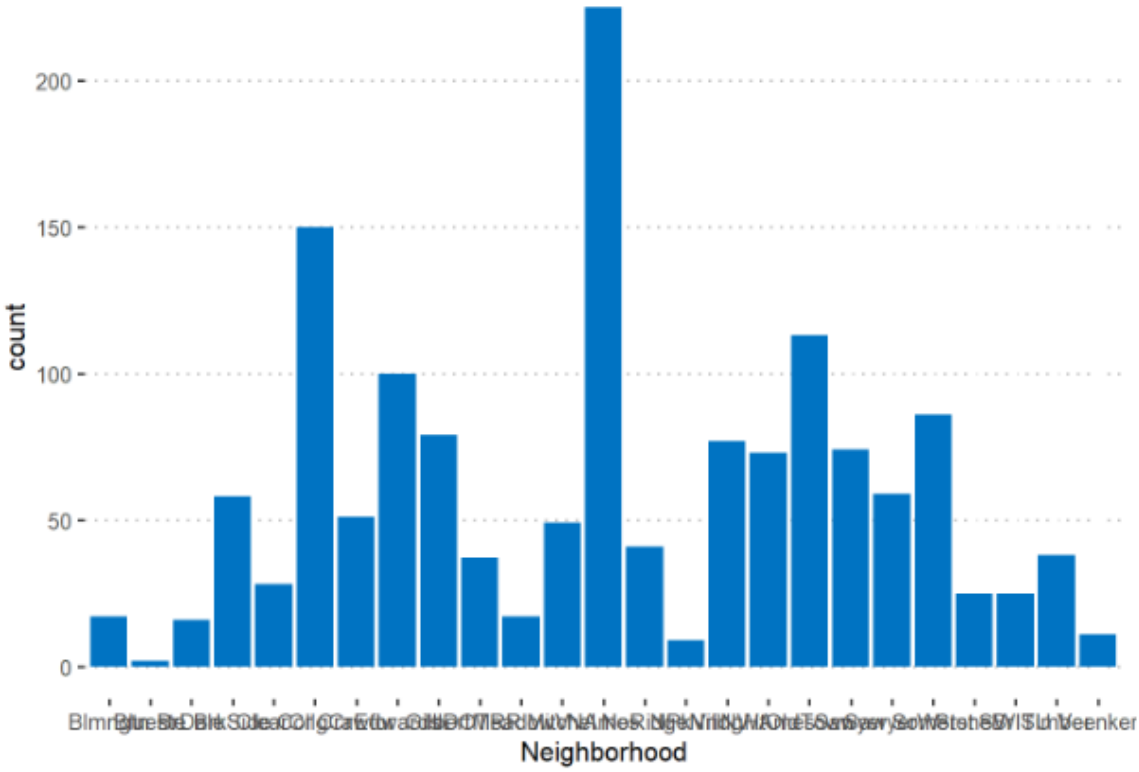

750
500
250
0
count
IR1
IR2
IR3
Reg
LotShape

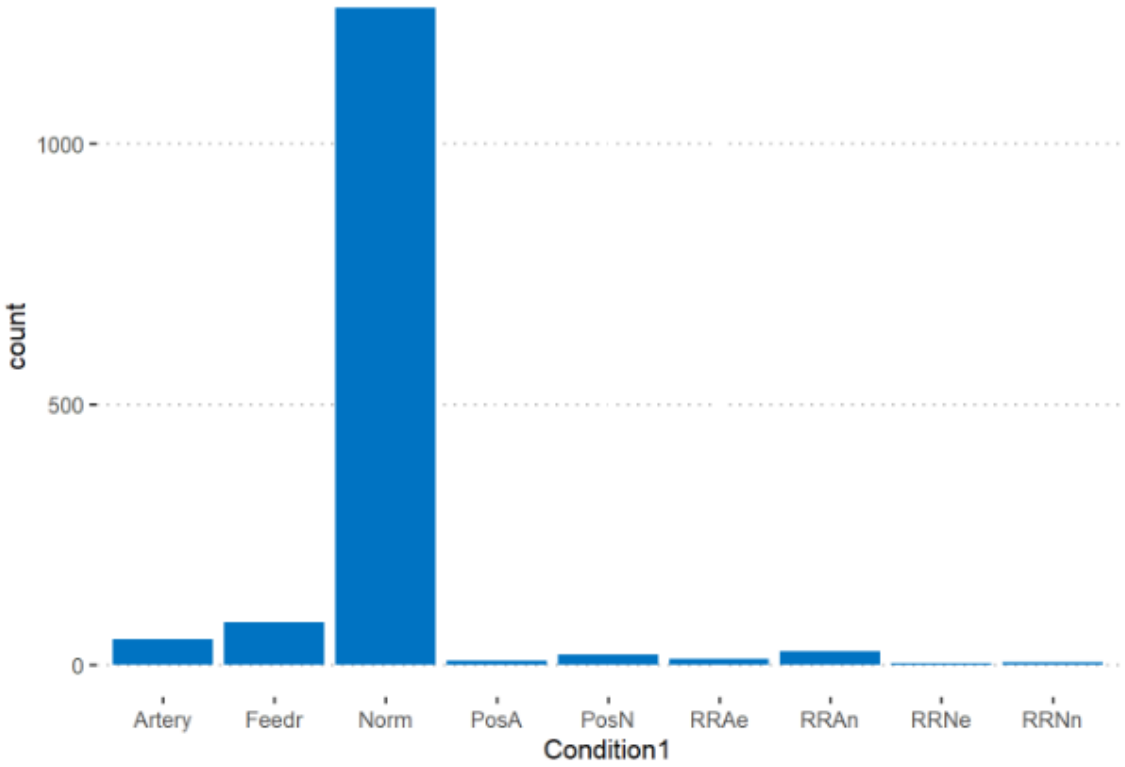

1000
500
0
count
Bnk
HLS
Low
Lvl
LandContour

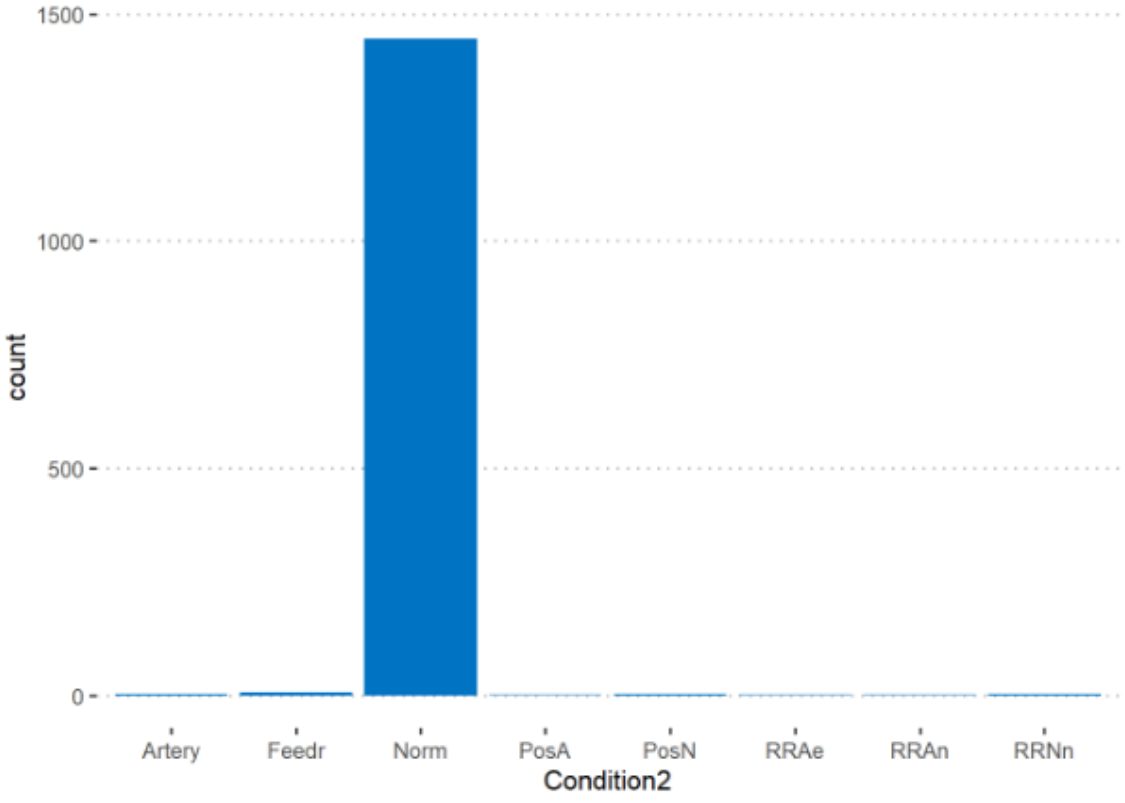

1500
1000
500
0
count
AllPub
NoSeWa
Utilities

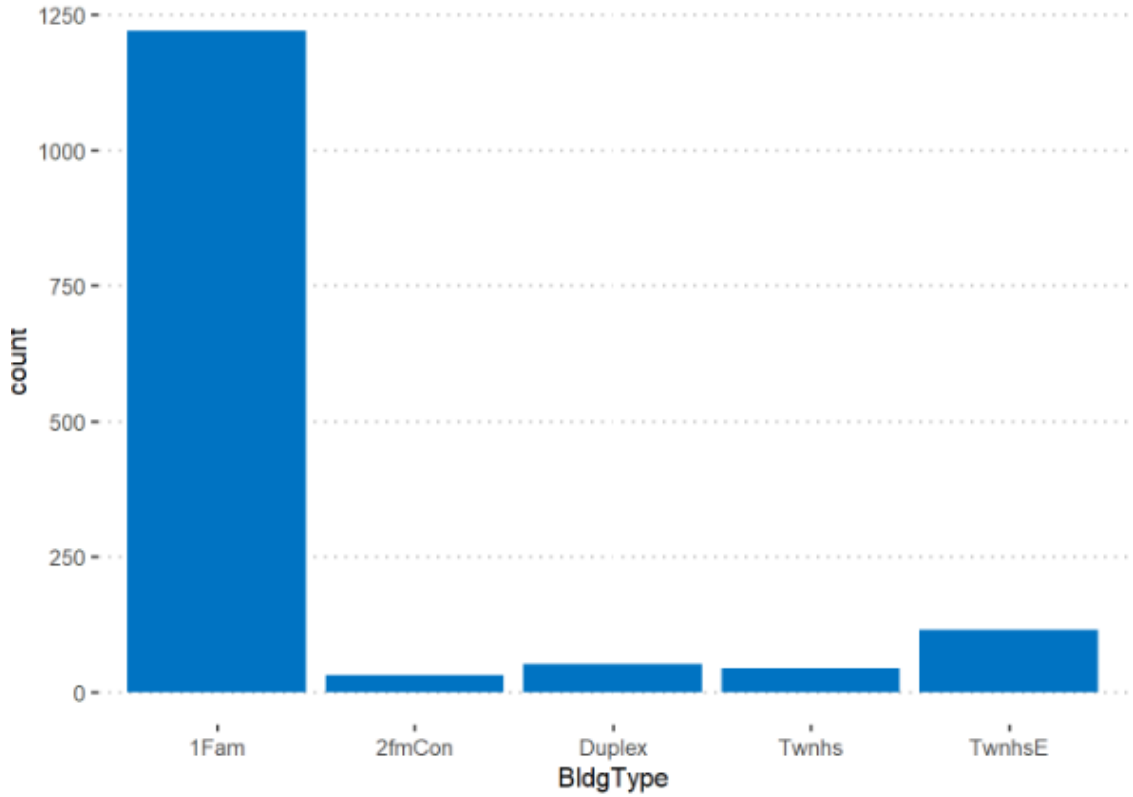

1000
750
500
250
0
count
Corner
CulDSac
FR2
FR3
Inside
LotConfig

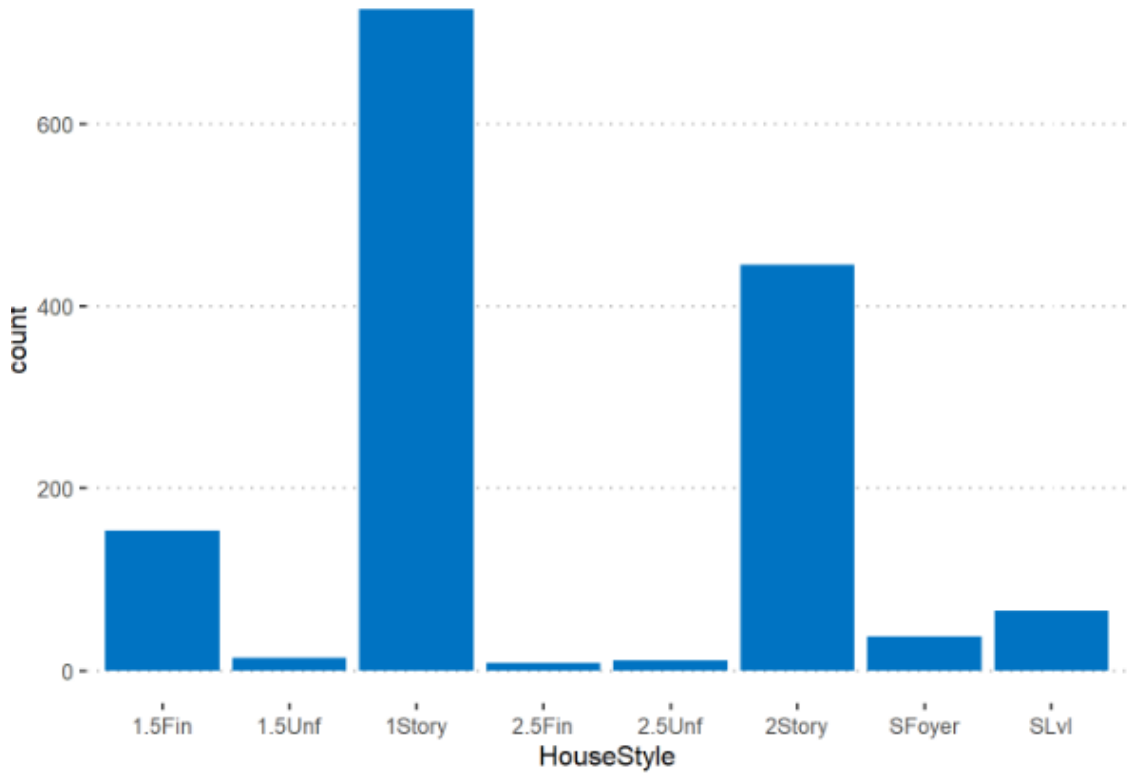

count
1000
500
0
Gtl
Mod
Sev
LandSlope

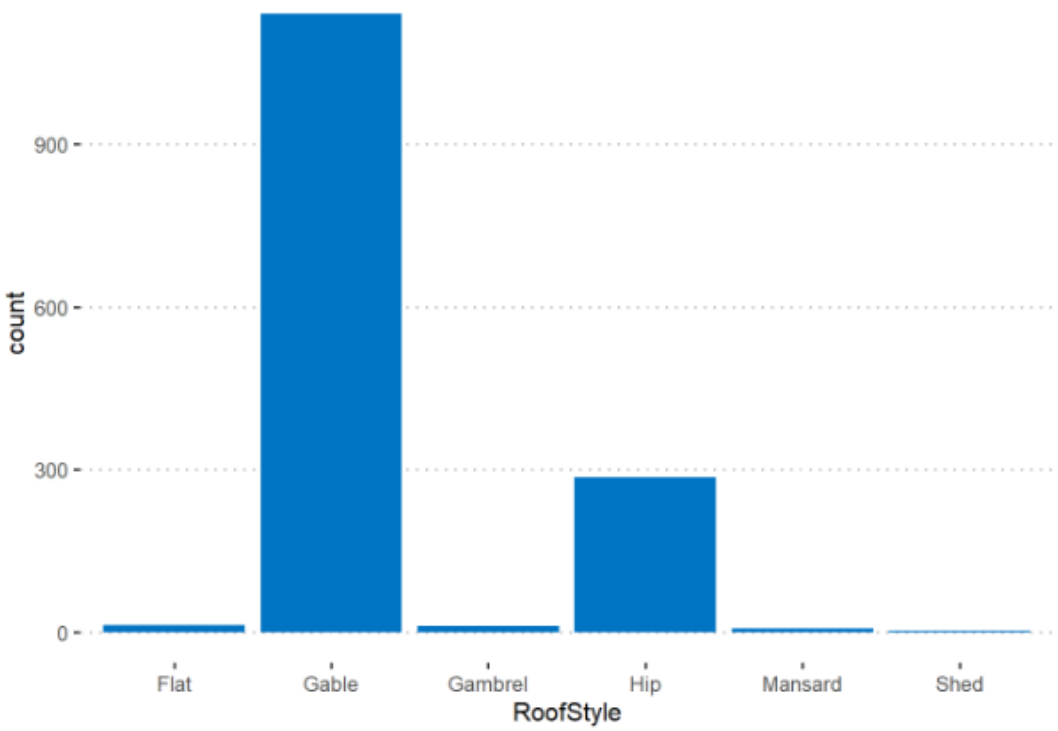

count
200
150
100
50
0
Neighborhood

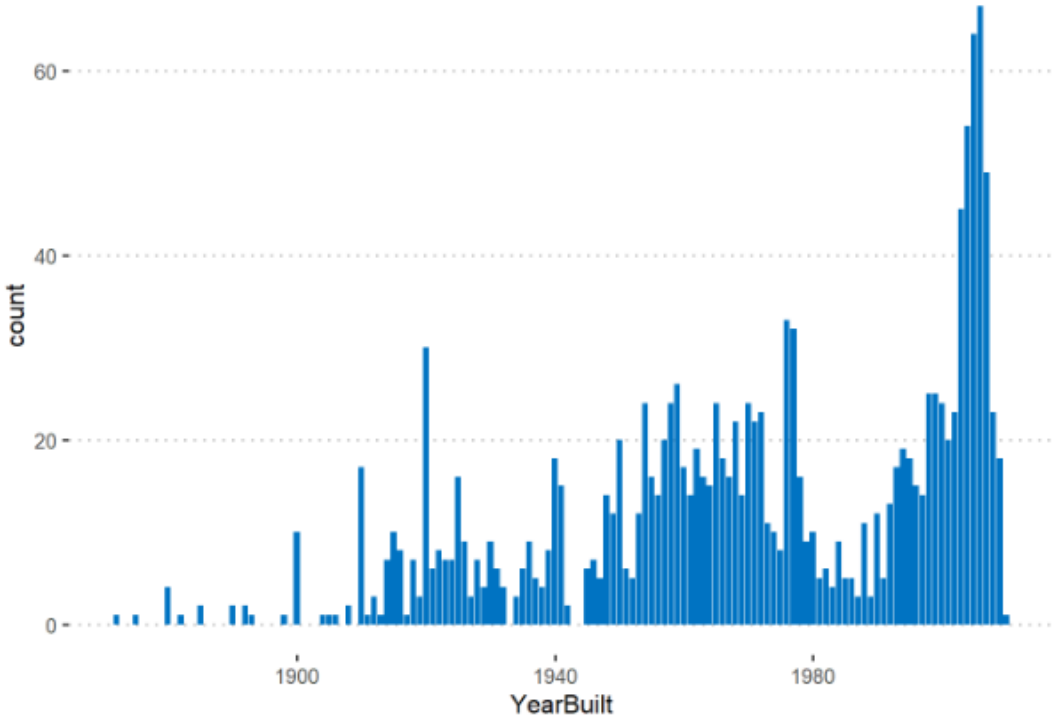

1000
500
0
count
Artery
Feedr
Norm
PosA
PosN
RRAe
RRAn
RRNe
RRNn
Condition1


1500
1000
500
0
count
Artery
Feedr
Norm
PosA
PosN
RRAe
RRAn
RRNn
Condition2

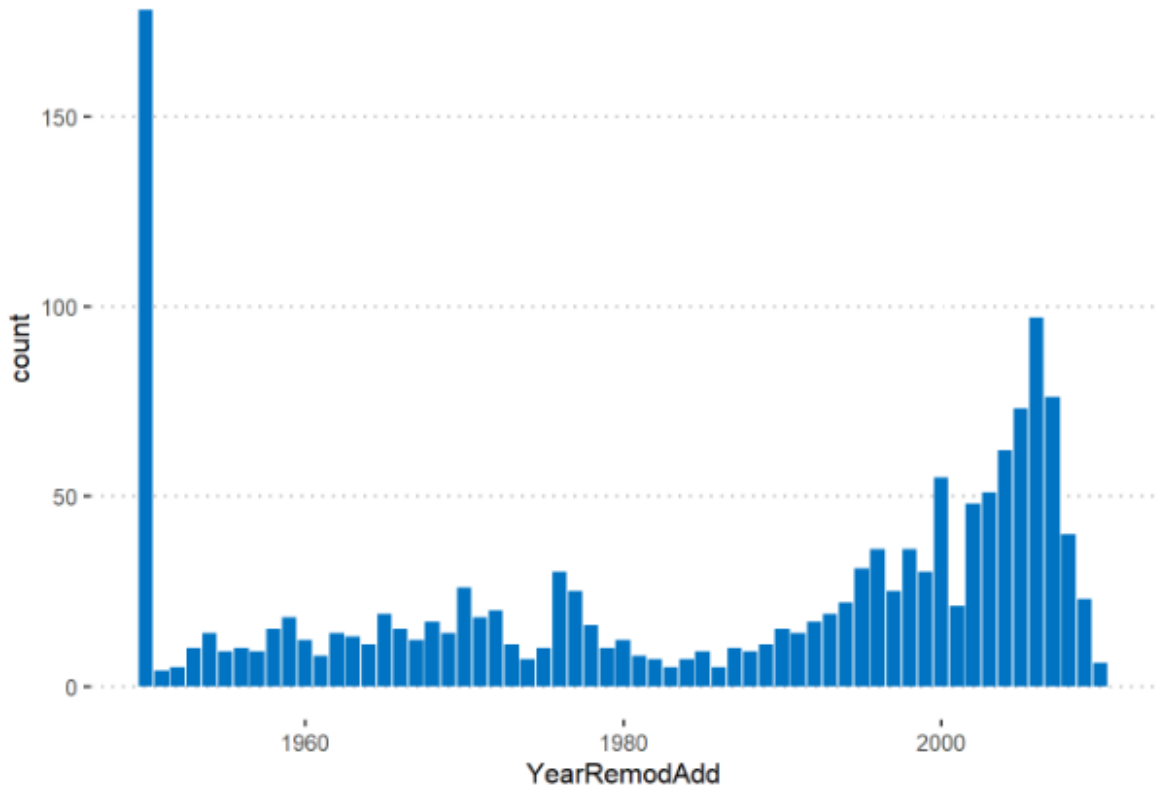

1250
1000
750
500
250
0
count
1Fam
2fmCon
Duplex
Twnhs
TwnhsE
BldgType

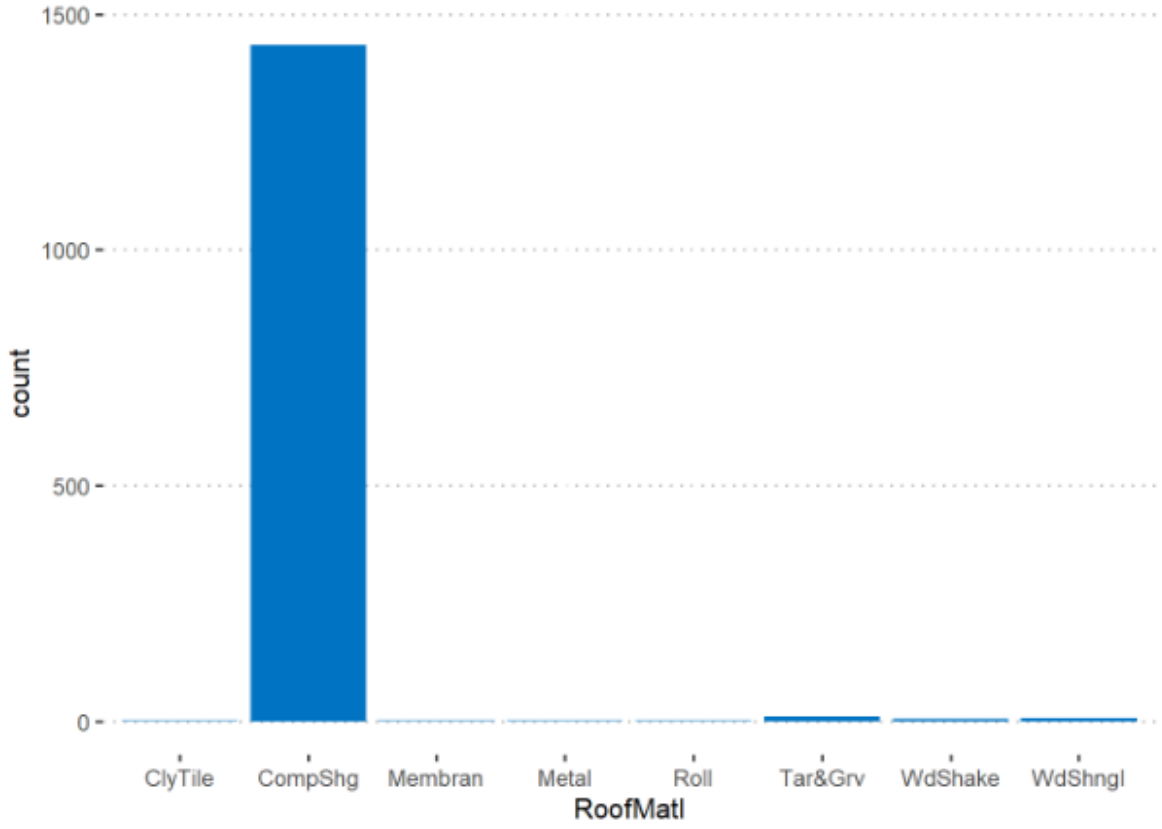

600
400
200
0
count
1.5Fin
1.5Unf
1Story
2.5Fin
2.5Unf
2Story
SFoyer
SLvl
HouseStyle

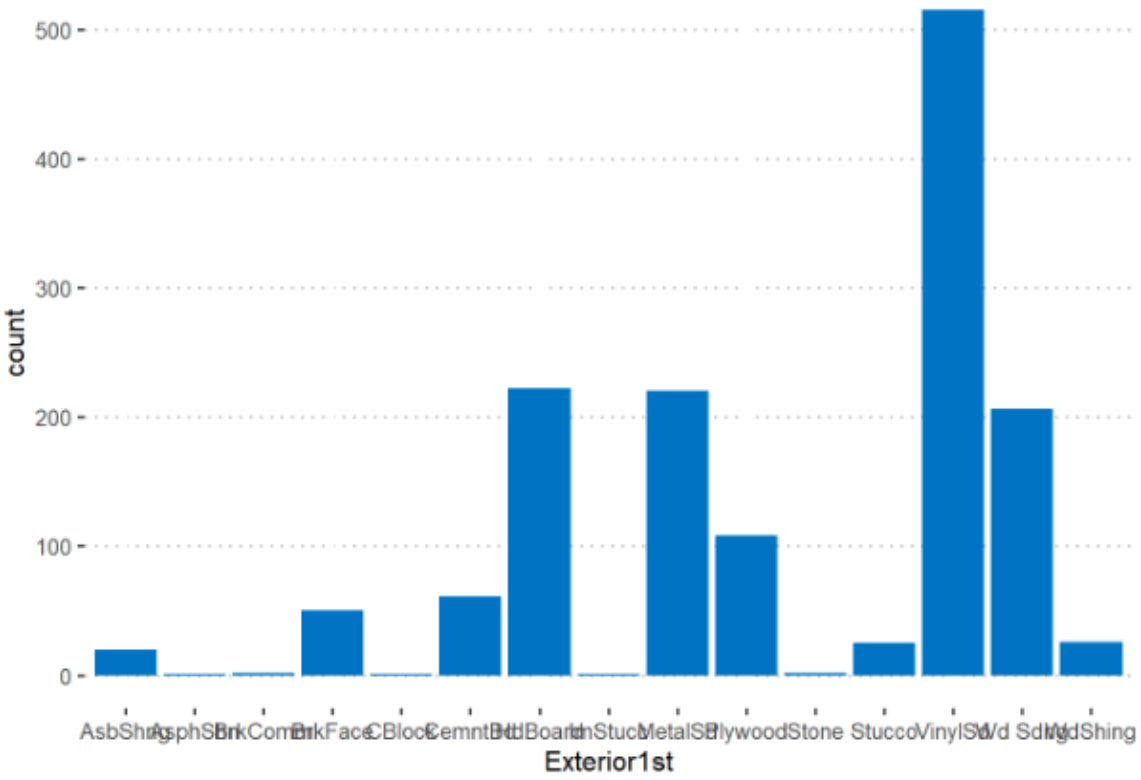


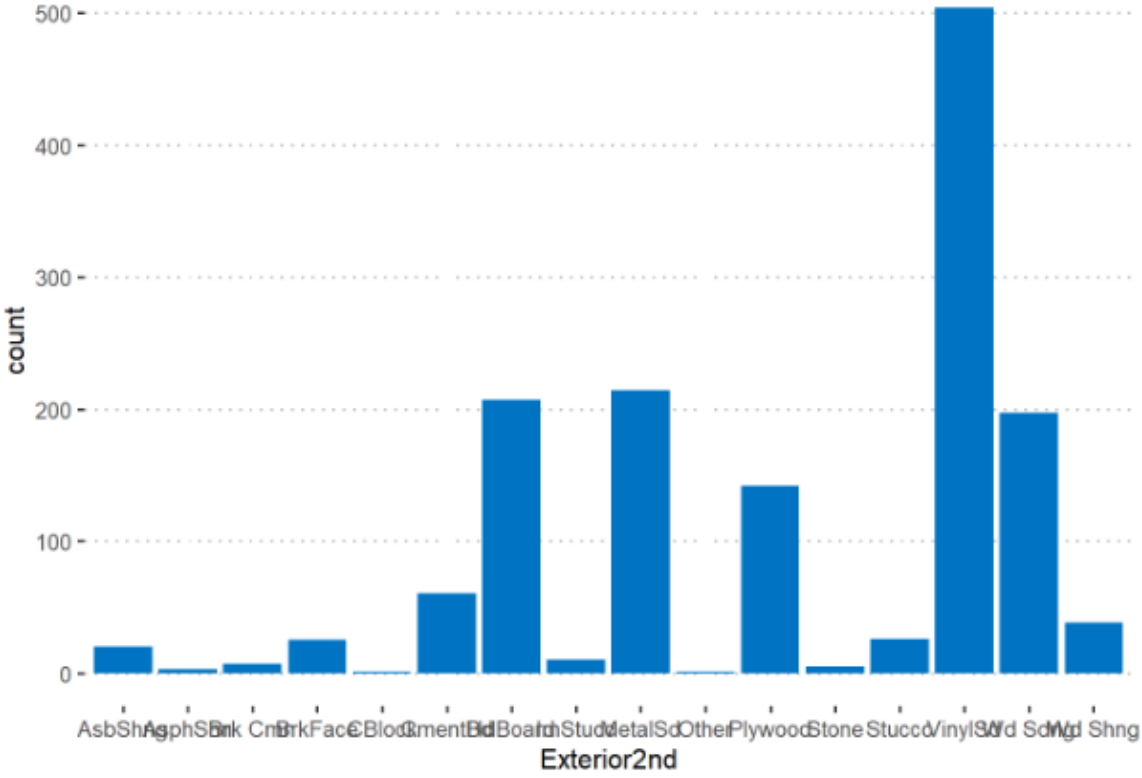


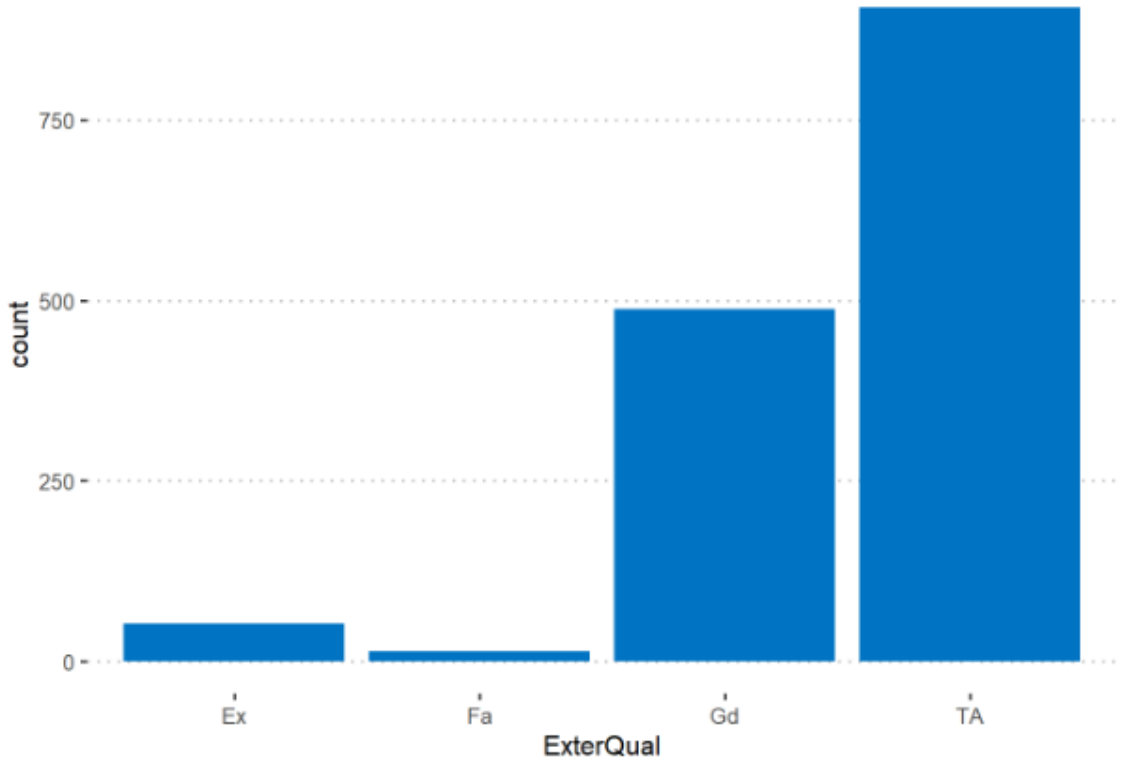
En el año de construcción (YearBuilt) se observa un claro crecimiento en los años recientes, probablemente por el crecimiento poblacional.

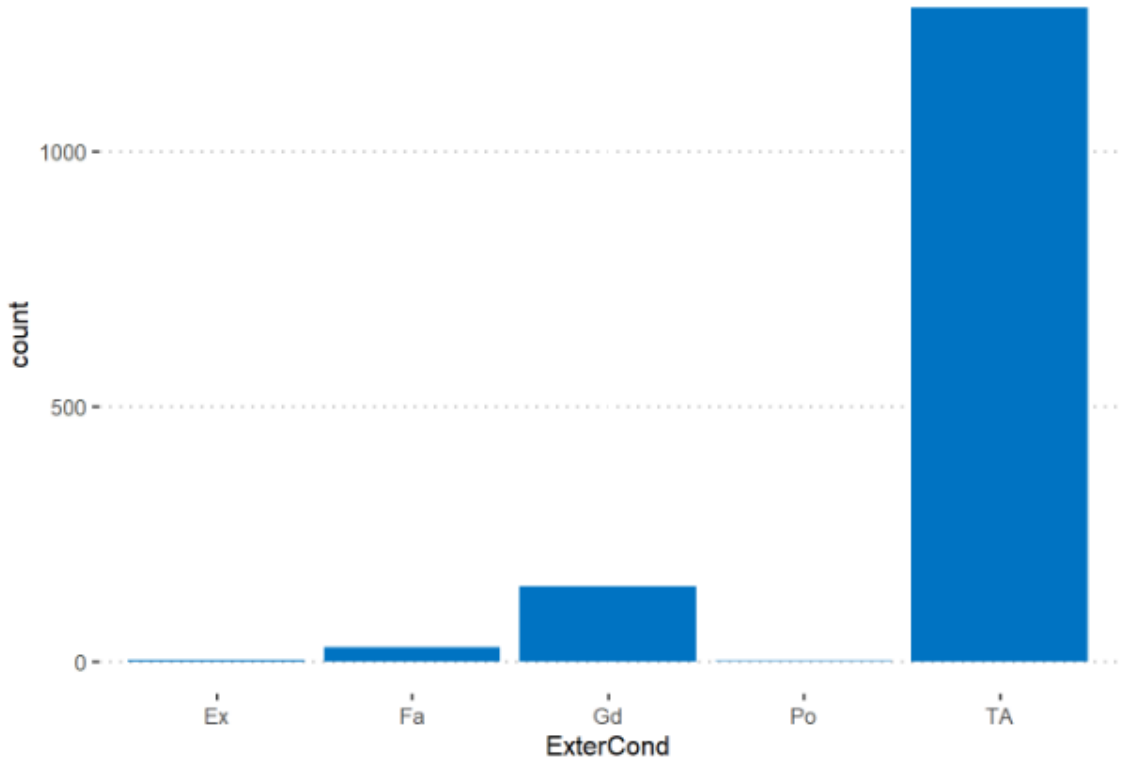

150
100
50
0
count
1960
1980
2000
YearRemodAdd

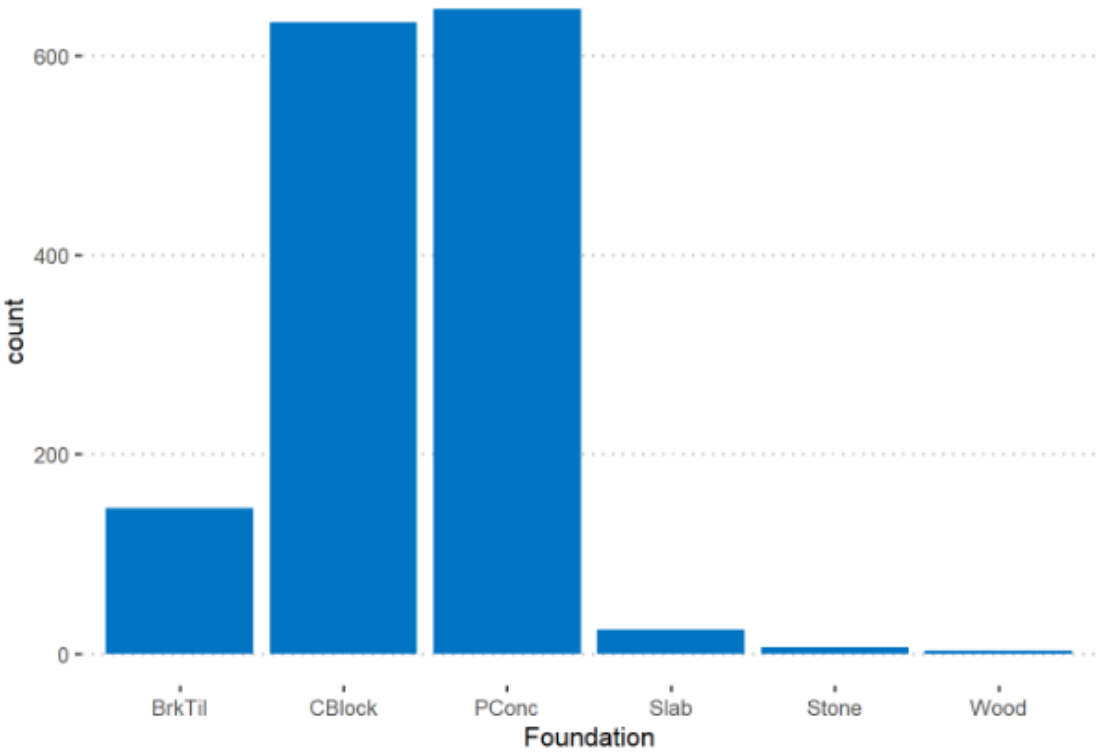

1500
1000
500
0
count
ClyTile
CompShg
Membran
Metal
Roll
Tar&Grv
WdShake
WdShngl
RoofMatl

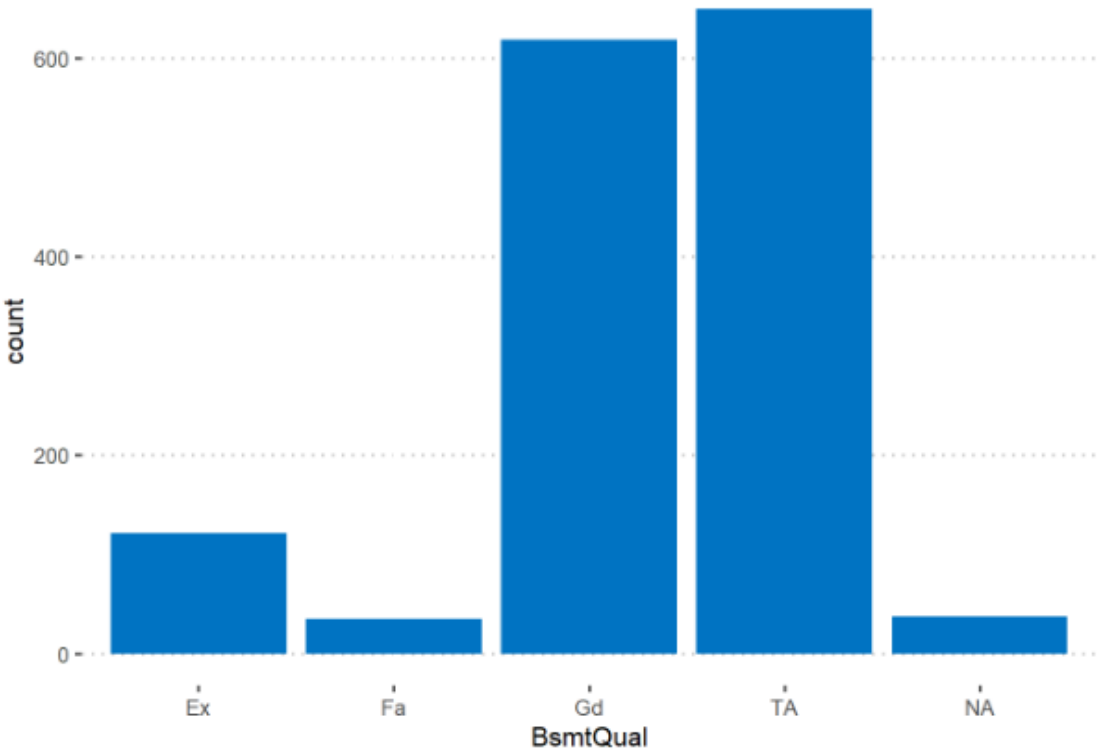

500
400
300
200
100
0
count
Exterior1st

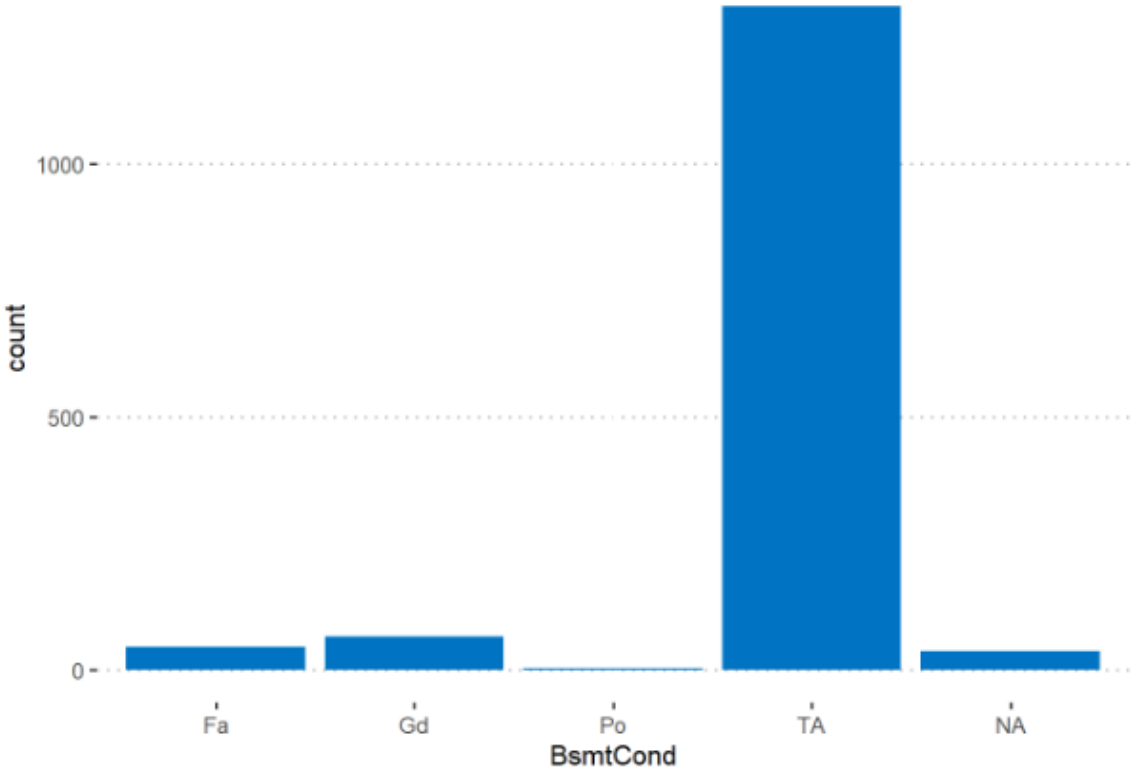

500
400
300
200
100
0
count
Exterior2nd

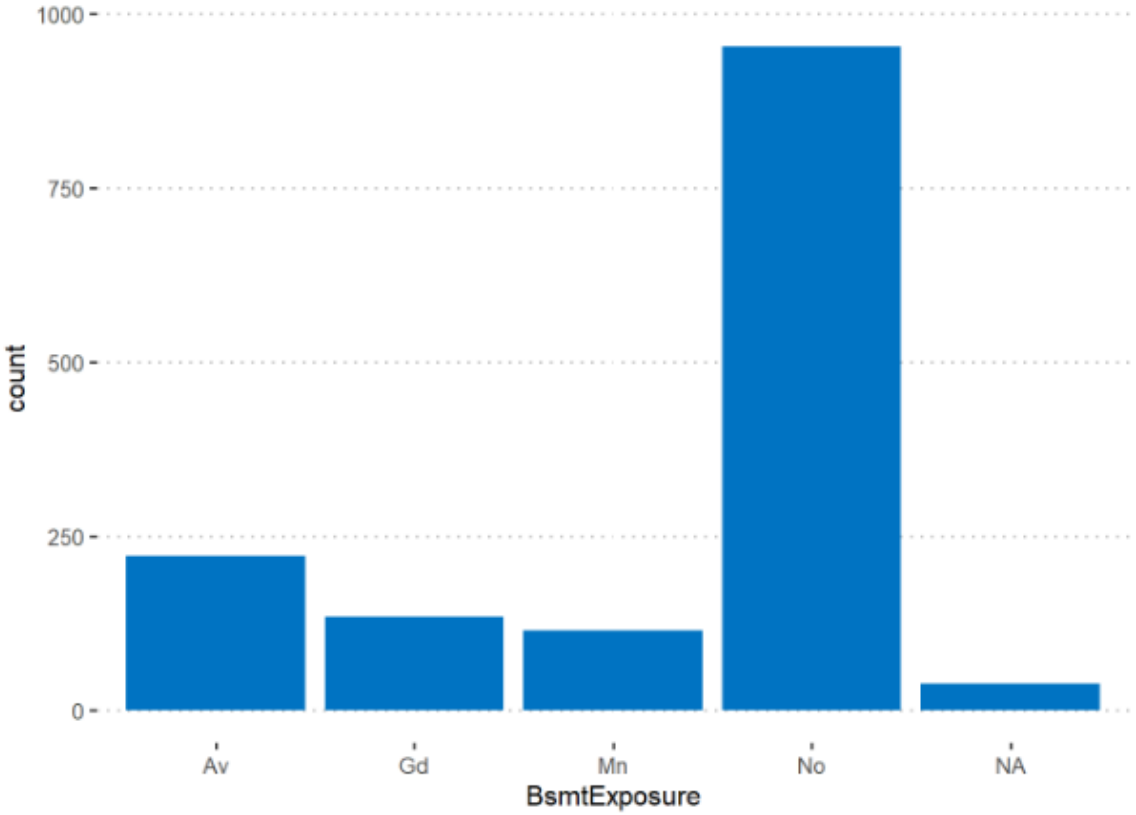

750
500
250
0
count
Ex
Fa
Gd
TA
ExterQual

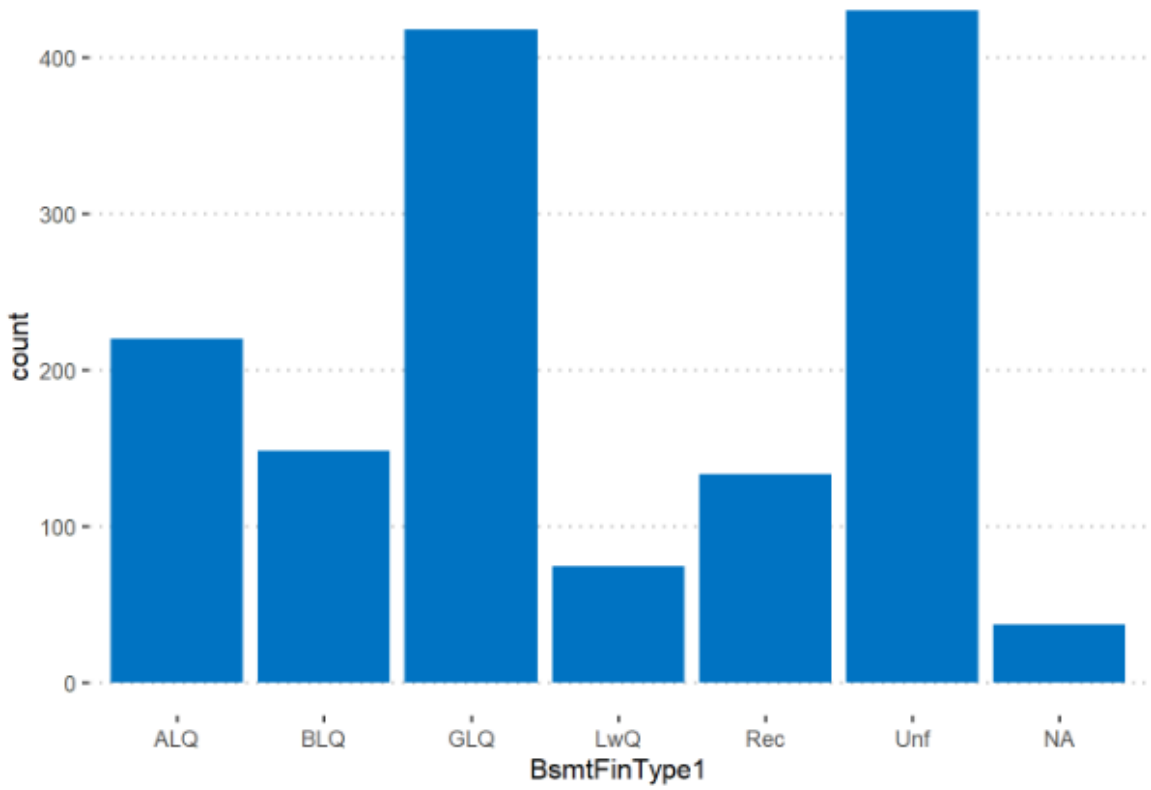

1000
500
0
count
Ex
Fa
Gd
Po
TA
ExterCond

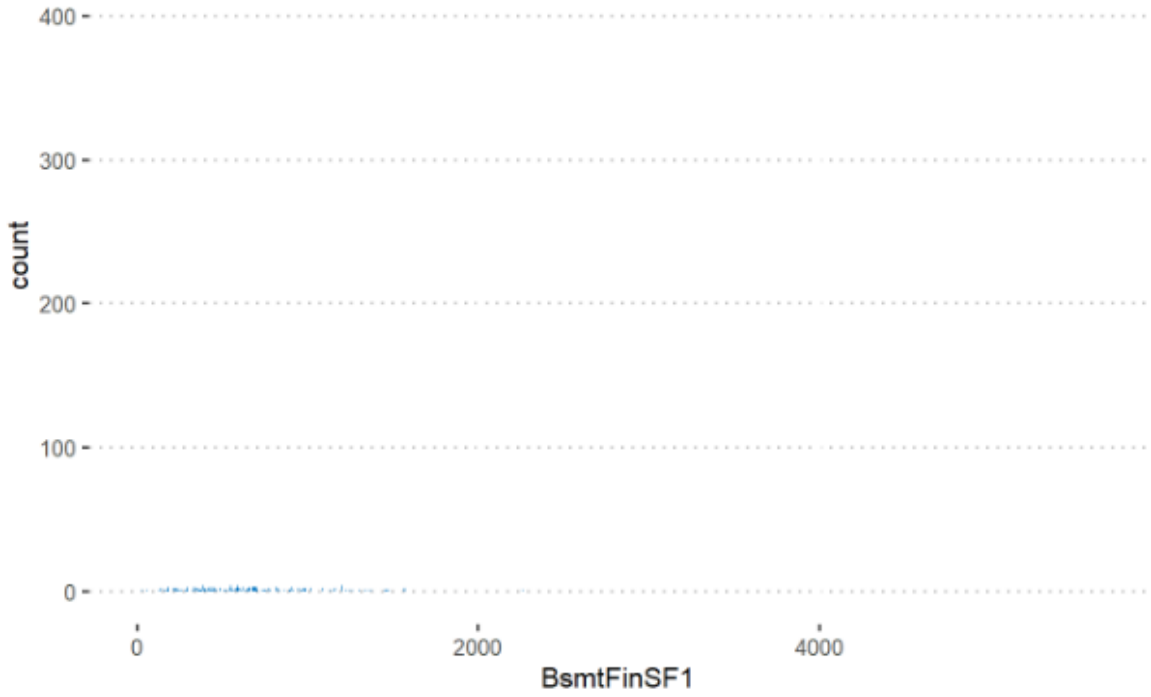

600
400
200
0
count
BrkTil
CBlock
PConc
Slab
Stone
Wood
Foundation

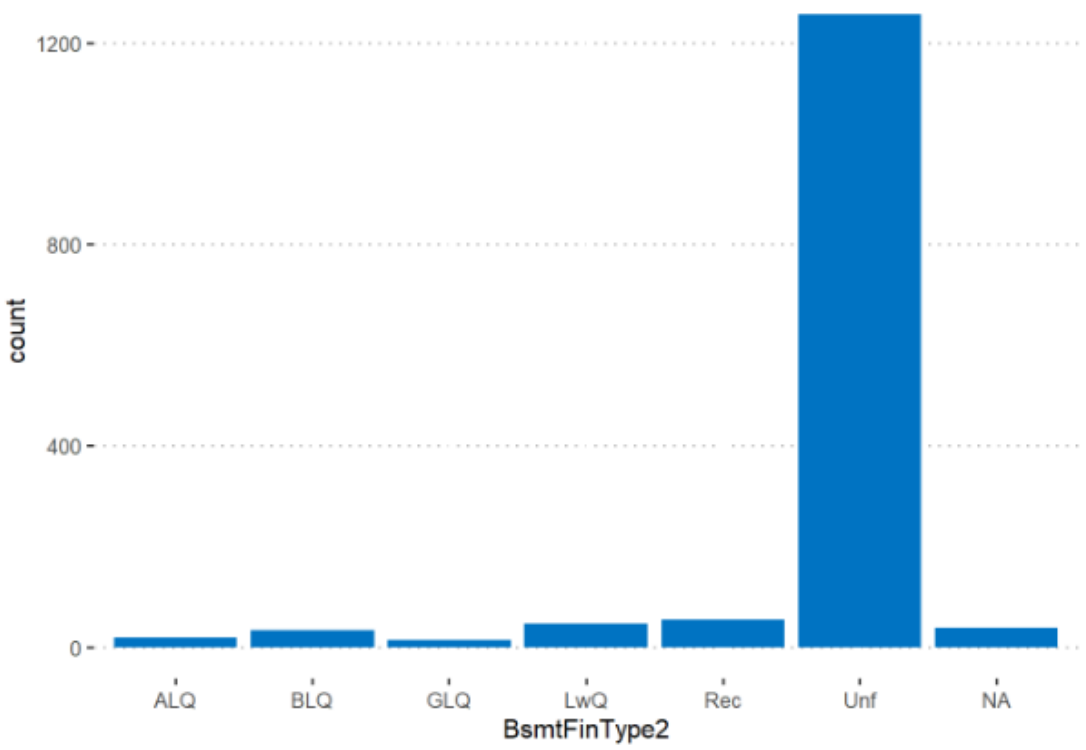

600
400
200
0
count
Ex
Fa
Gd
TA
NA
BsmtQual

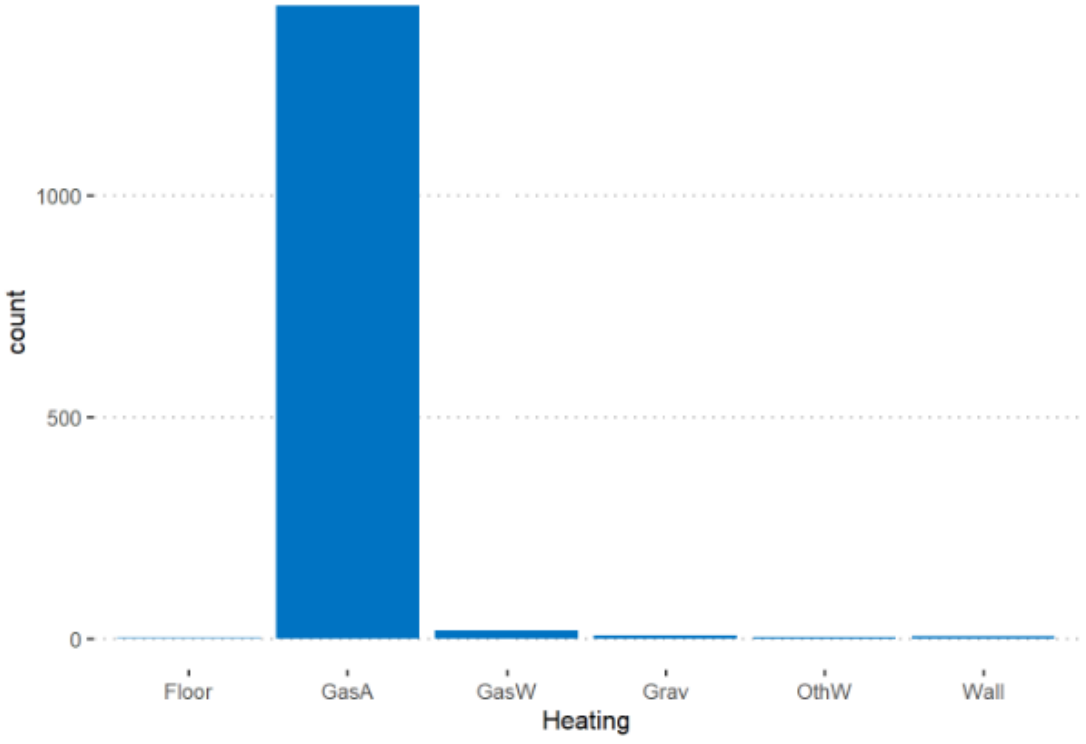

1000
500
0
count
Fa
Gd
Po
TA
NA
BsmtCond

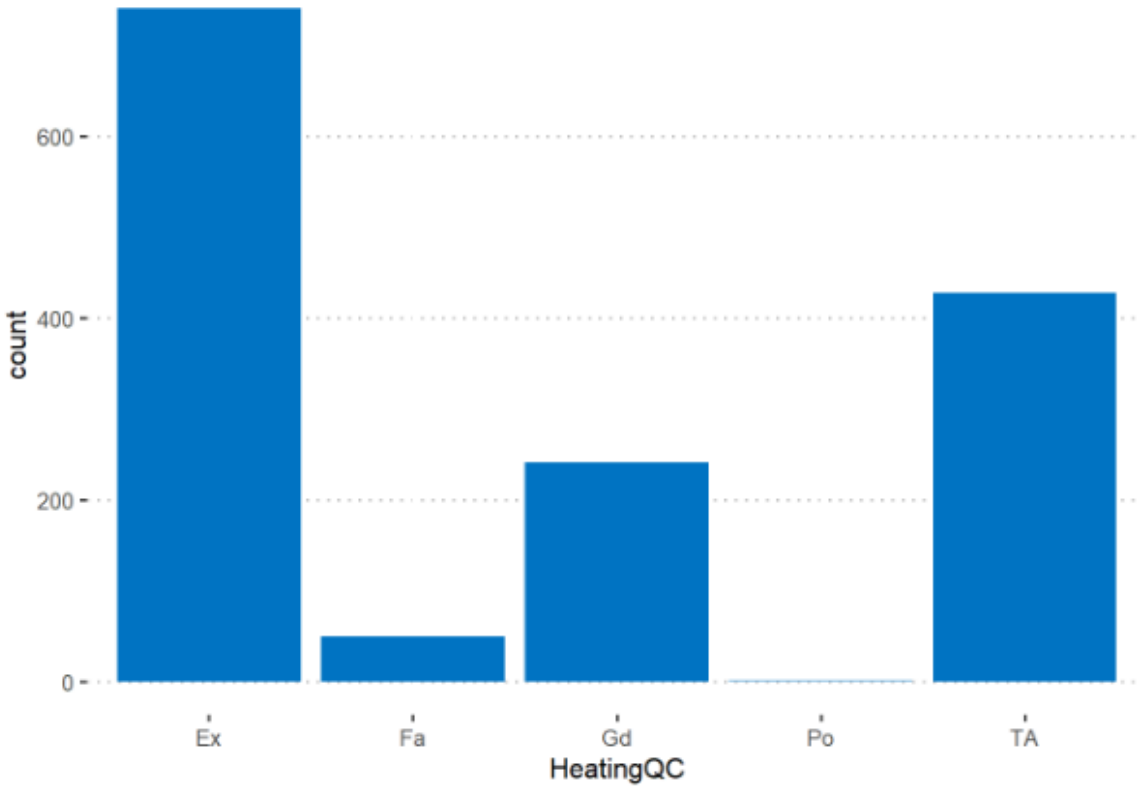

1000
750
500
250
0
count
Av
Gd
Mn
No
NA
BsmtExposure

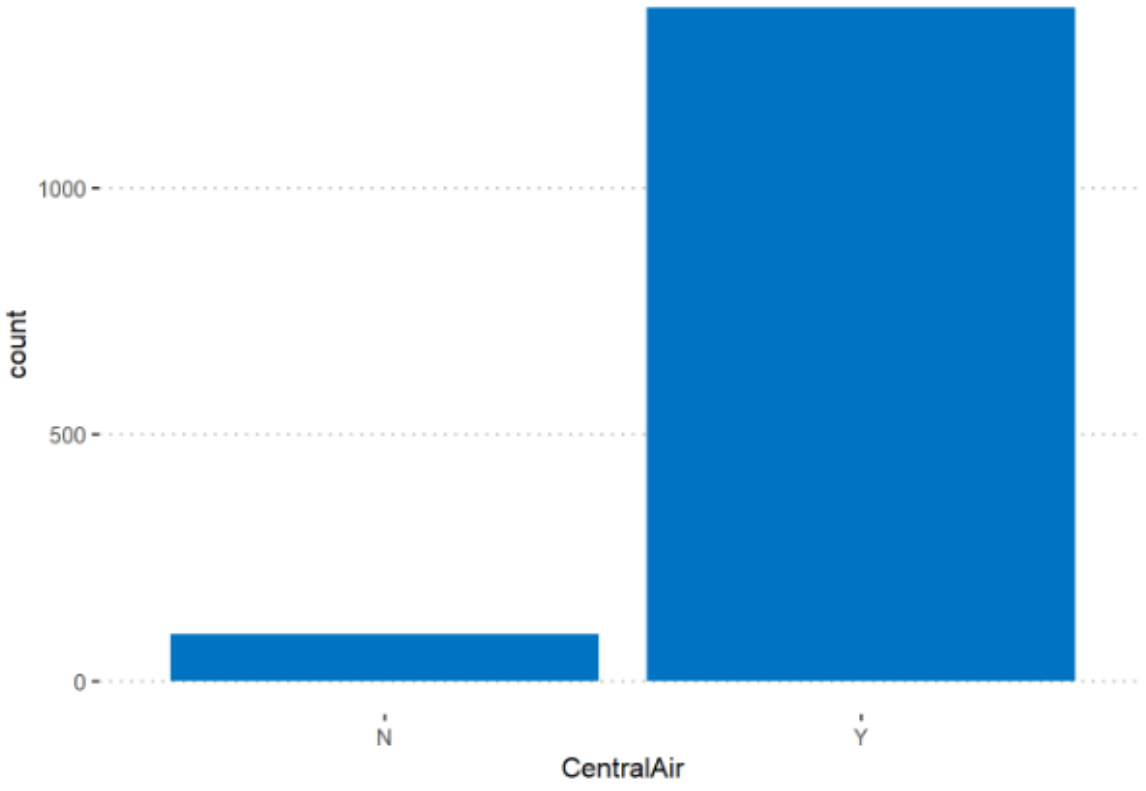

400
300
200
100
0
count
ALQ
BLQ
GLQ
LwQ
Rec
Unf
NA
BsmtFinType1

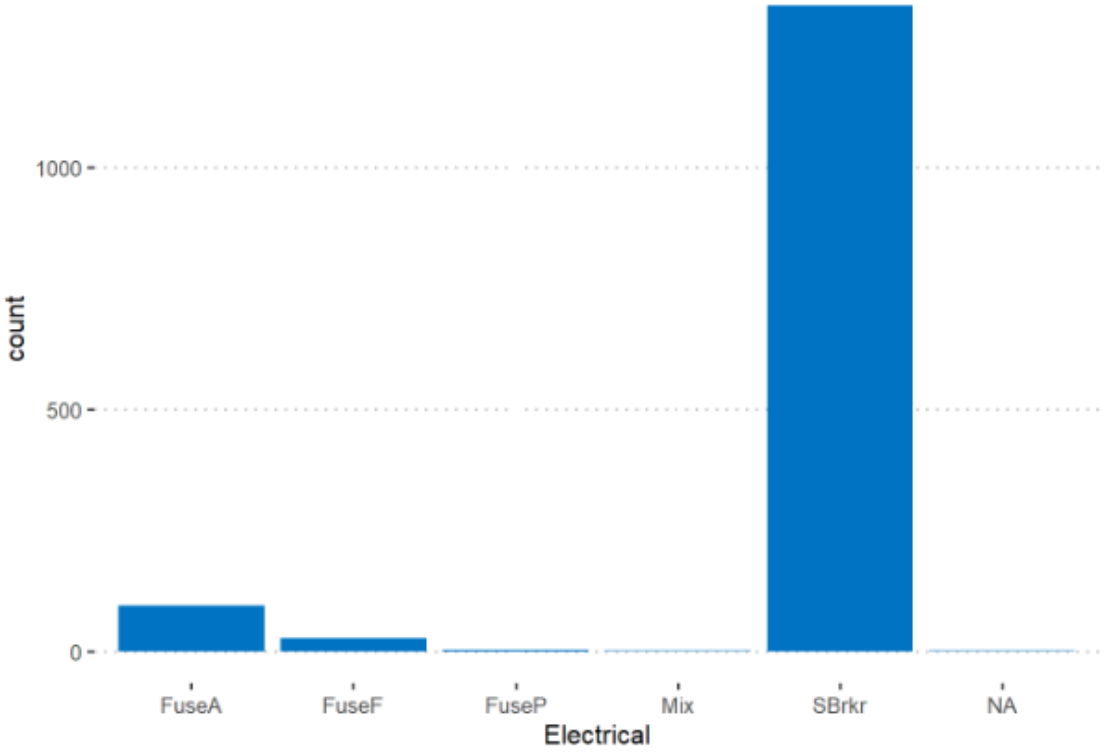

400
300
200
100
0
count
0
2000
4000
BsmtFinSF1

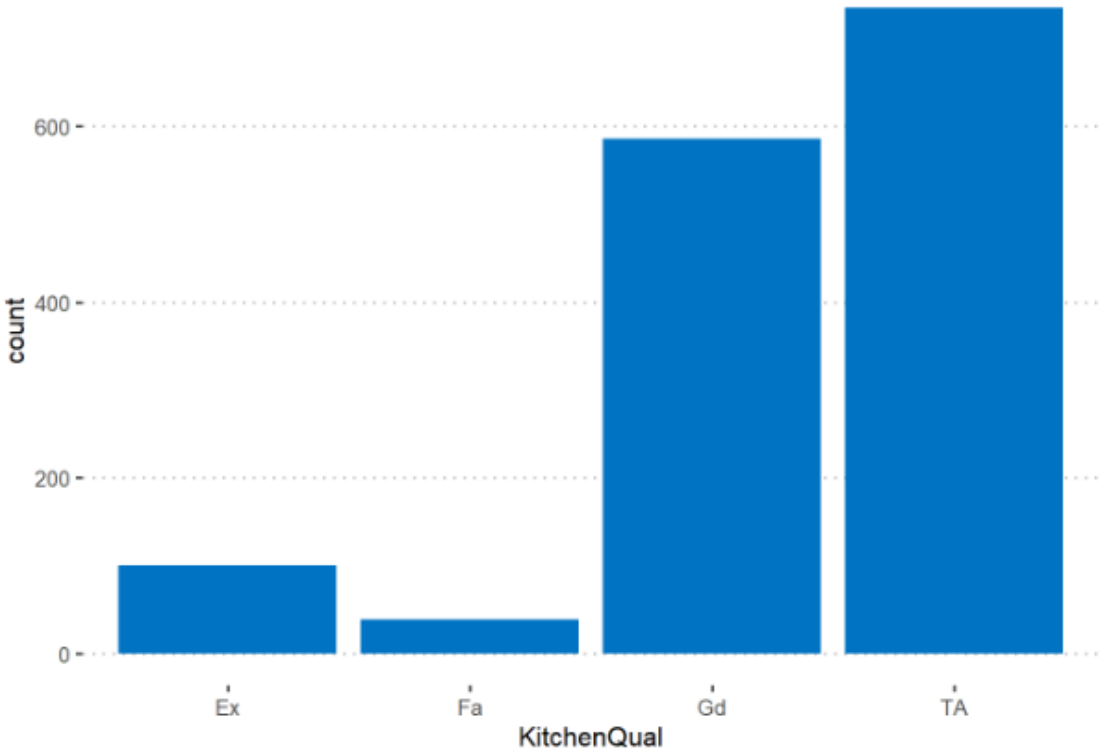

1200
800
400
0
count
ALQ
BLQ
GLQ
LwQ
Rec
Unf
NA
BsmtFinType2

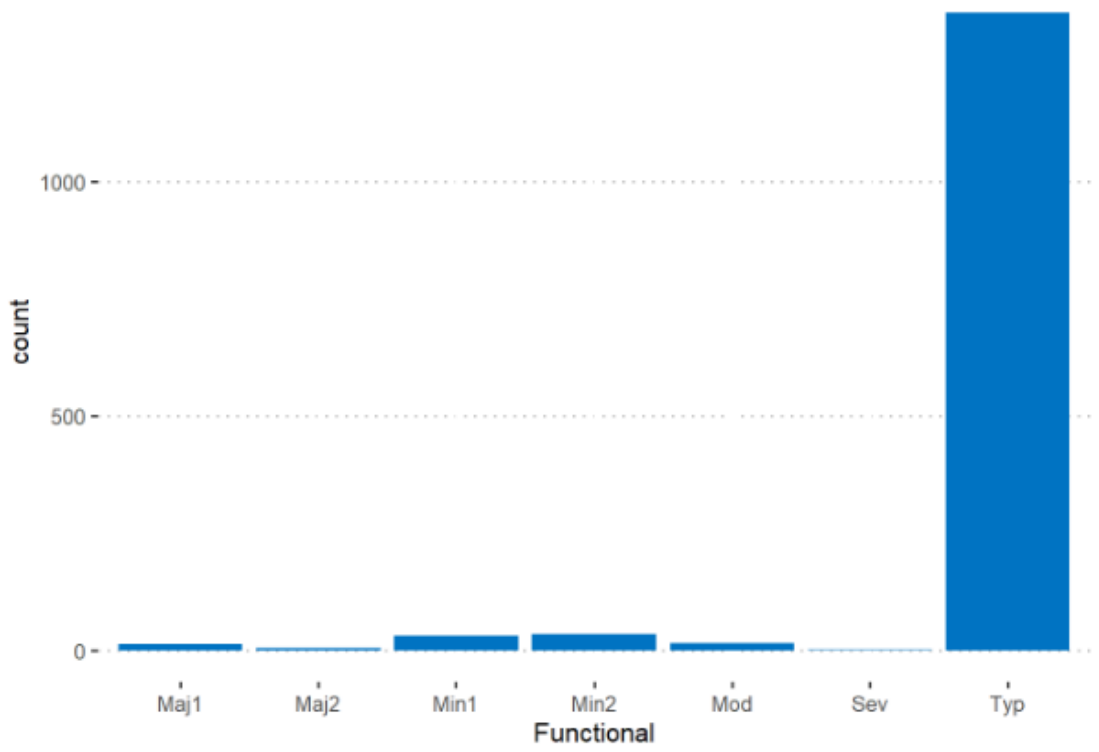

1000
500
0
count
Floor
GasA
GasW
Grav
OthW
Wall
Heating

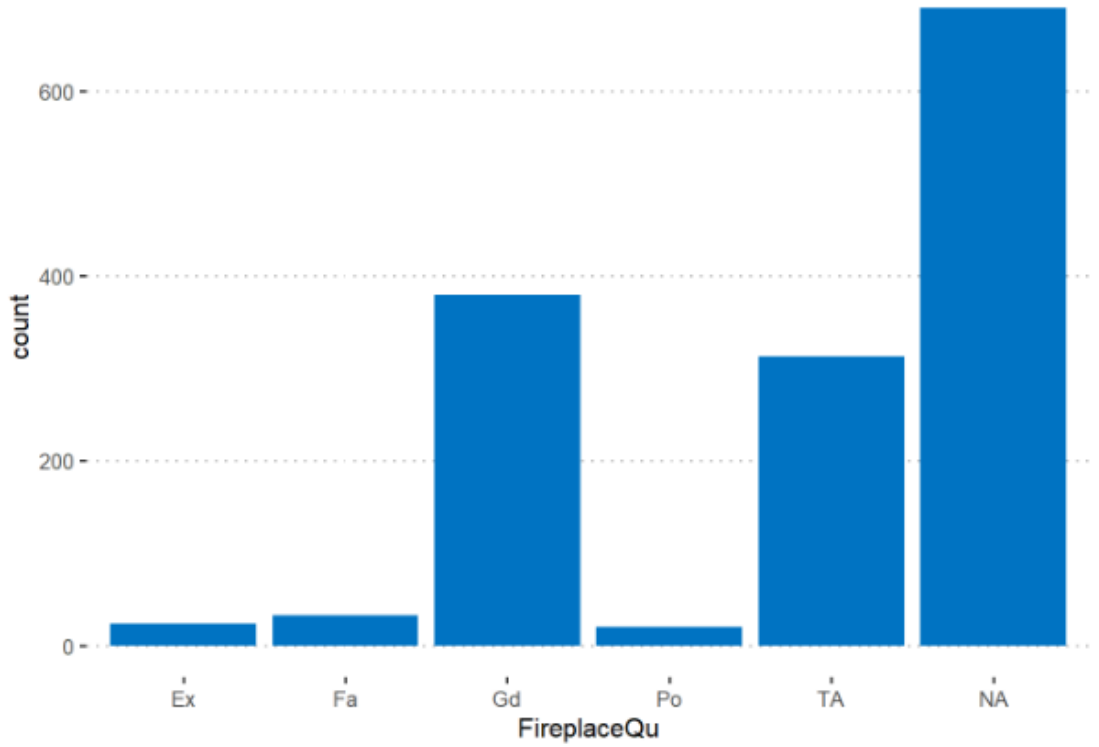

600
400
200
0
count
Ex
Fa
Gd
Po
TA
HeatingQC

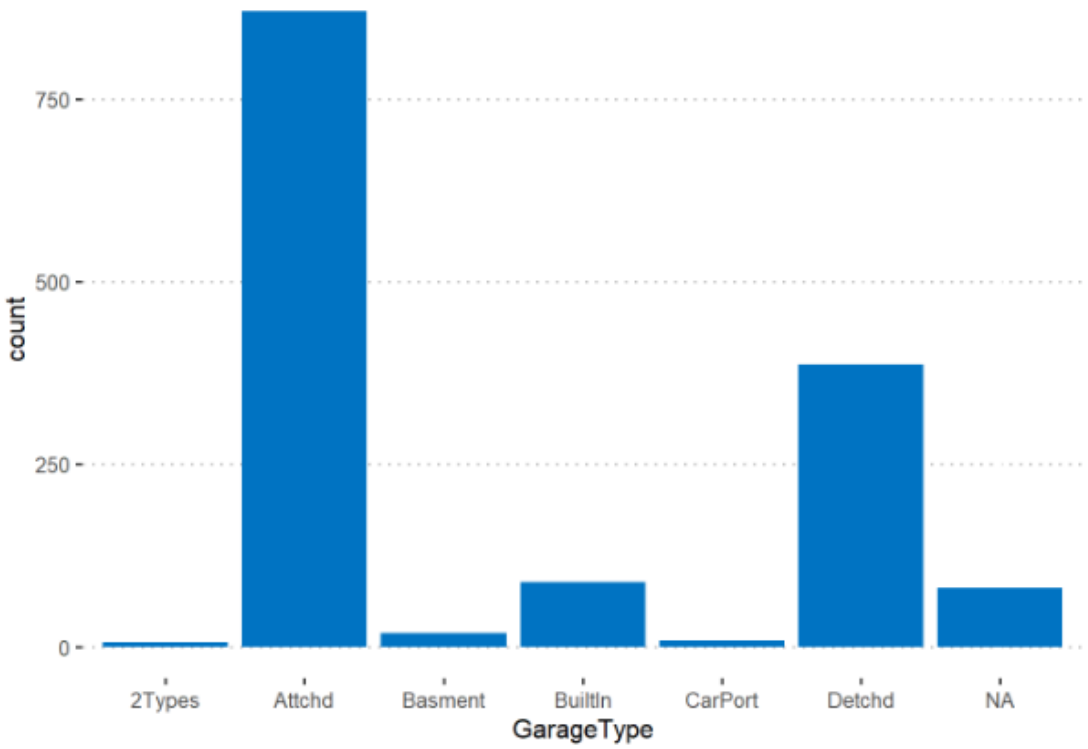

1000
500
0
count
N
Y
CentralAir

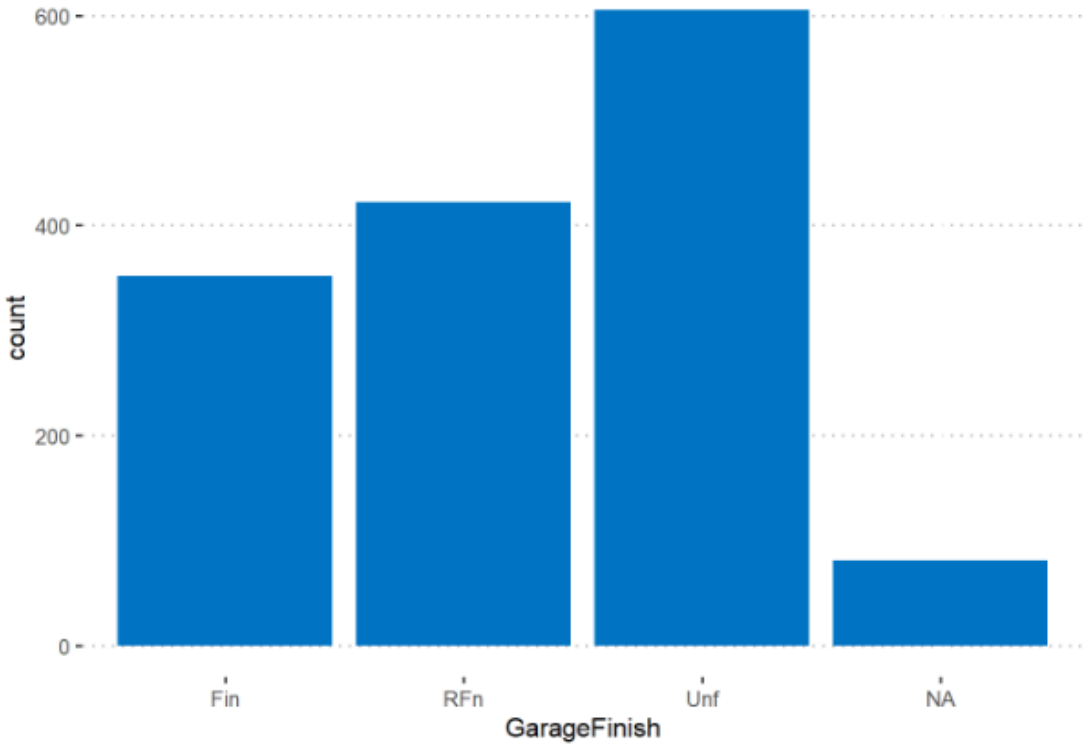

1000
500
0
count
FuseA
FuseF
FuseP
Mix
SBrkr
NA
Electrical

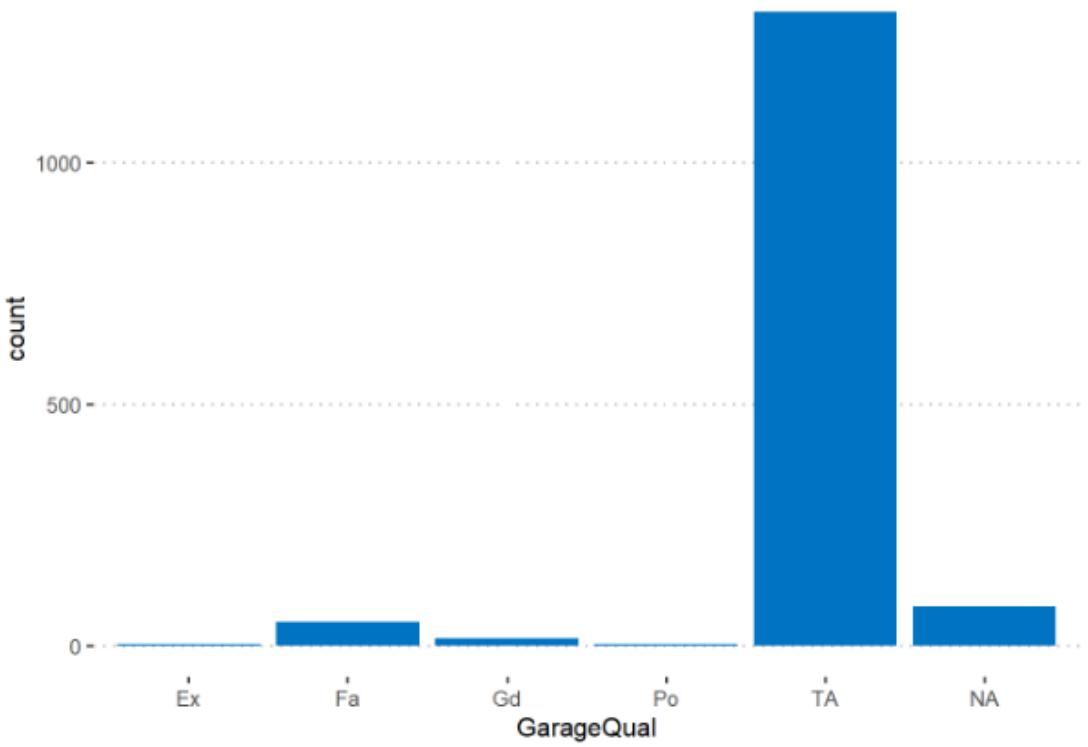

600
400
200
0
count
Ex
Fa
Gd
TA
KitchenQual

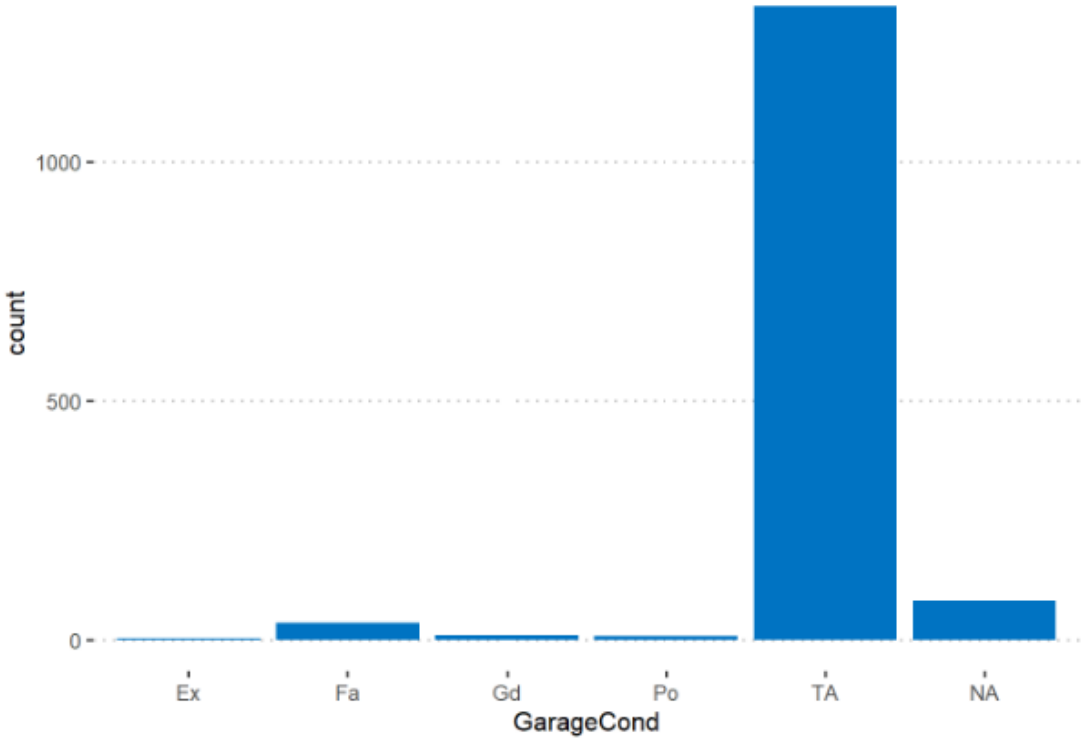

1000
500
0
count
Maj1
Maj2
Min1
Min2
Mod
Sev
Typ
Functional

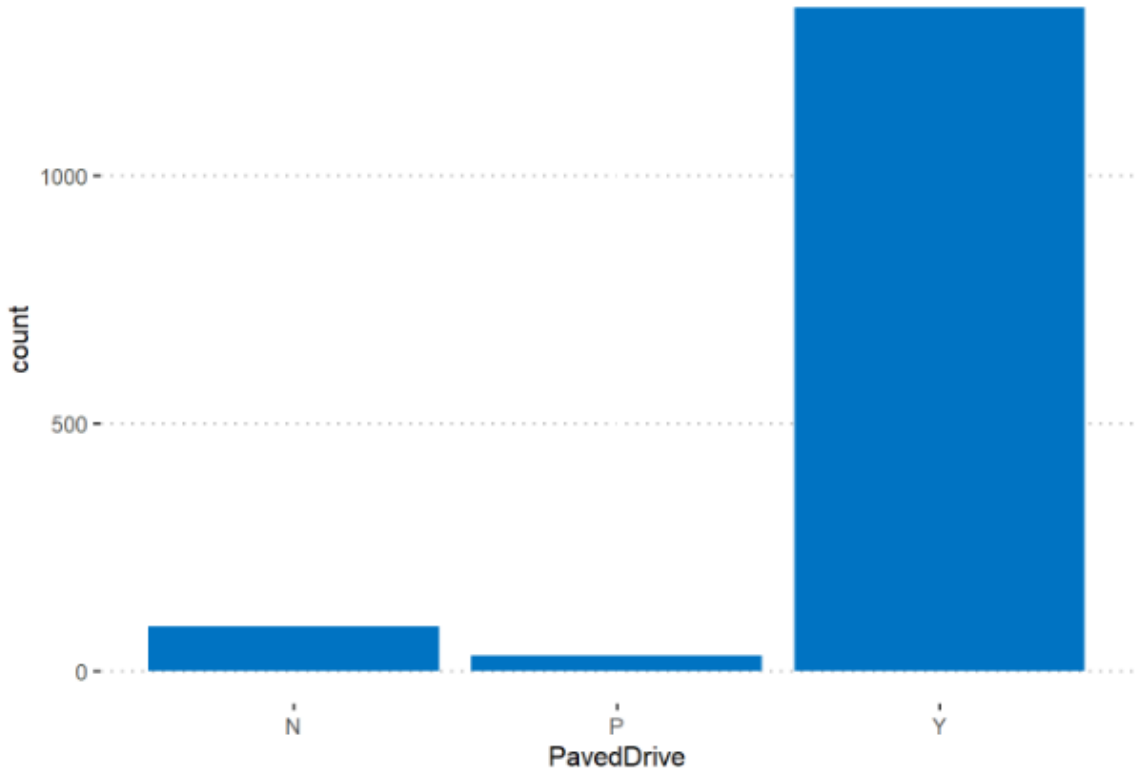

600
400
200
0
count
Ex
Fa
Gd
Po
TA
NA
FireplaceQu

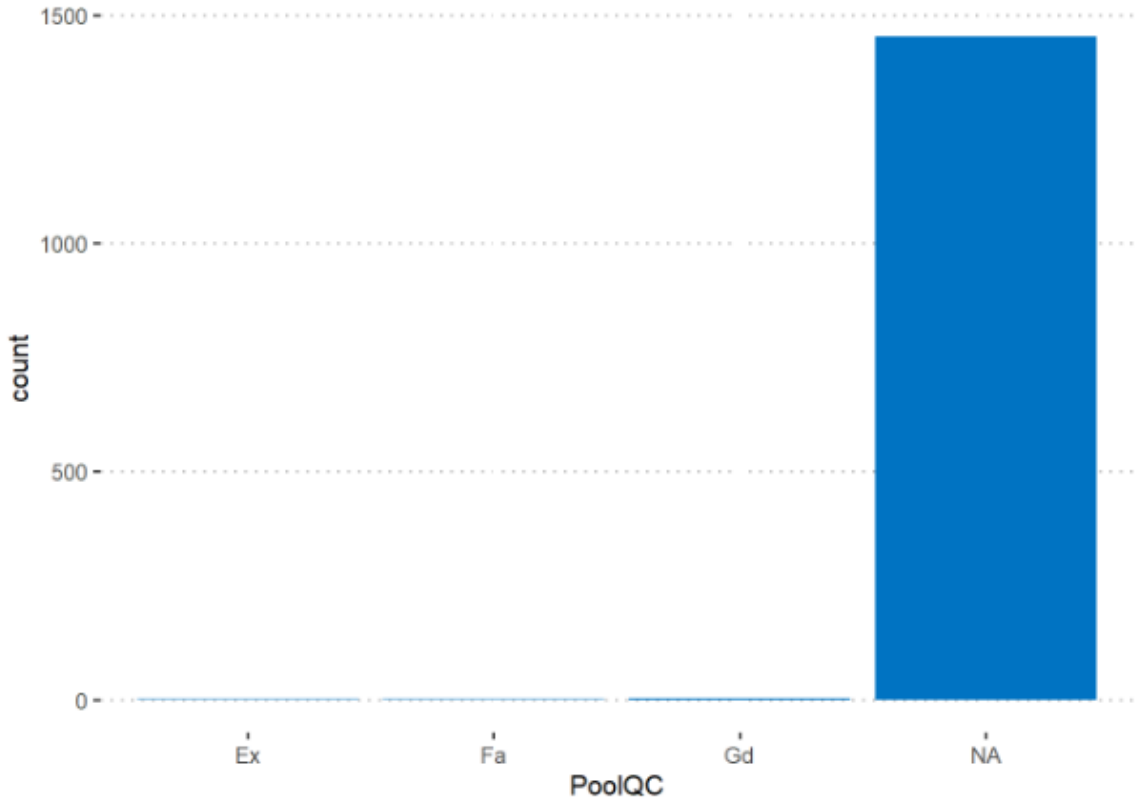

750
500
250
0
count
2Types
Attchd
Basment
BuiltIn
CarPort
Detchd
NA
GarageType

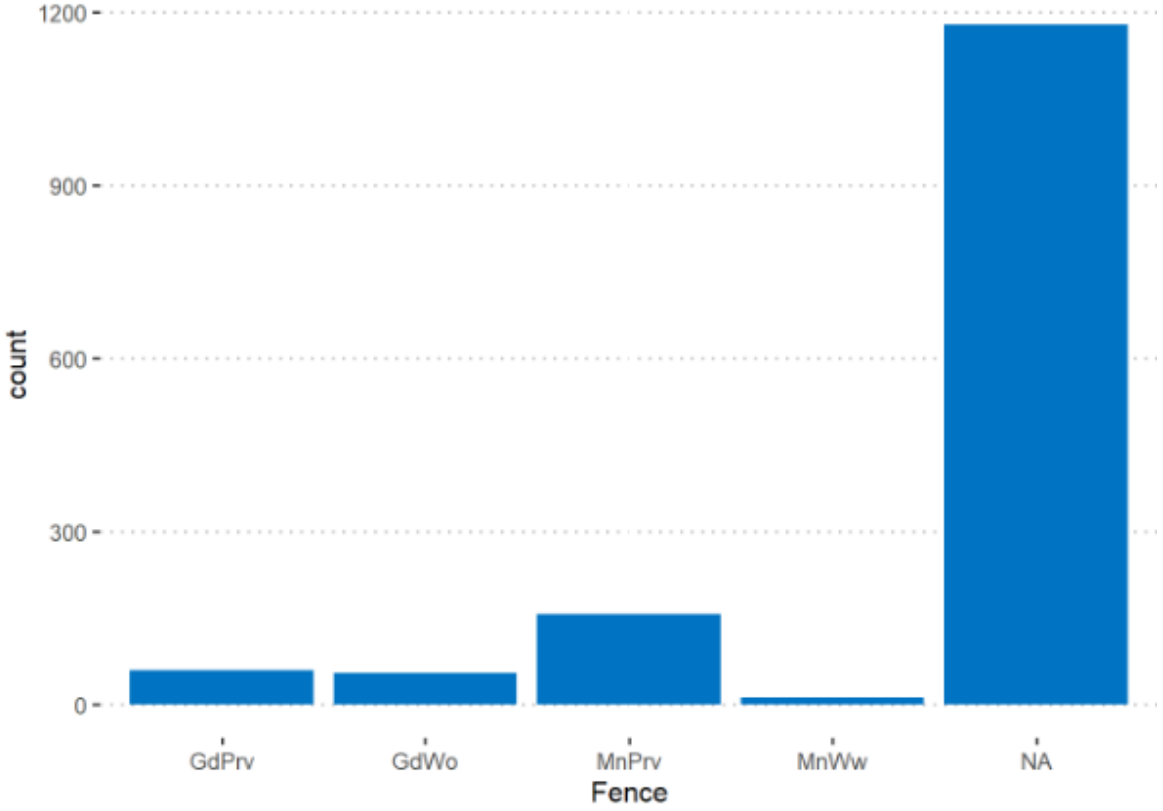

600
400
200
0
count
Fin
RFn
Unf
NA
GarageFinish

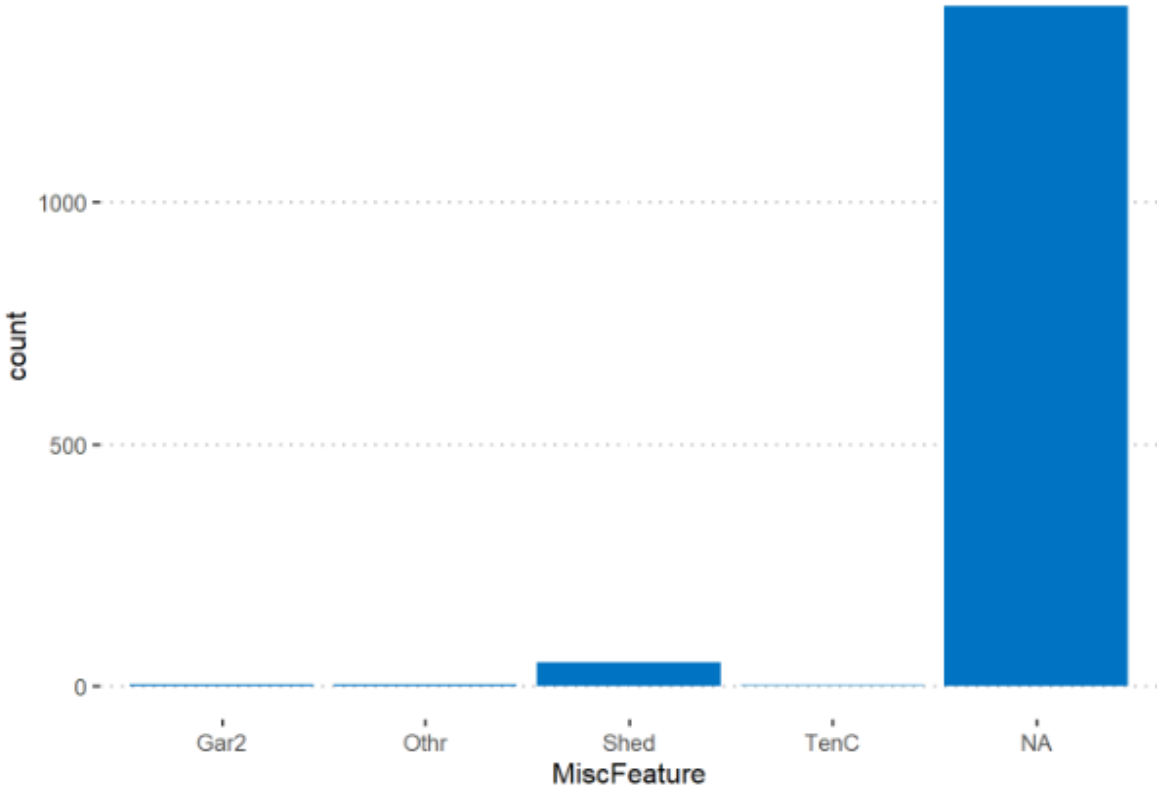

1000
500
0
count
Ex
Fa
Gd
Po
TA
NA
GarageQual

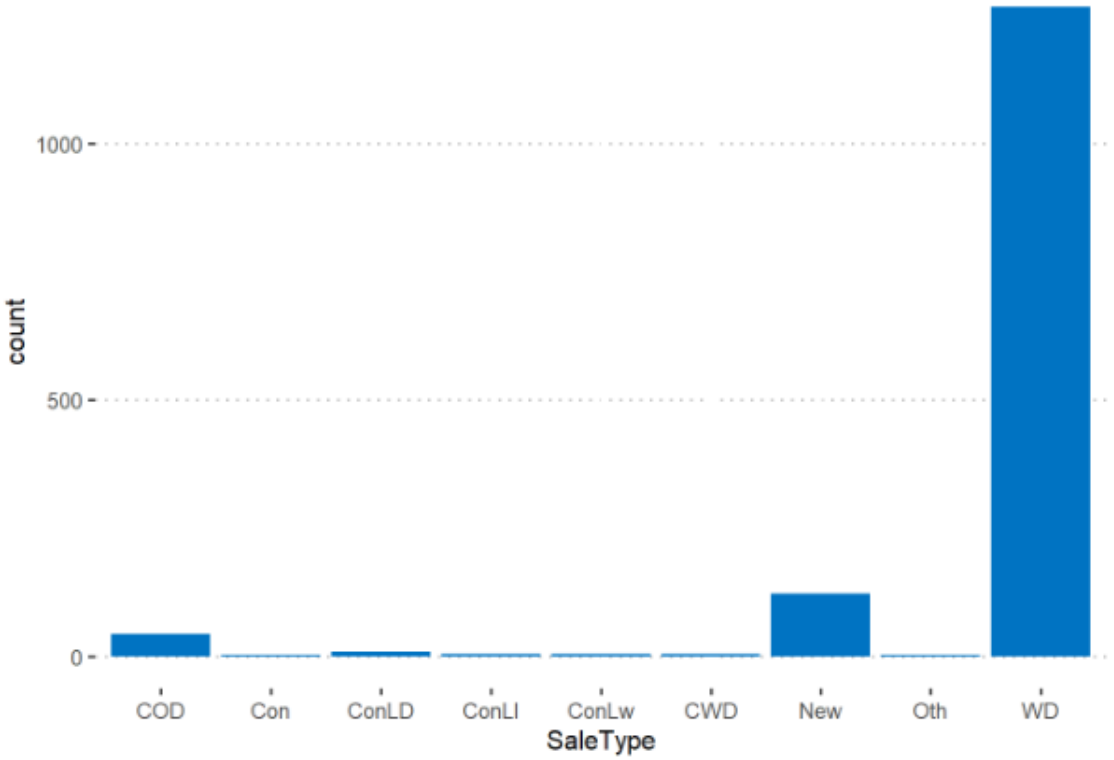

1000
500
0
count
Ex
Fa
Gd
Po
TA
NA
GarageCond

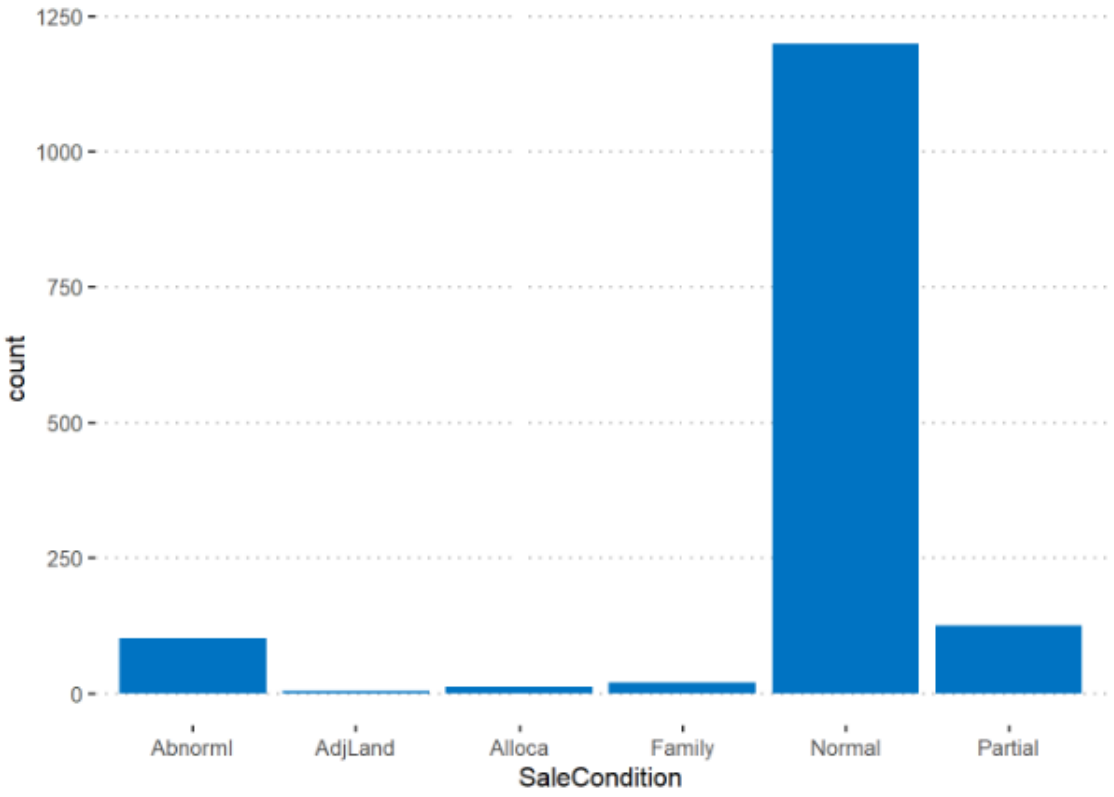

1000
500
0
count
N
P
Y
PavedDrive


1500
1000
500
0
count
Ex
Fa
Gd
NA
PoolQC

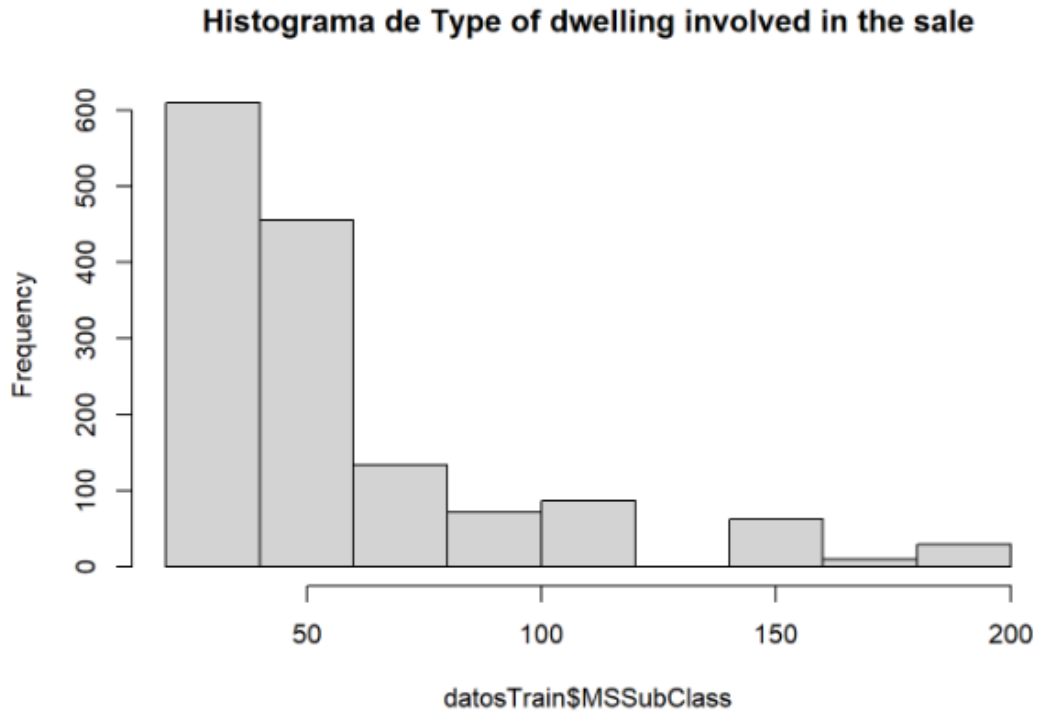

1200
900
600
300
0
count
GdPrv
GdWo
MnPrv
MnWw
NA
Fence

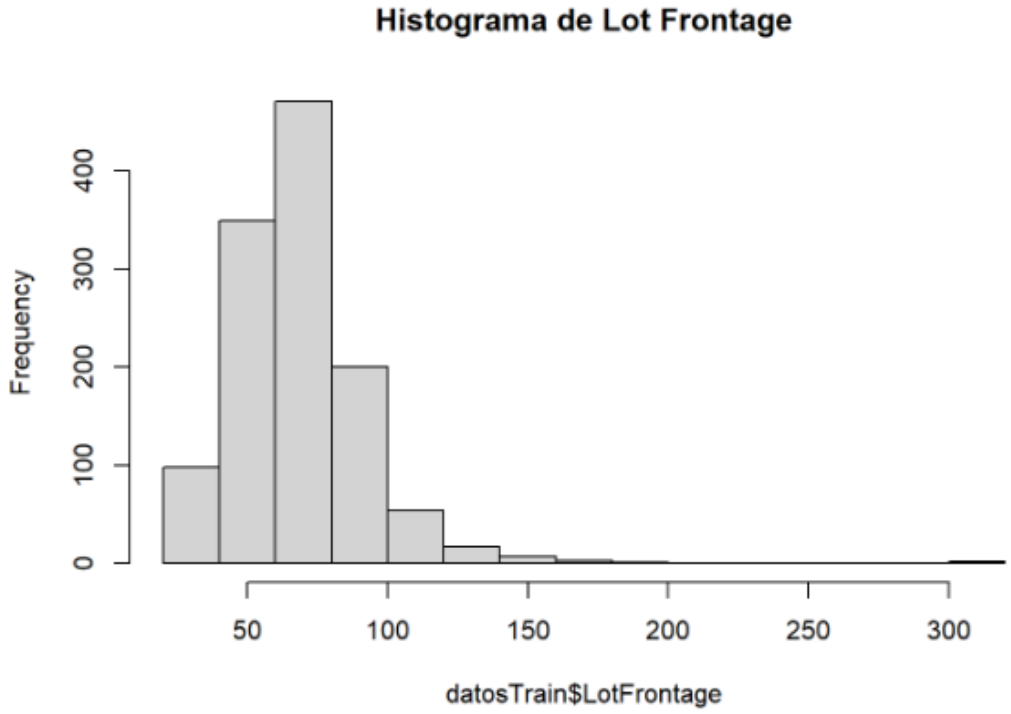

1000
500
0
count
Gar2
Othr
Shed
TenC
NA
MiscFeature


1000
500
0
count
COD
Con
ConLD
ConLI
ConLw
CWD
New
Oth
WD
SaleType

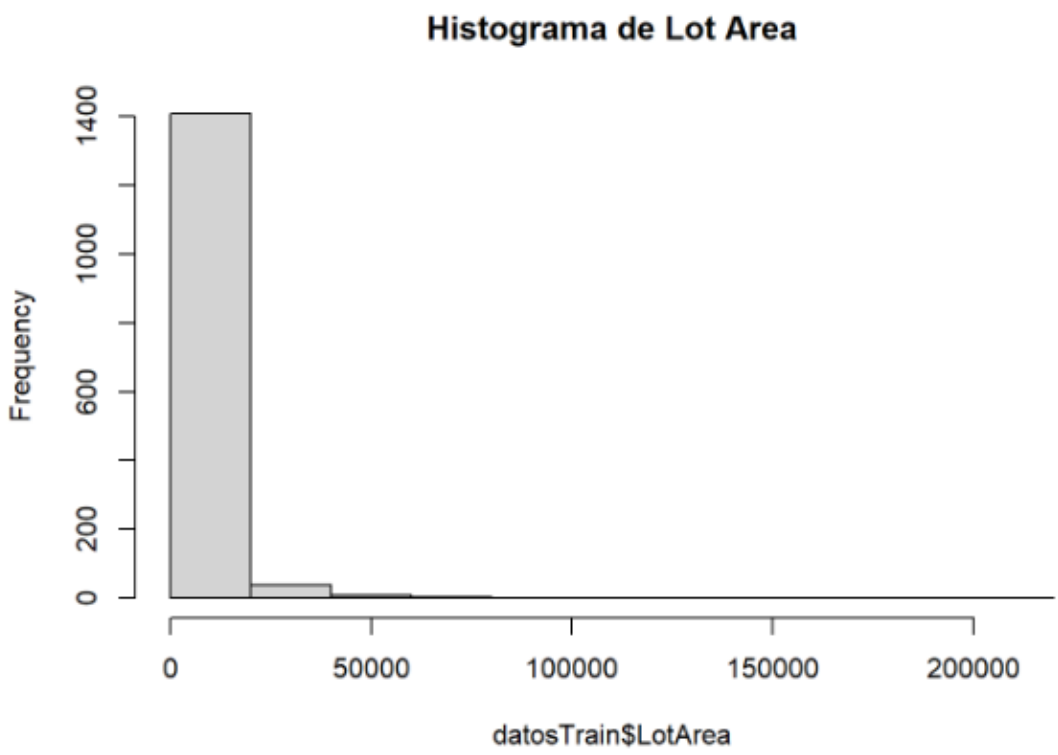

1250
1000
750
500
250
0
count
Abnorml
AdjLand
Alloca
Family
Normal
Partial
SaleCondition

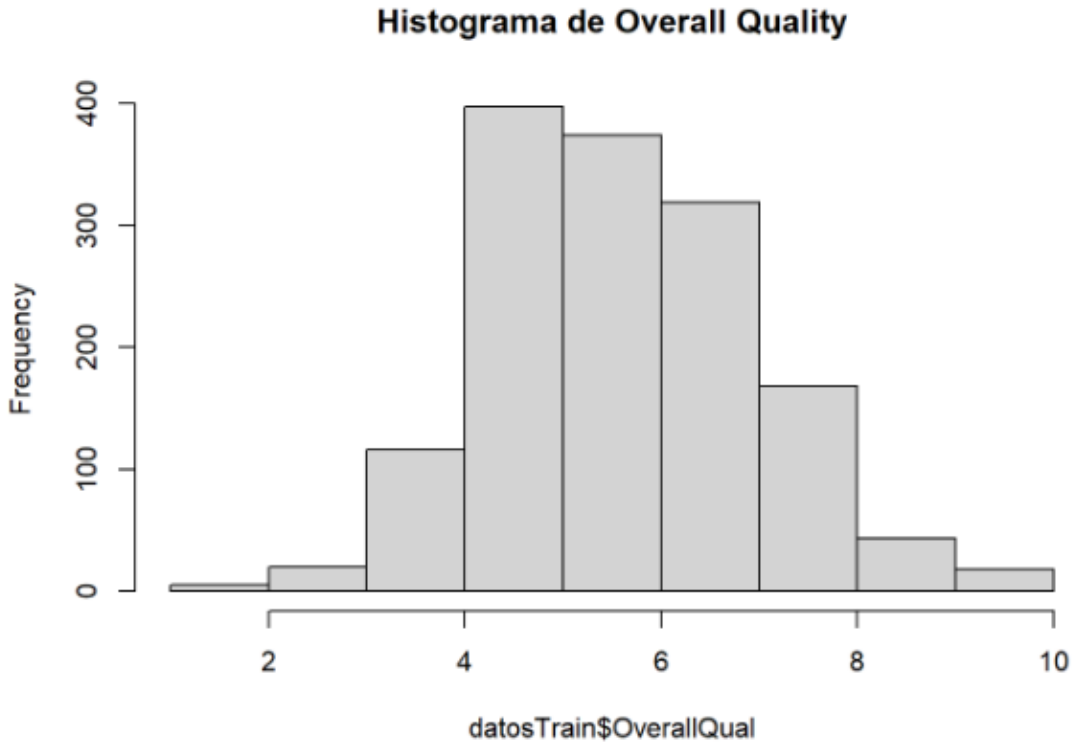
Por otro lado, para las variables cuantitativas, tanto continuas como discretas, se hicieron los siguientes histogramas:



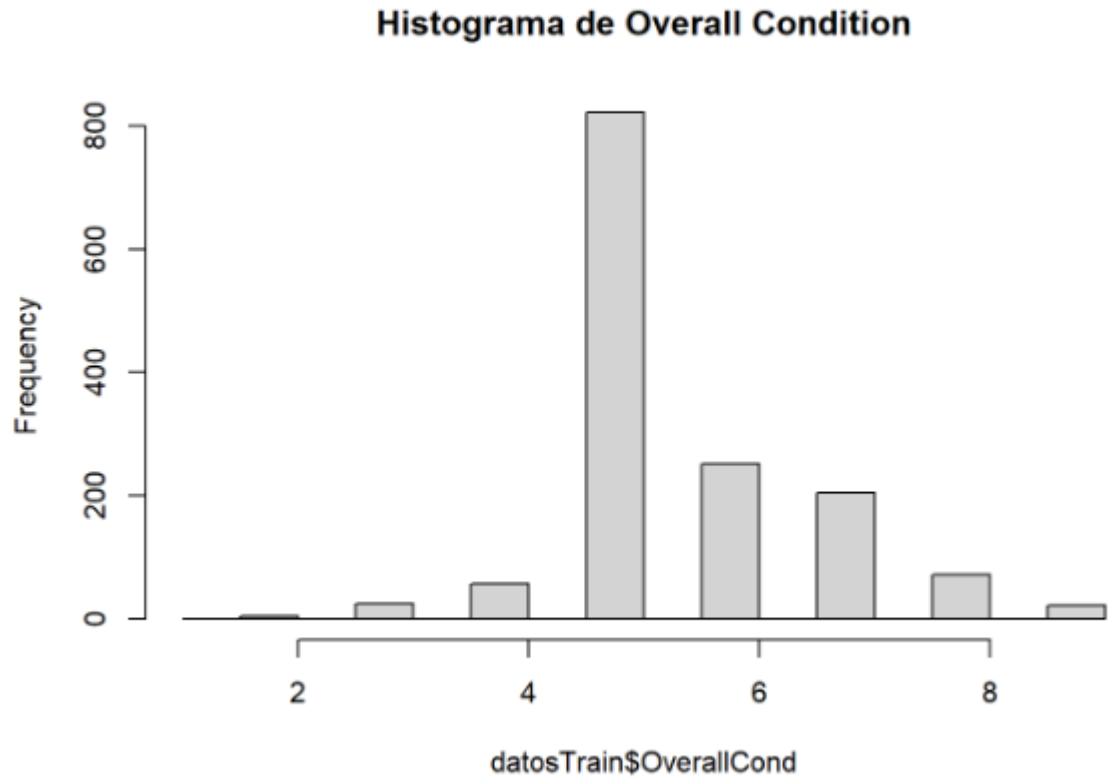


Se puede observar que LotFrontage sigue una distribución normal por su naturaleza continua, que es la distancia del hogar a la calle.

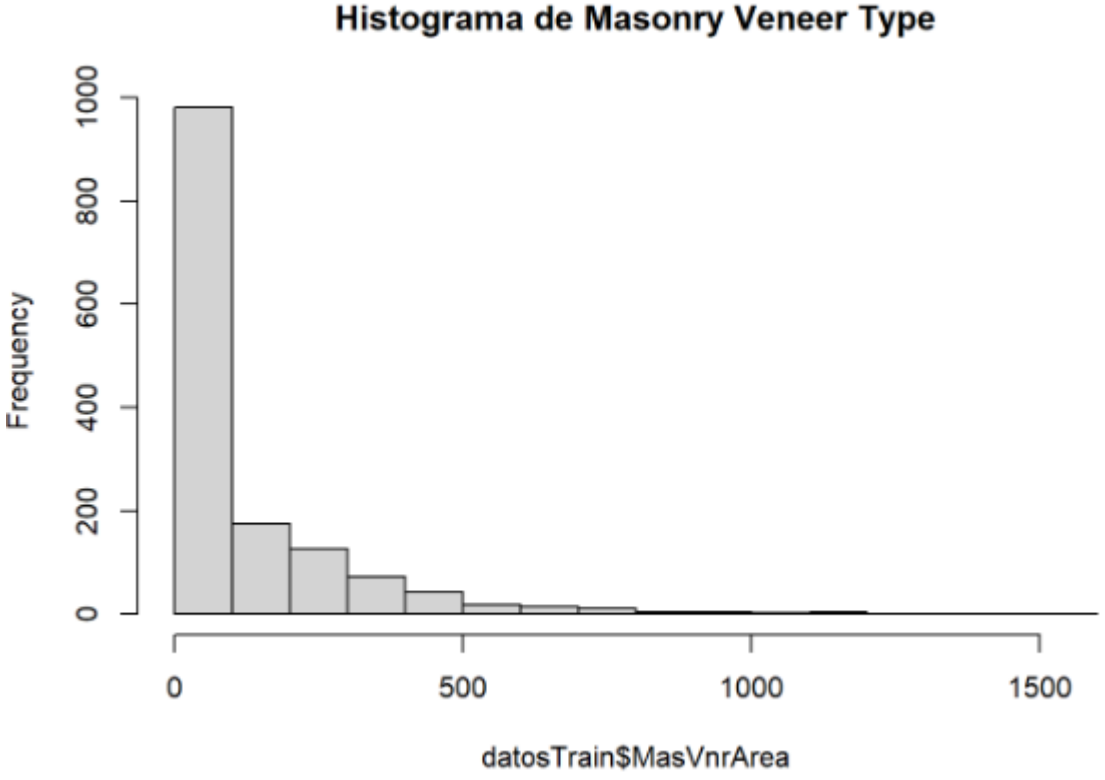


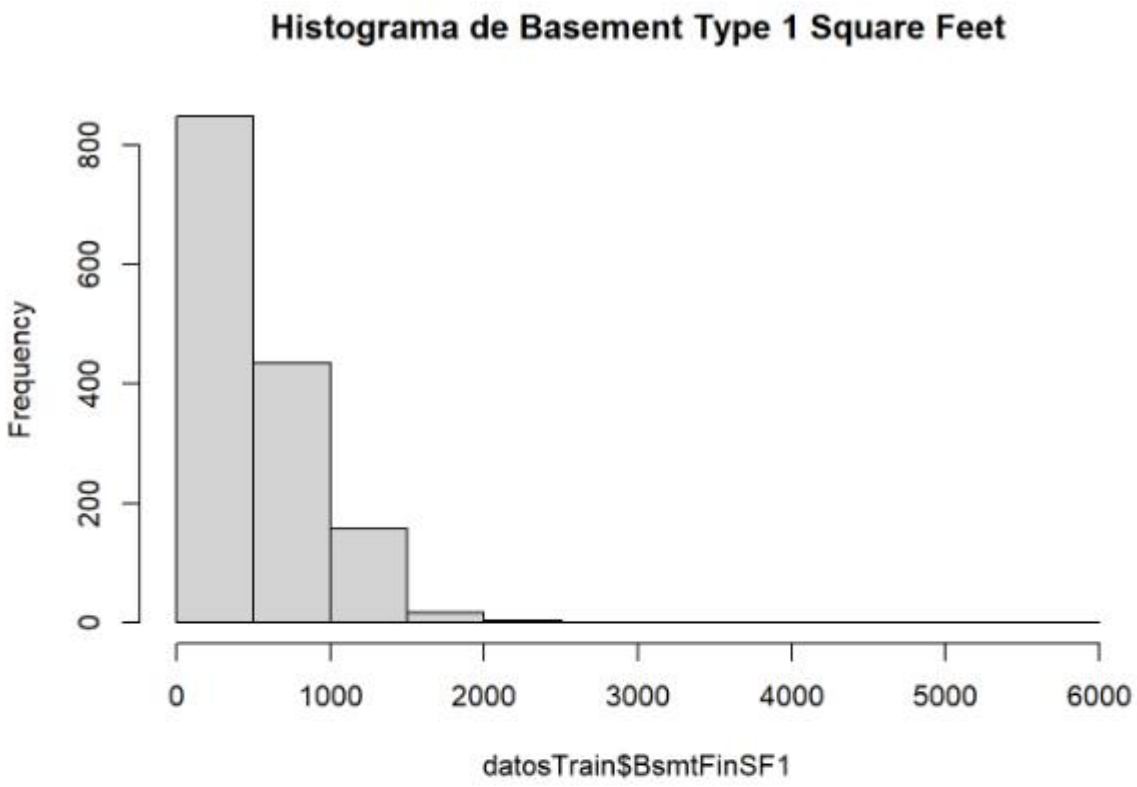


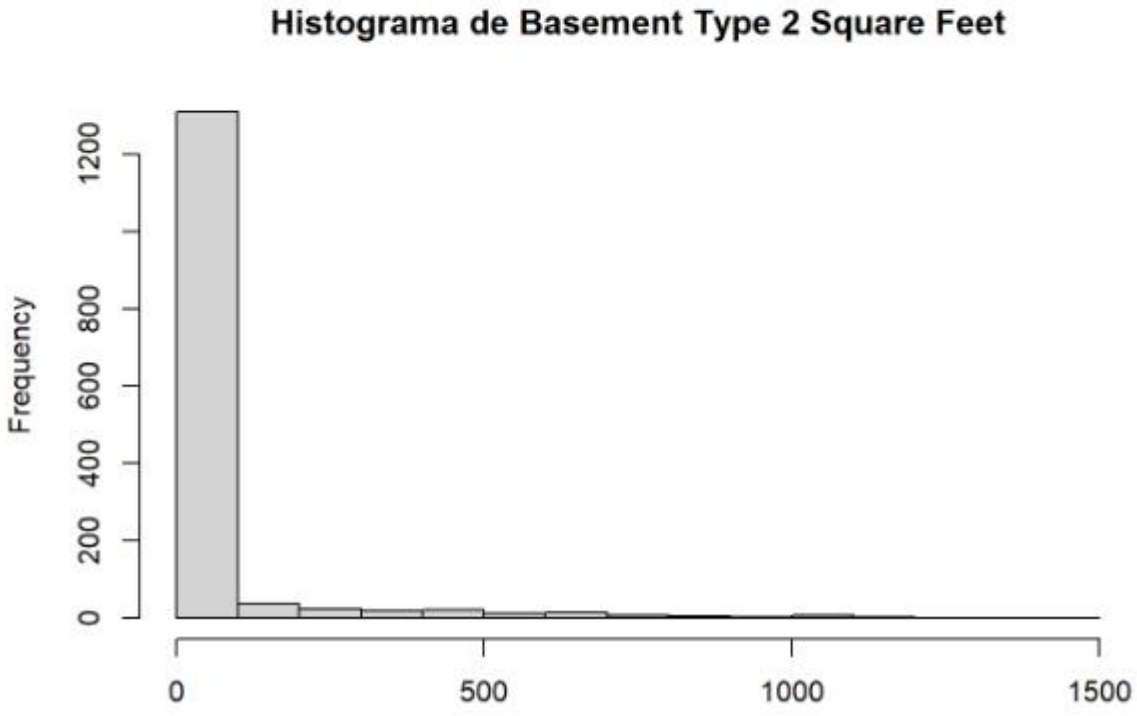
Se observa que Overall Quality también sigue una distribución normal.

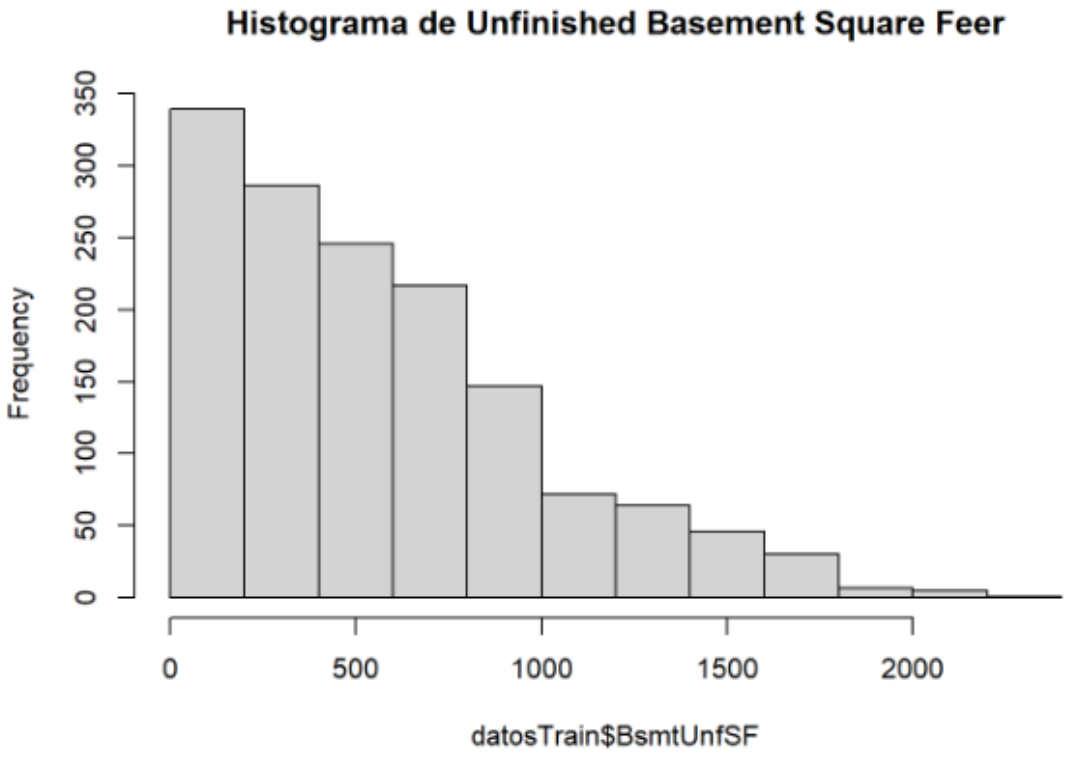


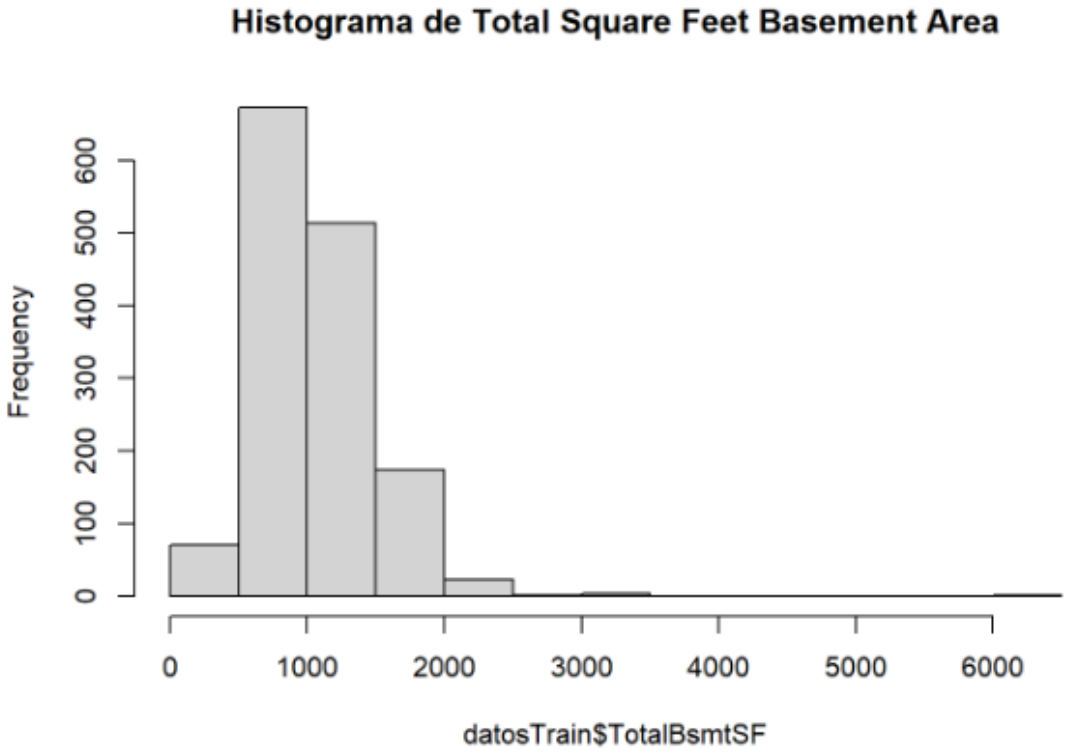
Overall Condition también sigue una distribución normal



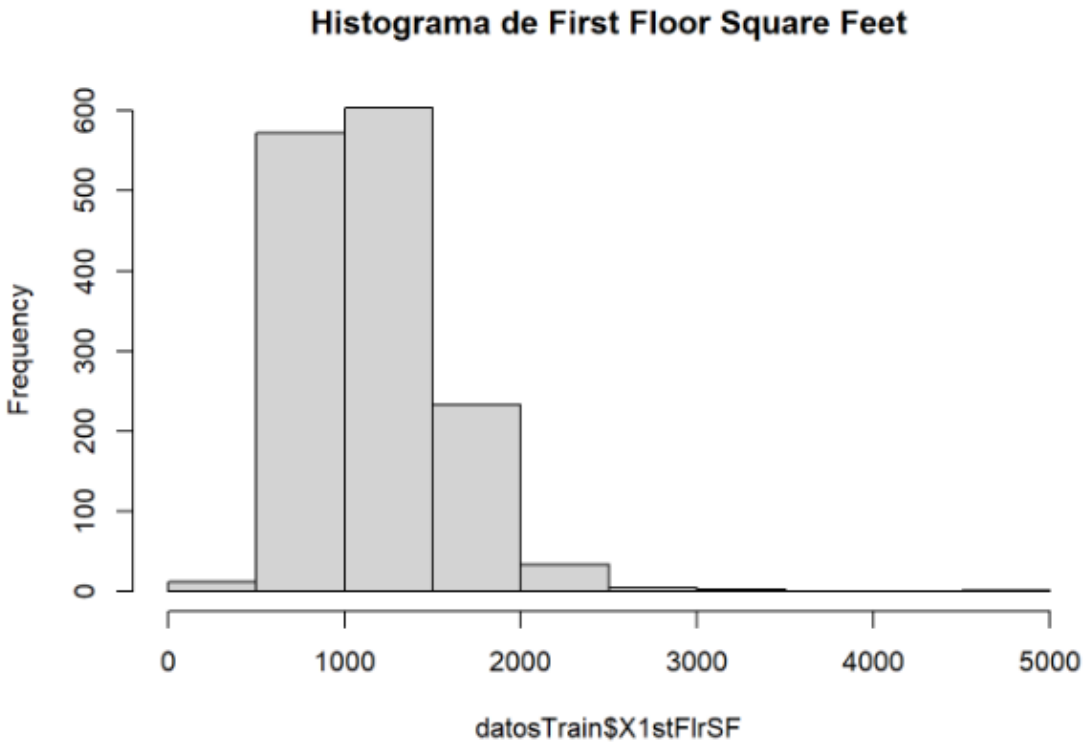

Histograma de Basement Type 1 Square Feet
Frequency
0
200
400
600
800
0
1000
2000
3000
4000
5000
6000
datosTrain$BsmtFinSF1

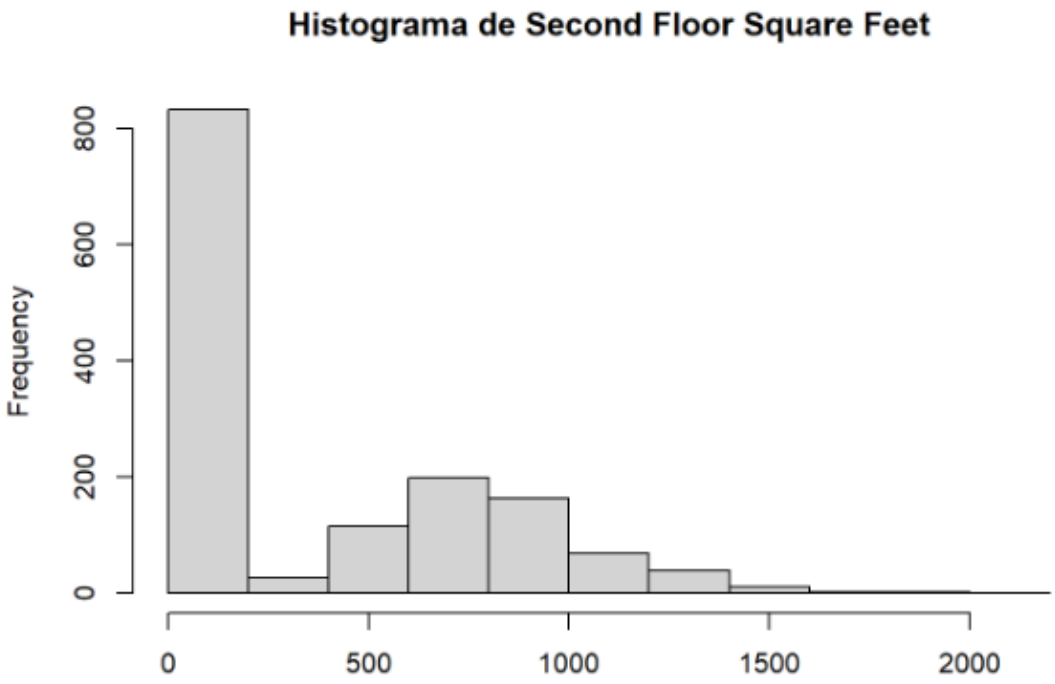

Histograma de Basement Type 2 Square Feet
Frequency
0
200
400
600
800
1200
0
500
1000
1500



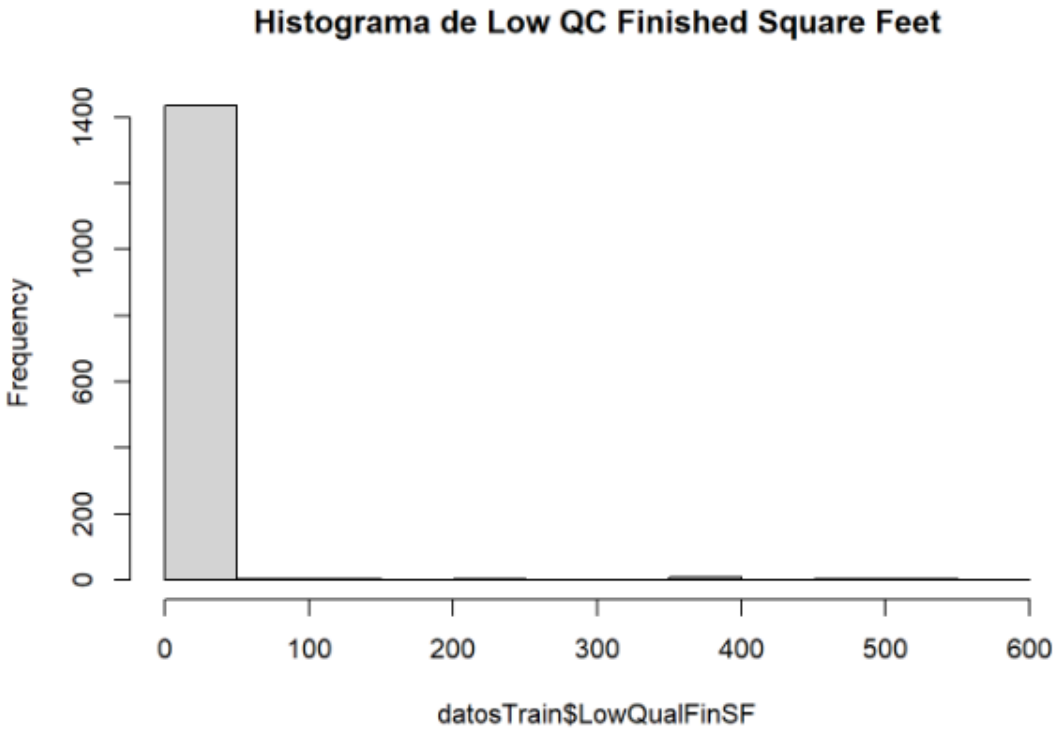


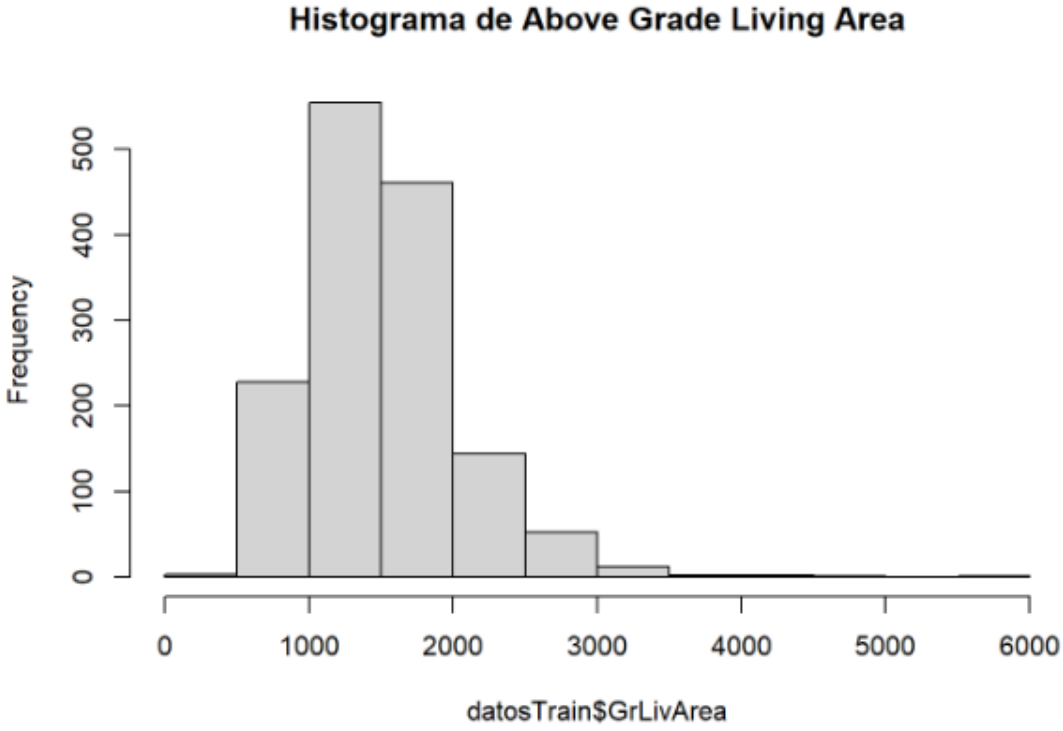
Se observa que Total Square Feet Basement Area también sigue una distribución normal.



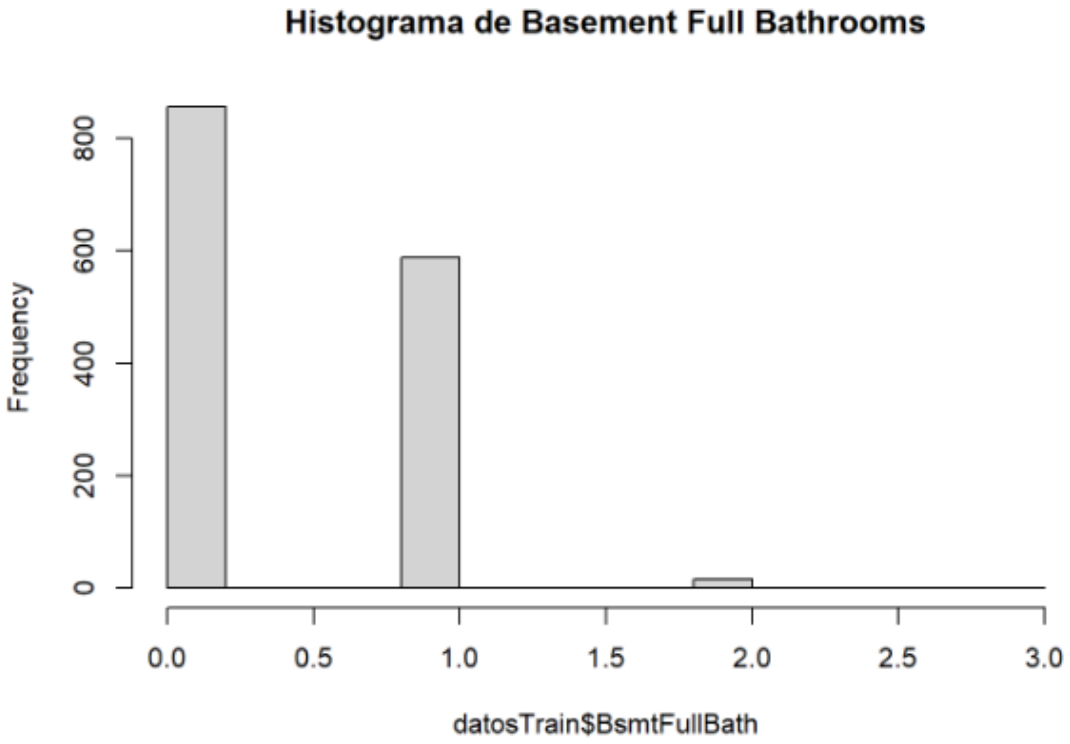


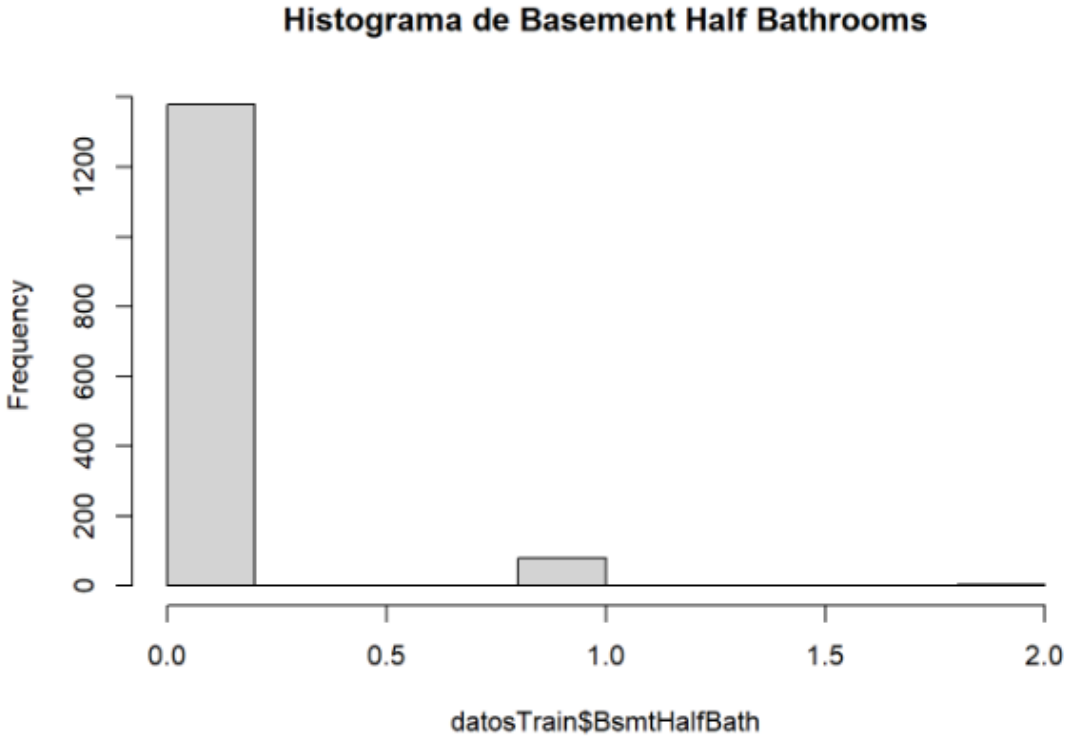
Se observa que First Floor Square Feet sigue una distribución normal.

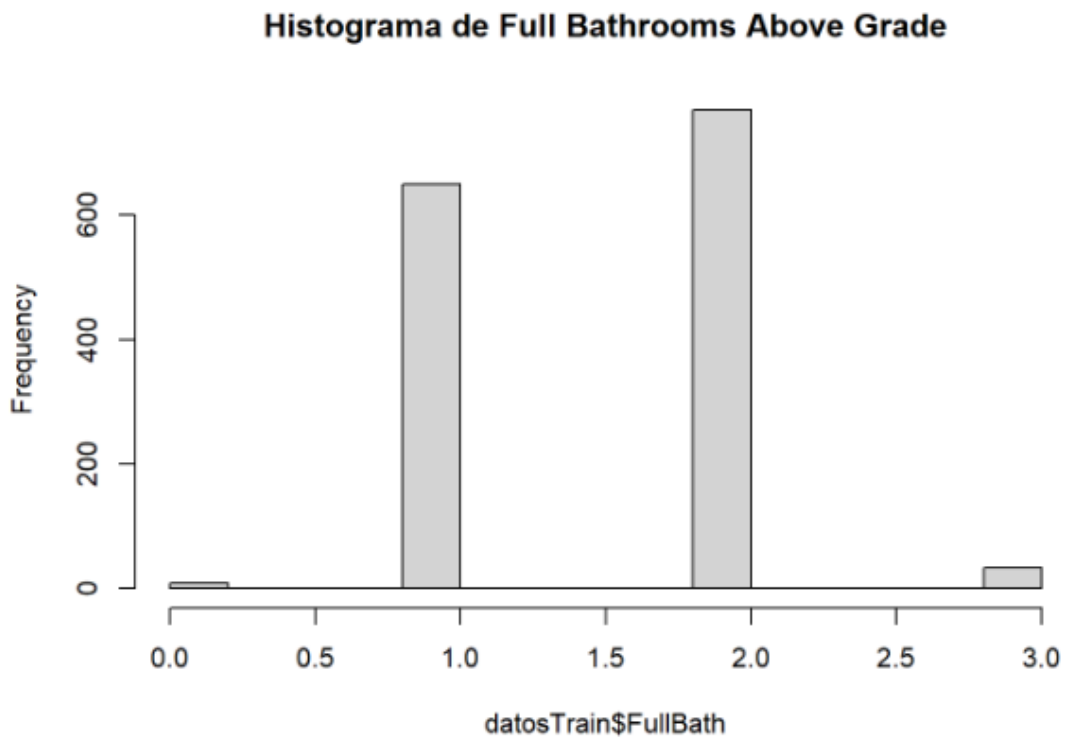


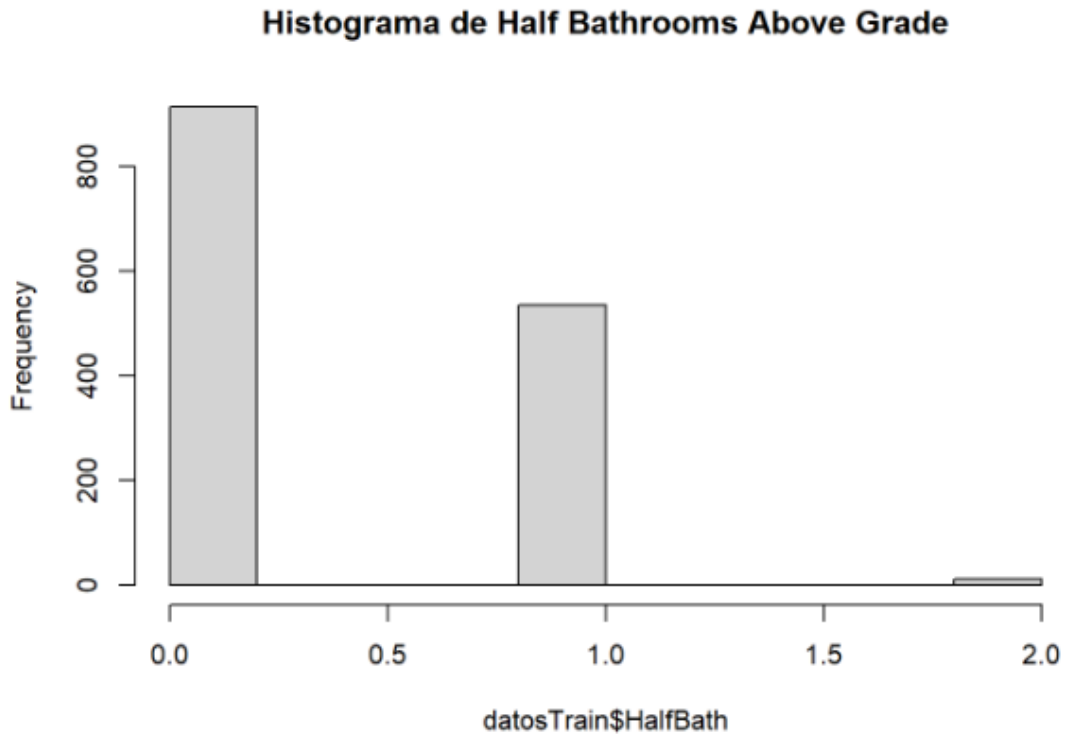


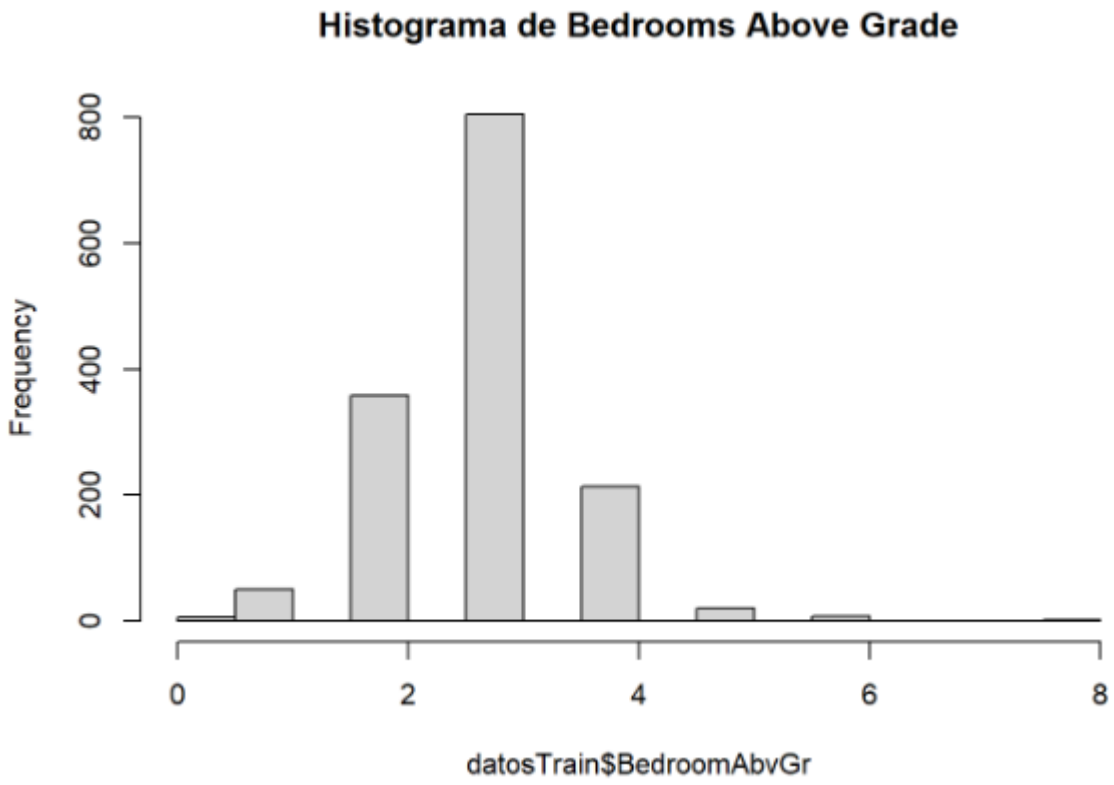
Se observa que Above Grade Living Area tiene una distribución normal.


Histograma de Basement Full Bathrooms
Frequency
0
200
400
600
800
0.0
0.5
1.0
1.5
2.0
2.5
3.0
datosTrain$BsmtFullBath

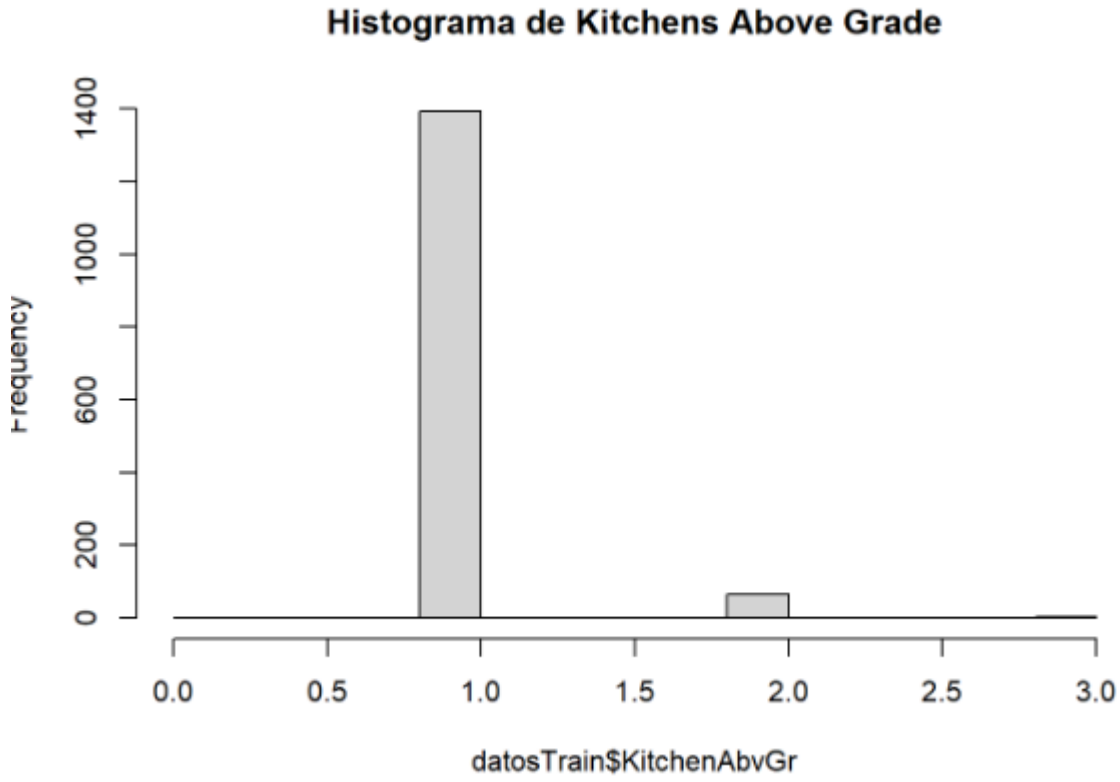

Histograma de Basement Half Bathrooms
Frequency
0
200
400
600
800
1200
0.0
0.5
1.0
1.5
2.0
datosTrain$BsmtHalfBath

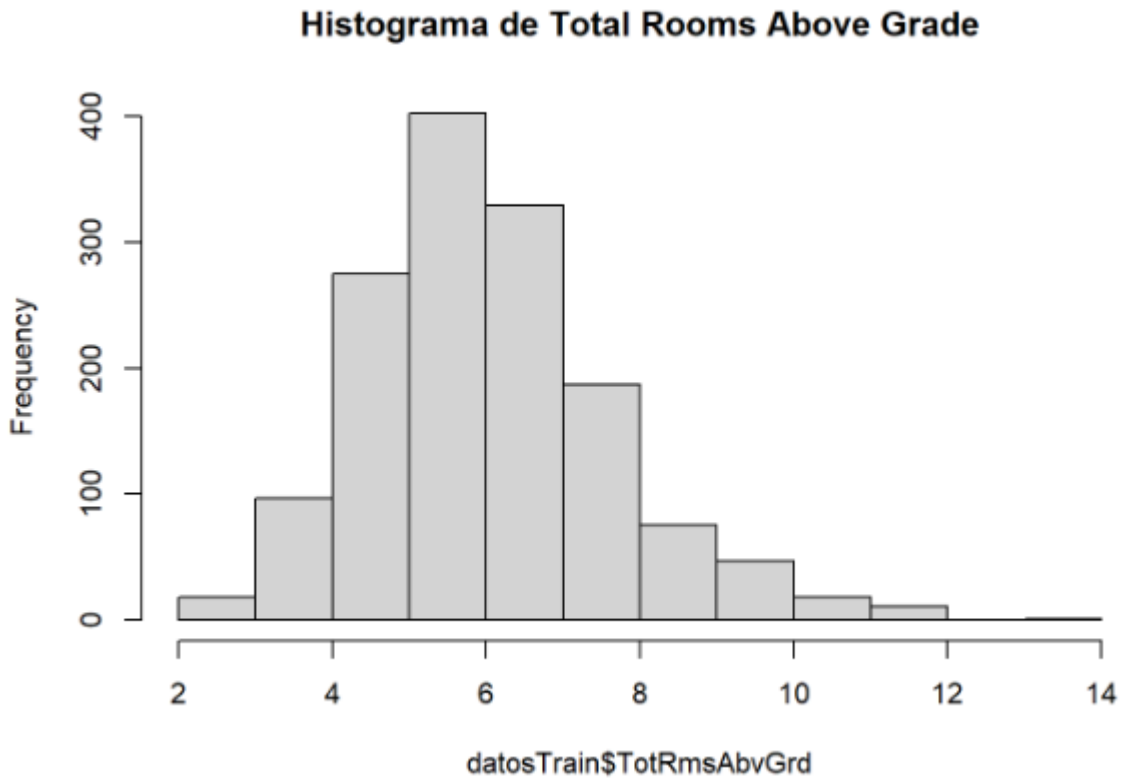

Histograma de Full Bathrooms Above Grade
Frequency
0
200
400
600
0.0
0.5
1.0
1.5
2.0
2.5
3.0
datosTrain$FullBath


Histograma de Half Bathrooms Above Grade
Frequency
0
200
400
600
800
0.0
0.5
1.0
1.5
2.0
datosTrain$HalfBath

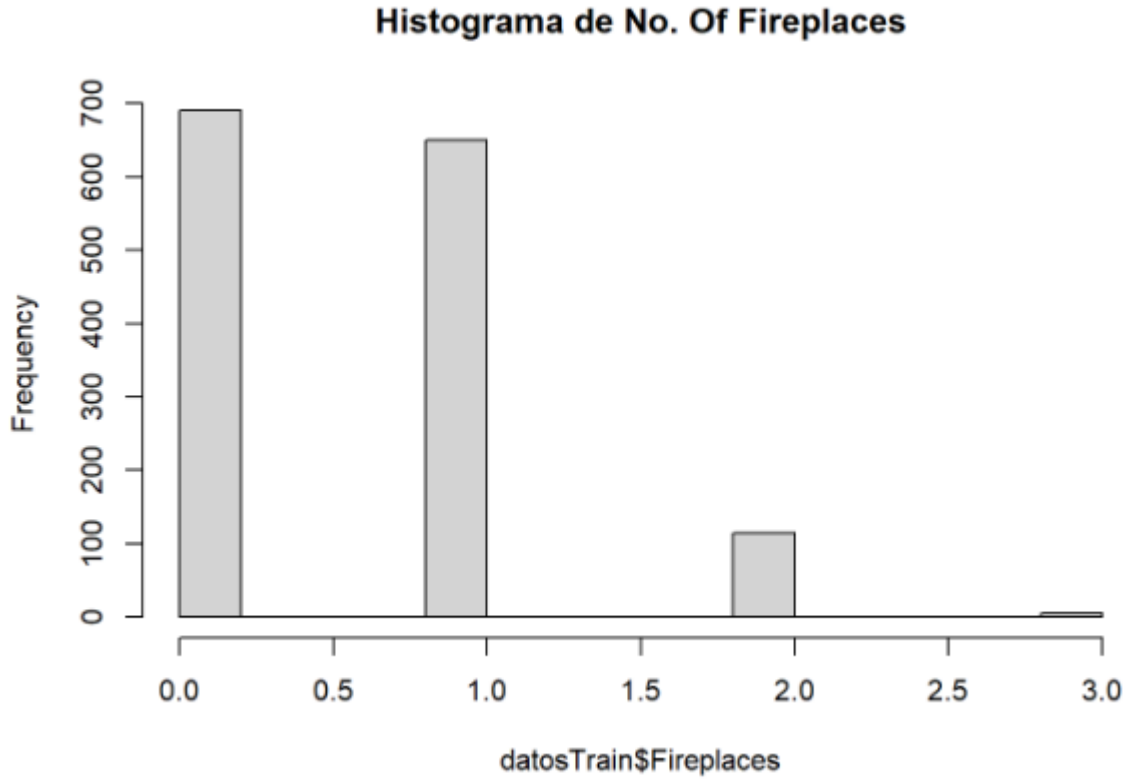


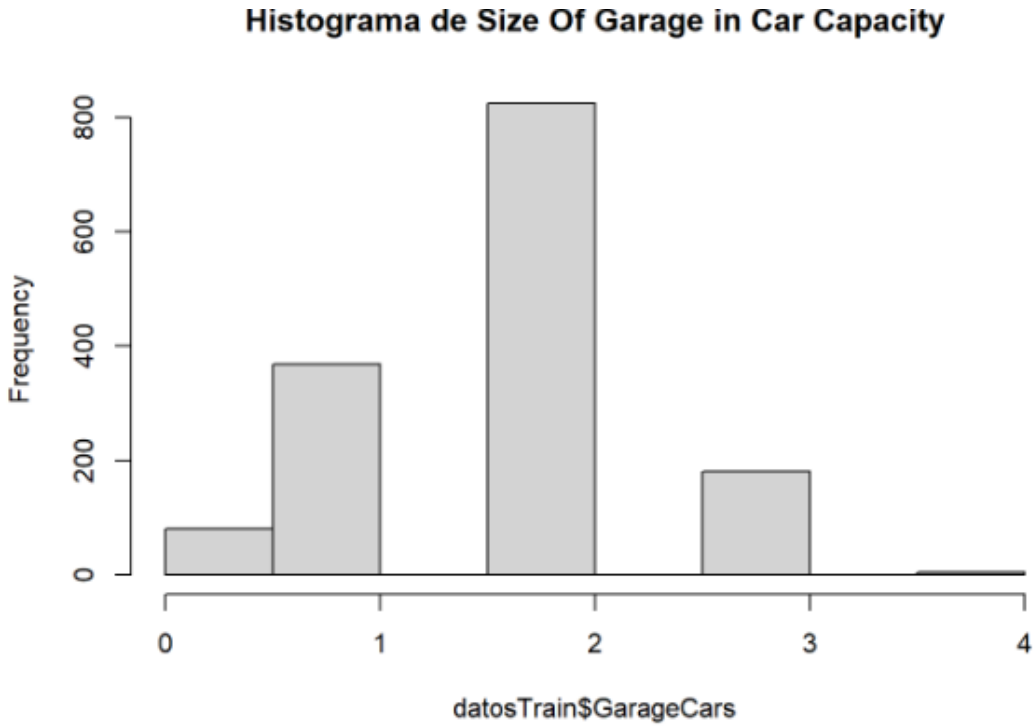
Se observa que Bedrooms Above Grade sigue una distribución normal.

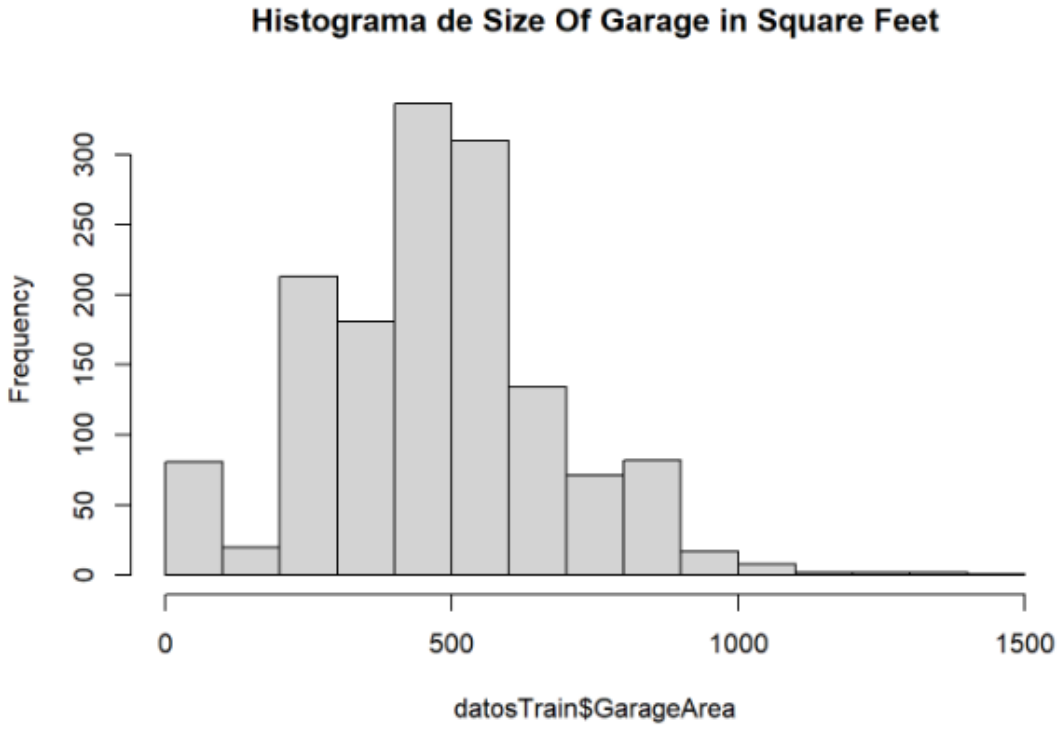




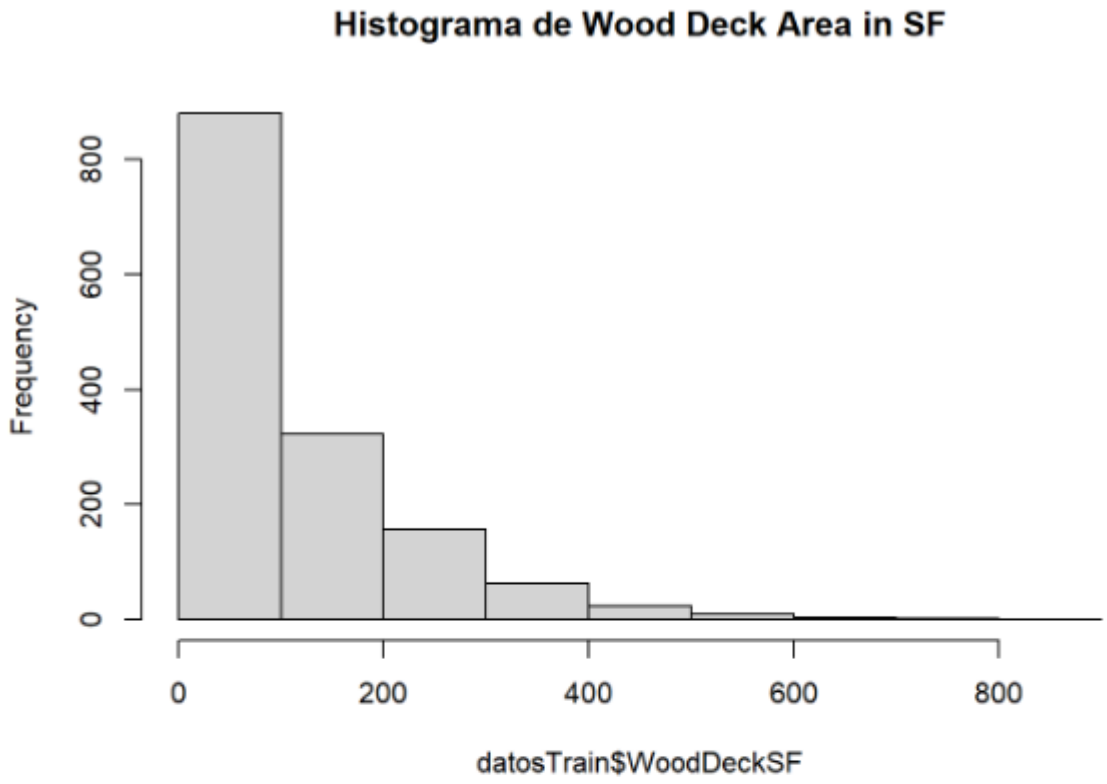
Se observa que Total Rooms Above Grade sigue una distribución normal

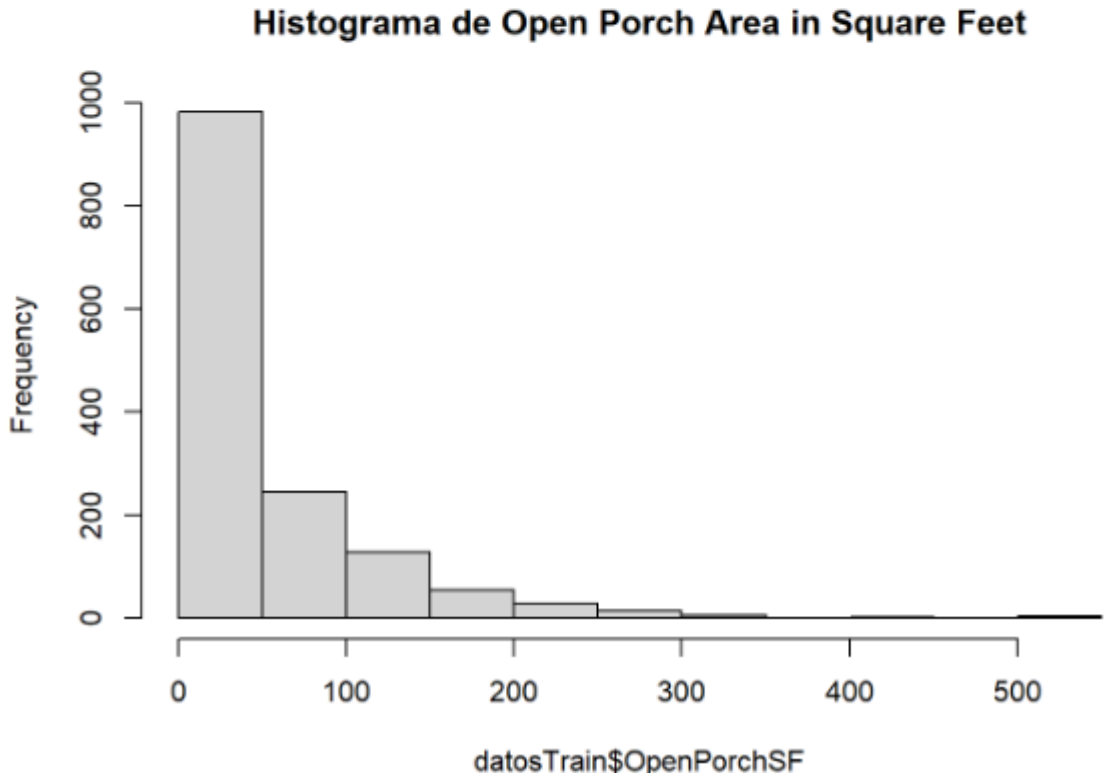


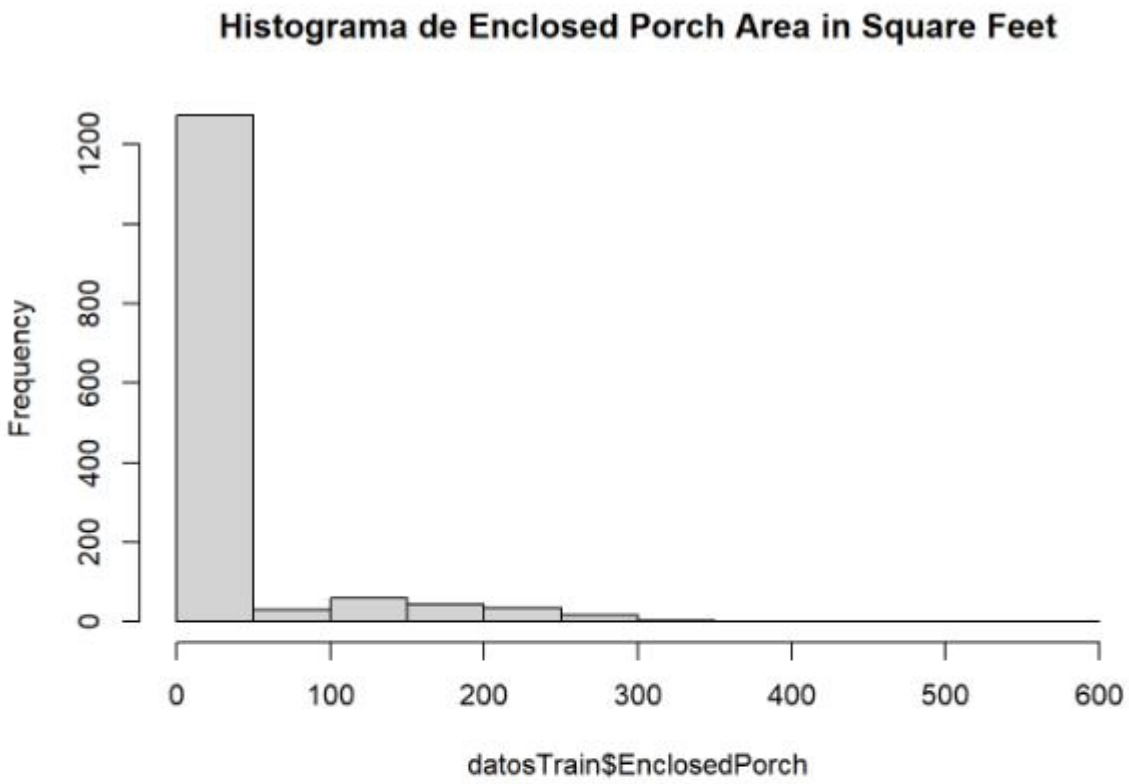


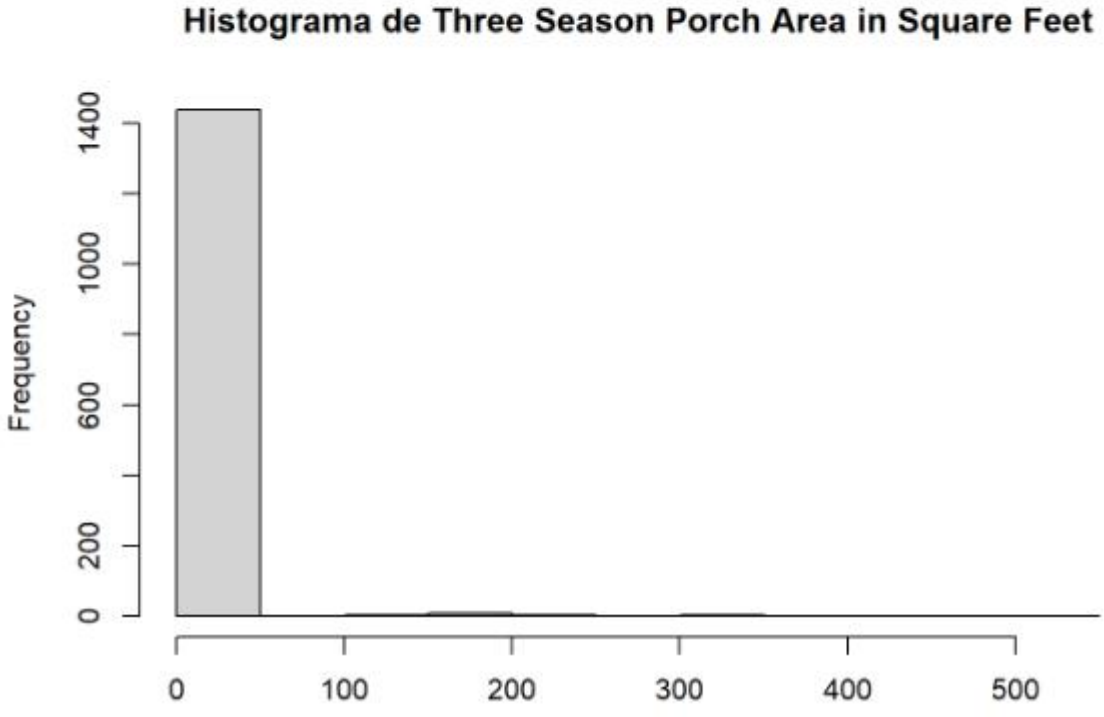


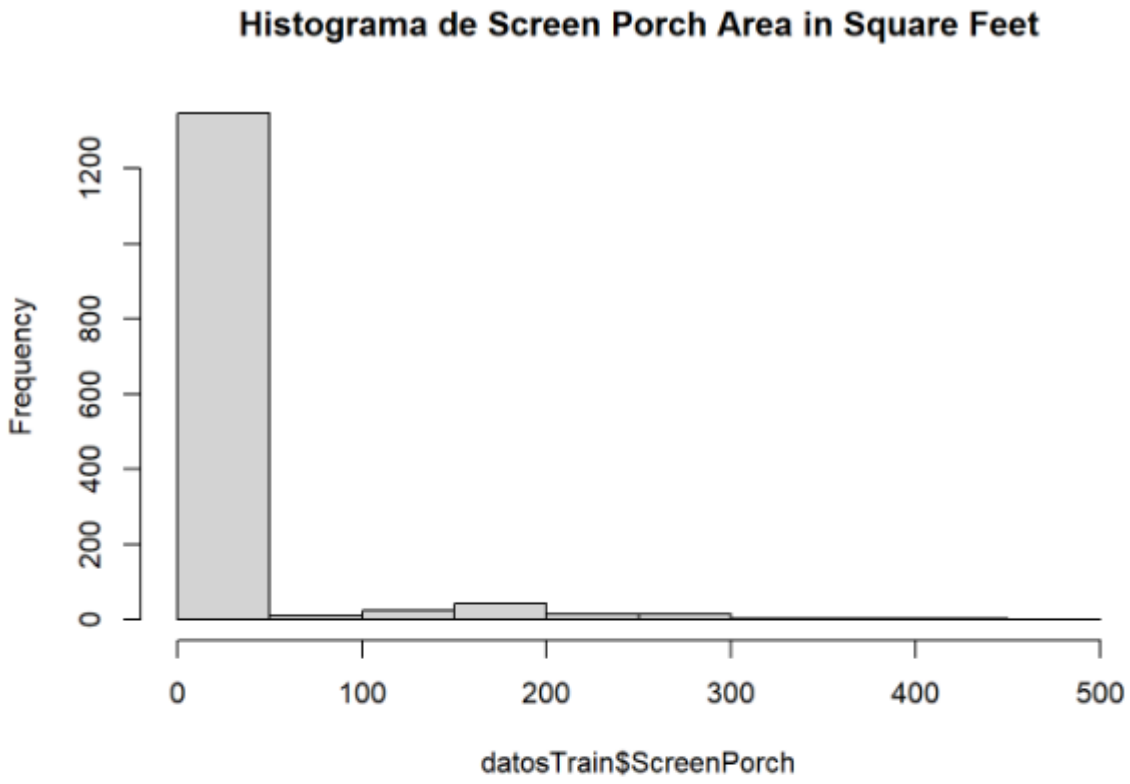
Ambos Size Of Garage in Car Capacity y Size Of Garage in Square Feet siguen una distribución normal.

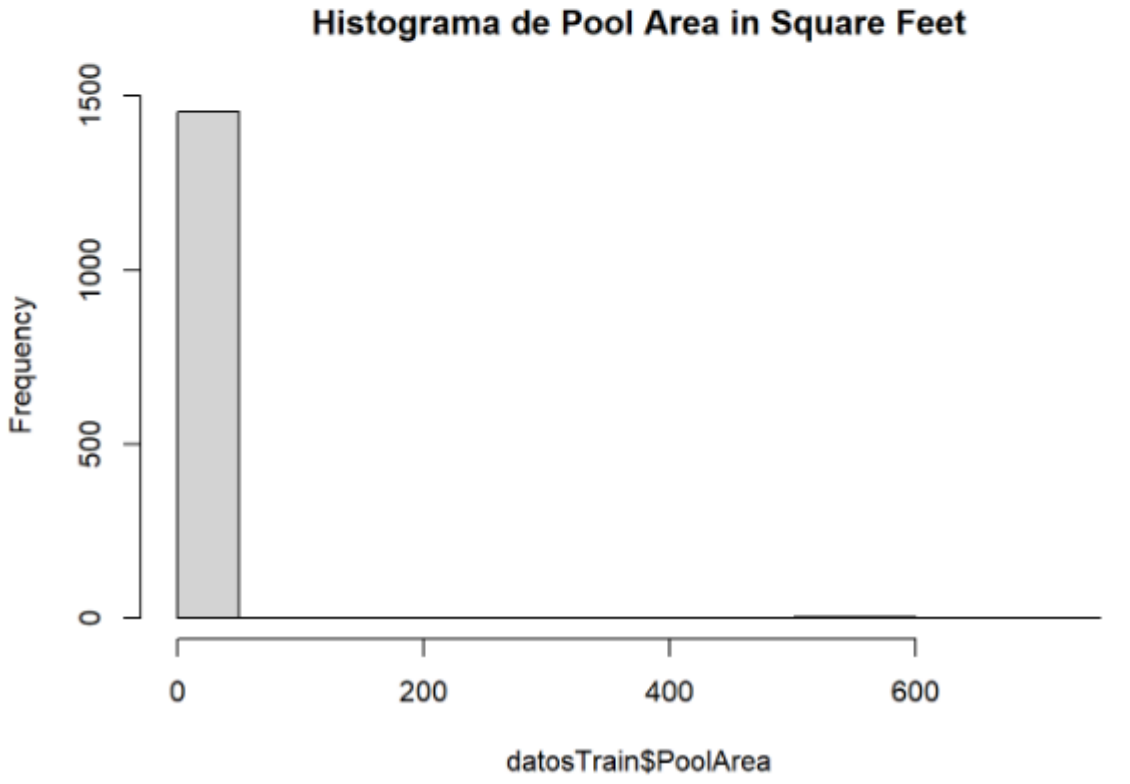

Histograma de Wood Deck Area in SF
Frequency
0
200
400
600
800
0
200
400
600
800
datosTrain$WoodDeckSF

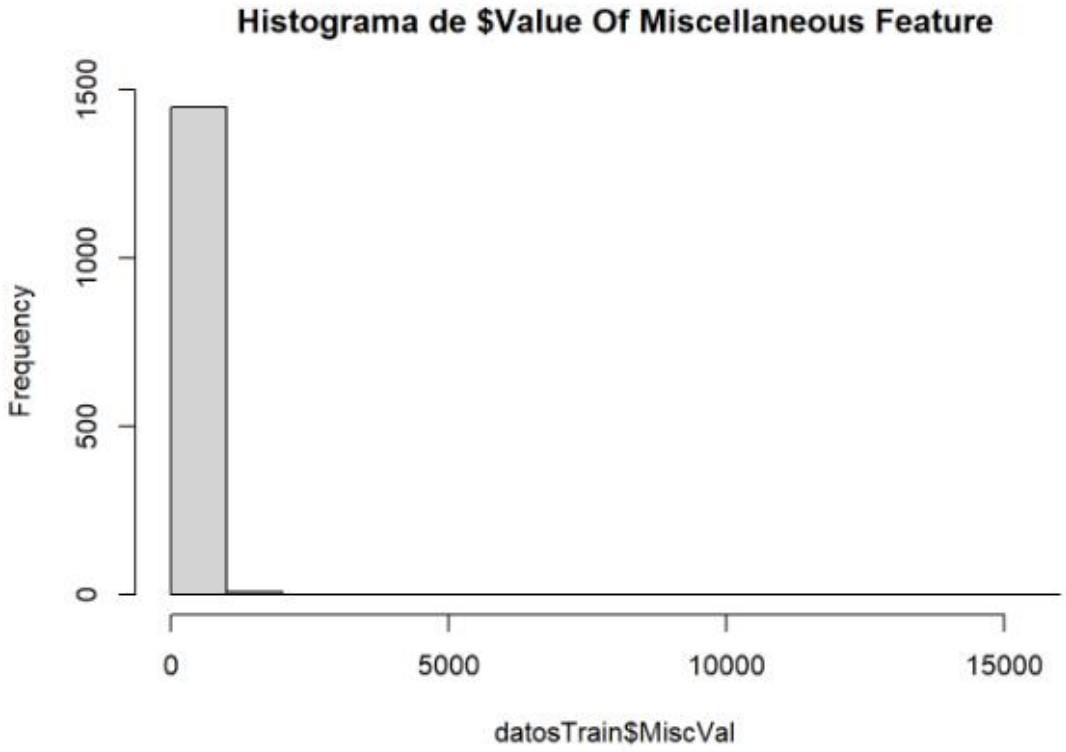

Histograma de Open Porch Area in Square Feet
Frequency
0
200
400
600
800
1000
0
100
200
300
400
500
datosTrain$OpenPorchSF

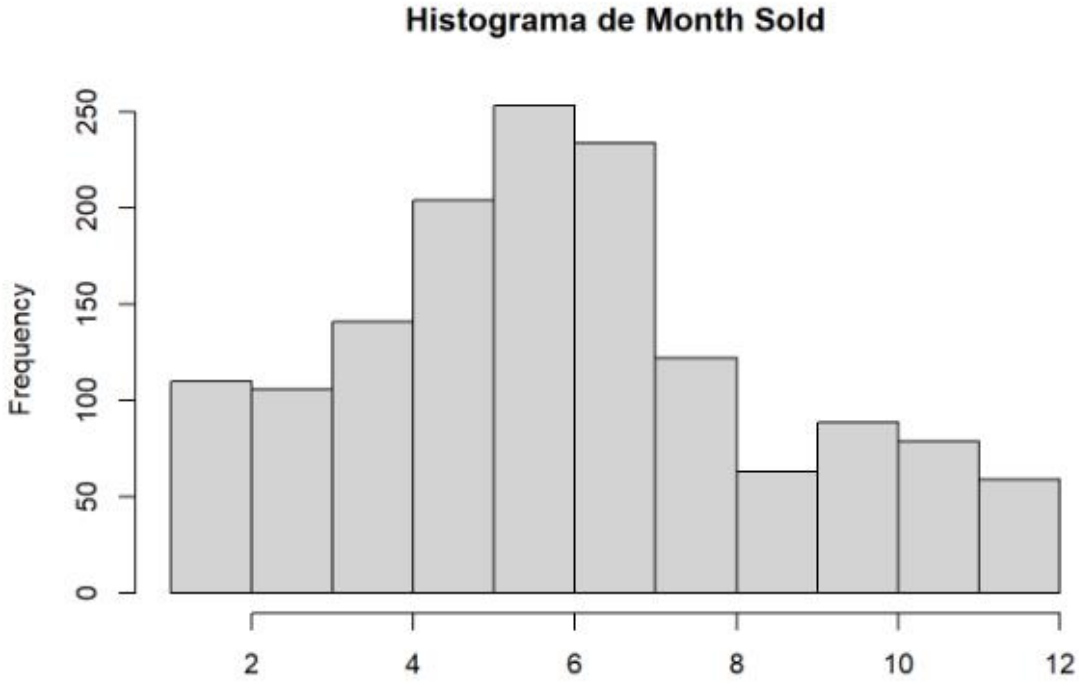

Histograma de Enclosed Porch Area in Square Feet
Frequency
0
200
400
600
800
1200
0
100
200
300
400
500
600
datosTrain$EnclosedPorch


Histograma de Three Season Porch Area in Square Feet
Frequency
0
200
600
1000
1400
0
100
200
300
400
500

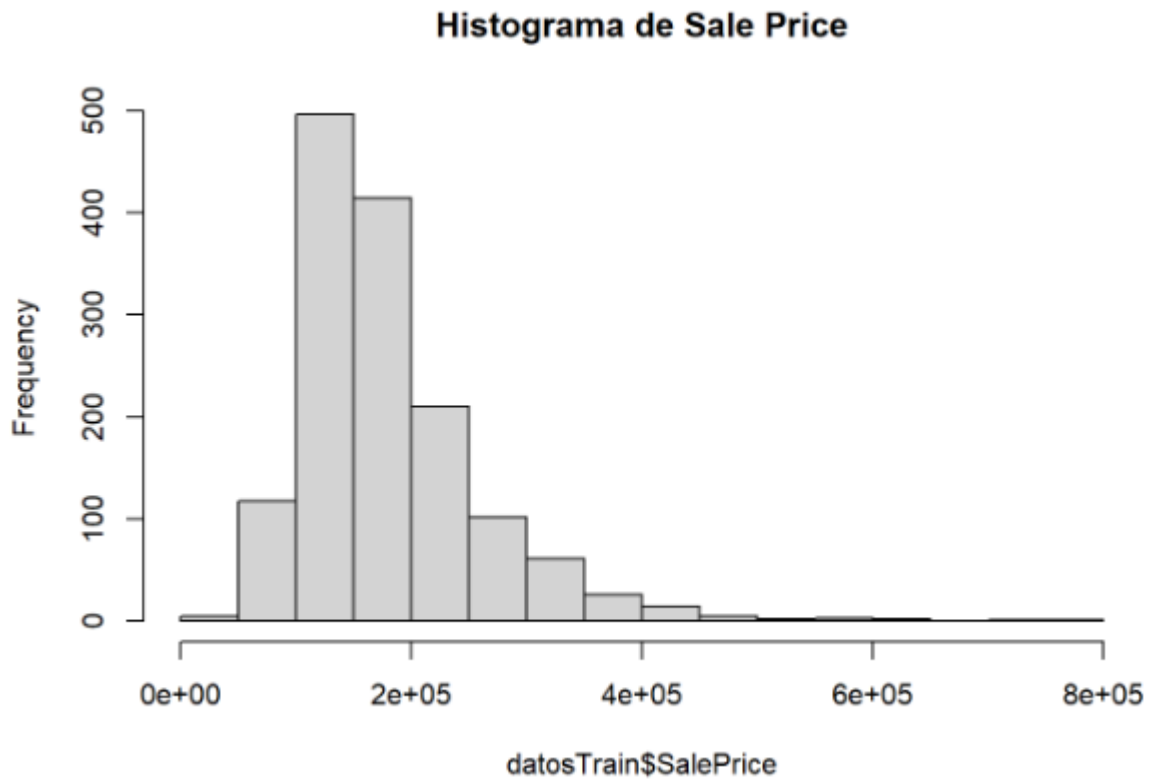

Histograma de Screen Porch Area in Square Feet
Frequency
0
200
400
600
800
1200
0
100
200
300
400
500
datosTrain$ScreenPorch


Histograma de Pool Area in Square Feet
Frequency
0
500
1000
1500
0
200
400
600
datosTrain$PoolArea





Sorprendentemente el mes vendido también parece tener una distribución normal.



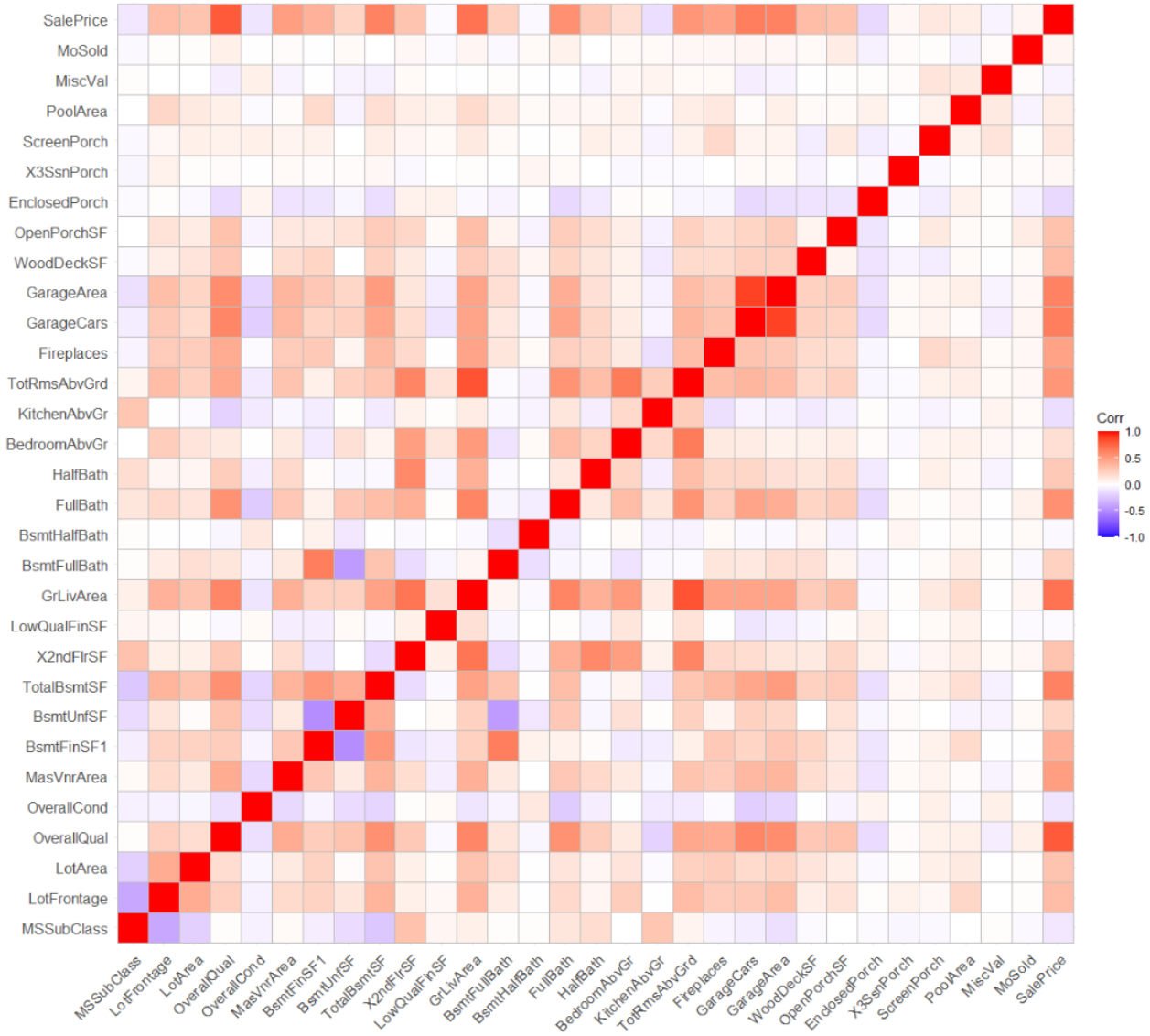
SalePrice también tiene una distribución normal.

Ya con los histogramas y las tablas de frecuencia, se pueden entender más los datos y se pudo observar que los datos con normalidad eran los siguientes: MonthSold, SalePrice, GarageArea, GarageCars, BedroomAbvGr, TotRmsAbvGrd, GrLivArea, X1stFlrSF, TotalBsmtSF, OverallQual, OverallCond, LotFrontage. Para un total de 12 variables.

4. Aísle las variables numéricas de las categóticas, haga un análisis de correlación entre las mismas.

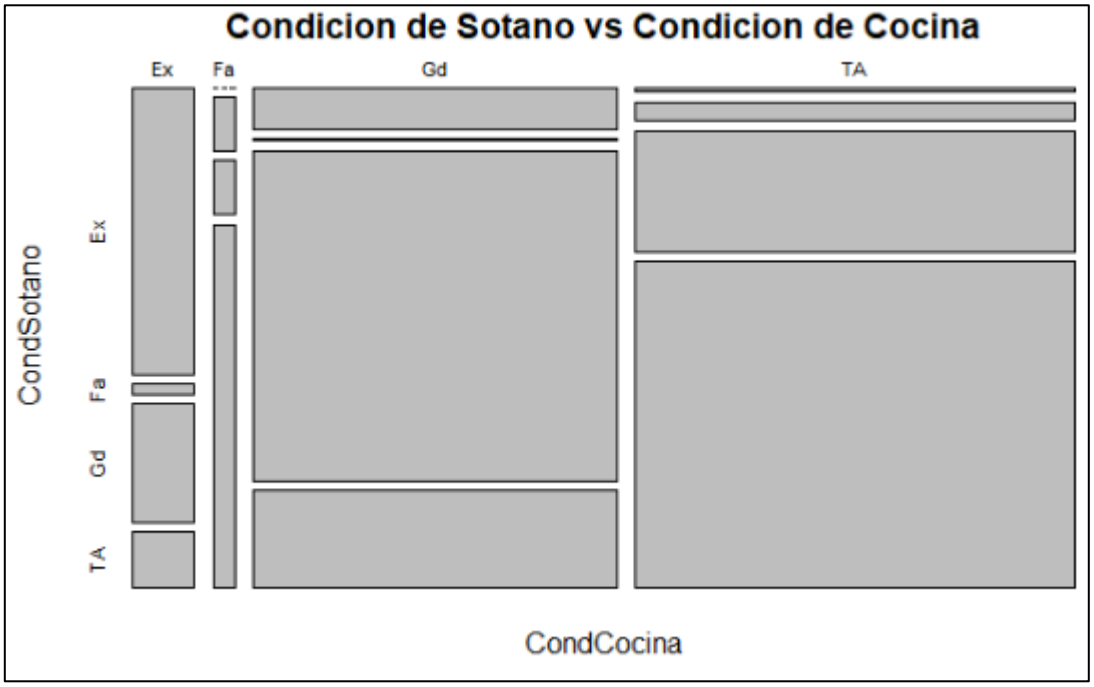
Se hizo una matriz de correlación con las variables numéricas y también se sacó el determinante de la misma el cuál es de: 1.1316e-07, que es básicamente 0. Esto indica que hay multicolinealidad entre las varibles y que más adelante será útil hacer un PCA, ya que para que un PCA sea efectivo, hay que tener correlacion de variables.

La matriz de correlacion cruda está en el RMarkDown, pero también se hizo un mapa de calor de correlación para verlo de una manera más facil.

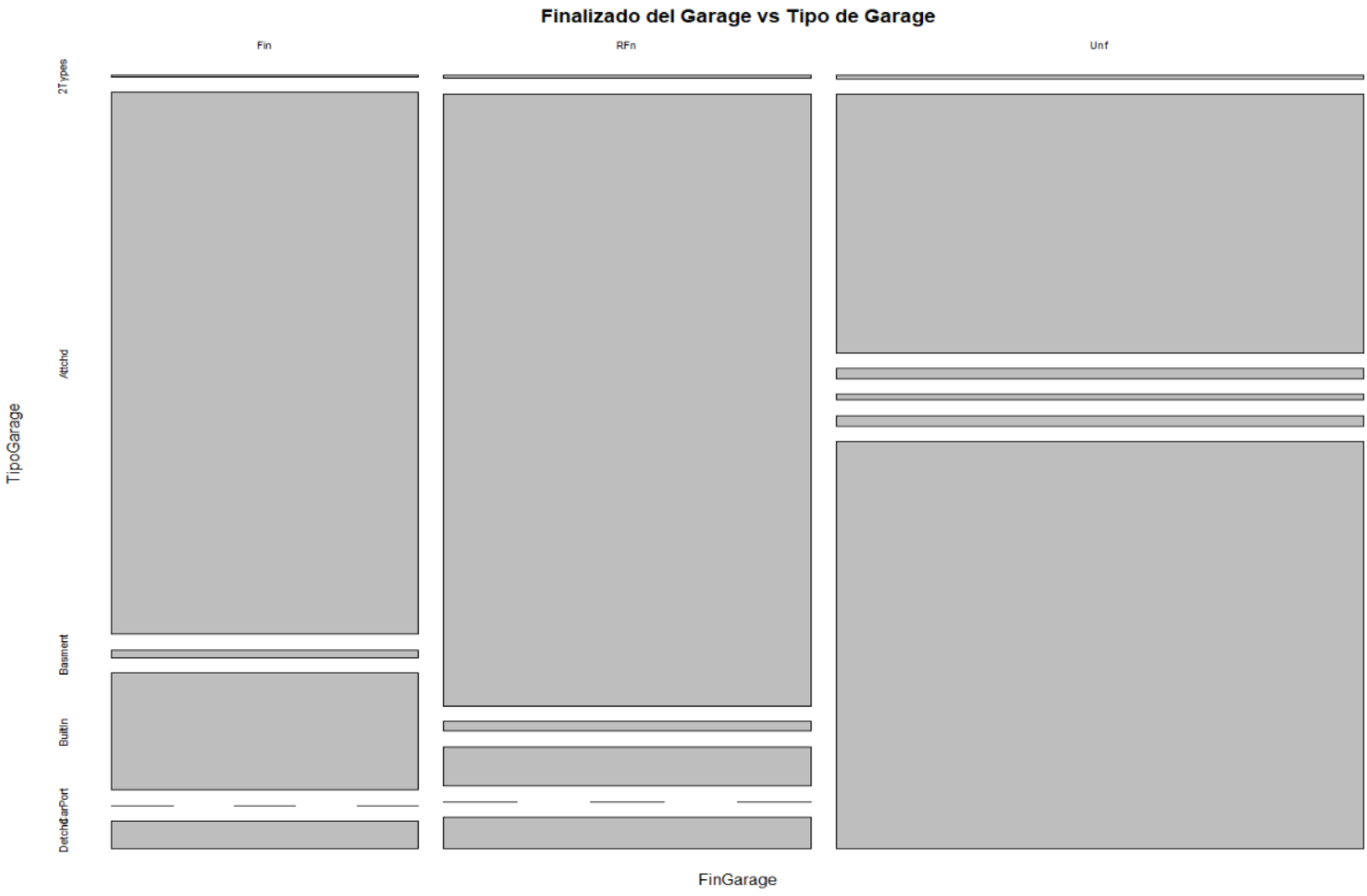


5. Utilice las variables categóricas, haga tablas de frecuencia, proporción, gráficas de barras o cualquier otra técnica que le permita explorar los datos

Como anteriormente, en el inciso 3, se muestran todas las tablas de frecuencia de las variables categóricas, ahora se buscará si hay relación entre algunos pares de variables que nos parezcan interesantes. Se comprobará la correlación con el algoritmo de Cramer-v



Se encontró una correlación moderada en la condición de las cocinas y la condición de los sótanos, ya que podemos observar que la mayoría de los sótanos excelentes (Ex) también son cocinas excelentes. El valor de Cramer-V es de 0.4214.



También se encontró una correlación moderada entre el finalizado del interior del garaje y el tipo del garaje ya que la mayoría de garajes sin finalizar (Unf) son garajes separados de la casa (Detch) y la mayoría de los que están junto a la casa o son parte de la casa (Attchd) estan Finalizados (Fin). Se obtuvo un valor Cramer-V de 0.4564.

6. Estudie si es conveniente hacer un Análisis de Componentes Principales. Recuerde que puede usar el índice KMO y el test de esfericidad de Bartlett. Haga un análisis de componentes principales con las variables numéricas, discuta los resultados e interprete los componentes.

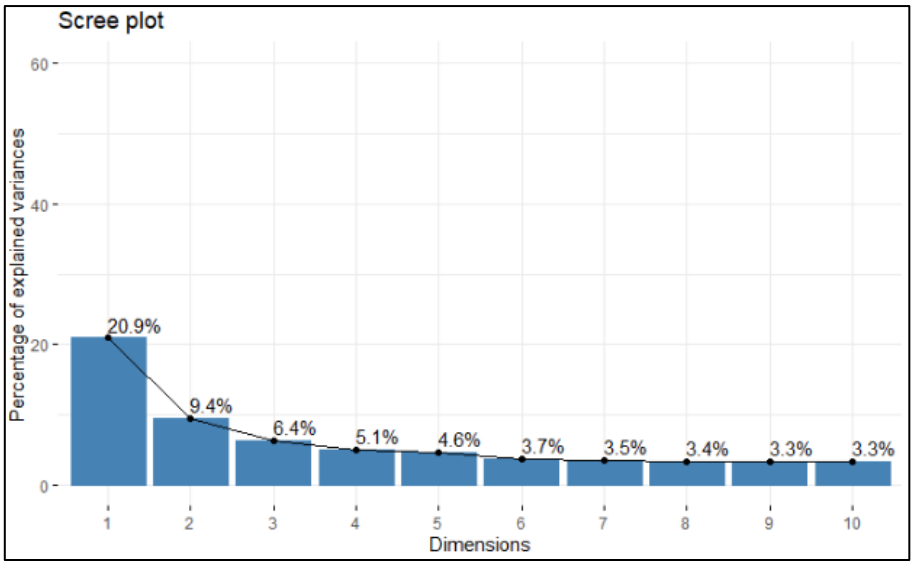
Para comenzar, nos podemos dar una idea de que sí es conveniente, ya que la determinante de la matriz de correlación era casi 0. Sin embargo, también se usó el índice de KMO y el test de esfericidad de Bartlett y el resultado fue de 0.78349 y 18919, respectivamente. Estos valores nos dicen que existe una aceptable adecuación muestral y es objetivamente útil hacer PCA.

```
> pafDatos$KMO
[1] 0.78349
> pafDatos$Bartlett
[1] 18919
```

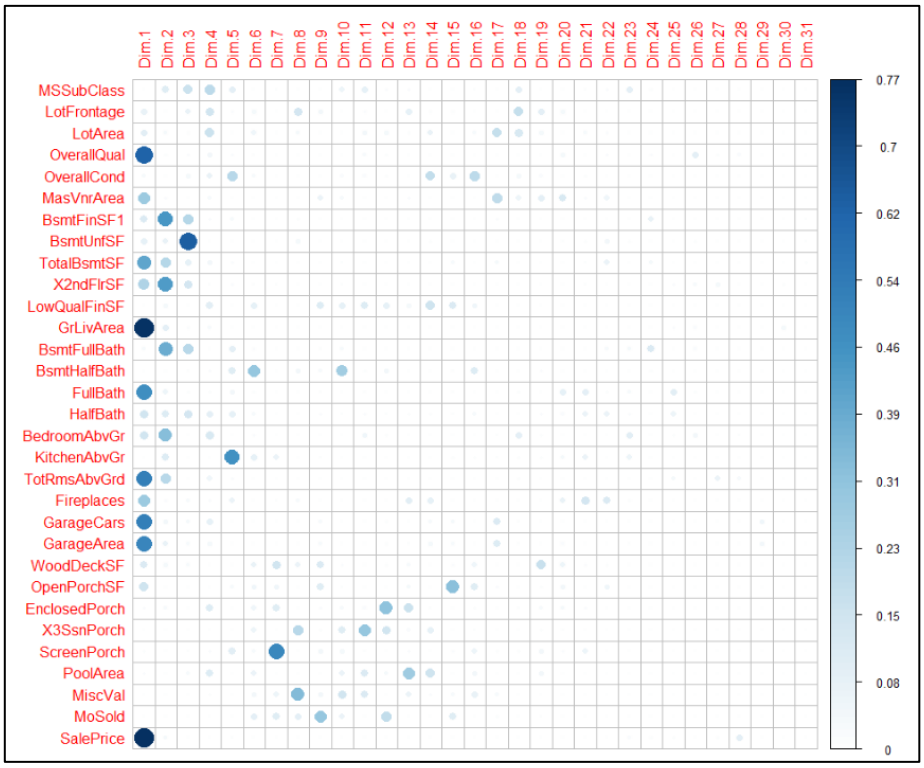
Aquí está el resumen de los componentes principales.

```
> summary(compPrinc)
Importance of components:
                          PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8    PC9   PC10   PC11
Standard deviation     2.547 1.7098 1.4034 1.2544 1.1934 1.0737 1.0476 1.027 1.0092 1.0070 1.0038
Proportion of Variance 0.209 0.0943 0.0635 0.0508 0.0459 0.0372 0.0354 0.034 0.0328 0.0327 0.0325
Cumulative Proportion  0.209 0.3036 0.3671 0.4178 0.4638 0.5010 0.5364 0.570 0.6033 0.6360 0.6685
                         PC12   PC13   PC14   PC15   PC16   PC17   PC18   PC19   PC20   PC21   PC22
Standard deviation     0.9849 0.9667 0.9556 0.9249 0.9137 0.9036 0.8475 0.8124 0.8043 0.7733 0.7387
Proportion of Variance 0.0313 0.0301 0.0295 0.0276 0.0269 0.0263 0.0232 0.0213 0.0209 0.0193 0.0176
Cumulative Proportion  0.6998 0.7299 0.7594 0.7870 0.8139 0.8403 0.8634 0.8847 0.9056 0.9249 0.9425
                         PC23   PC24    PC25    PC26    PC27    PC28    PC29   PC30    PC31
Standard deviation     0.6461 0.5627 0.52608 0.50872 0.40394 0.38395 0.31908 0.2549 0.19142
Proportion of Variance 0.0135 0.0102 0.00893 0.00835 0.00526 0.00476 0.00328 0.0021 0.00118
Cumulative Proportion  0.9559 0.9661 0.97507 0.98342 0.98868 0.99344 0.99672 0.9988 1.00000
```

Se hizo un screeplot y como se puede ver, el porcentaje deja de bajar significantemente en la dimensión número 3 (6.4%), por lo que podemos inferir que podríamos representar el data set con 3 componentes.

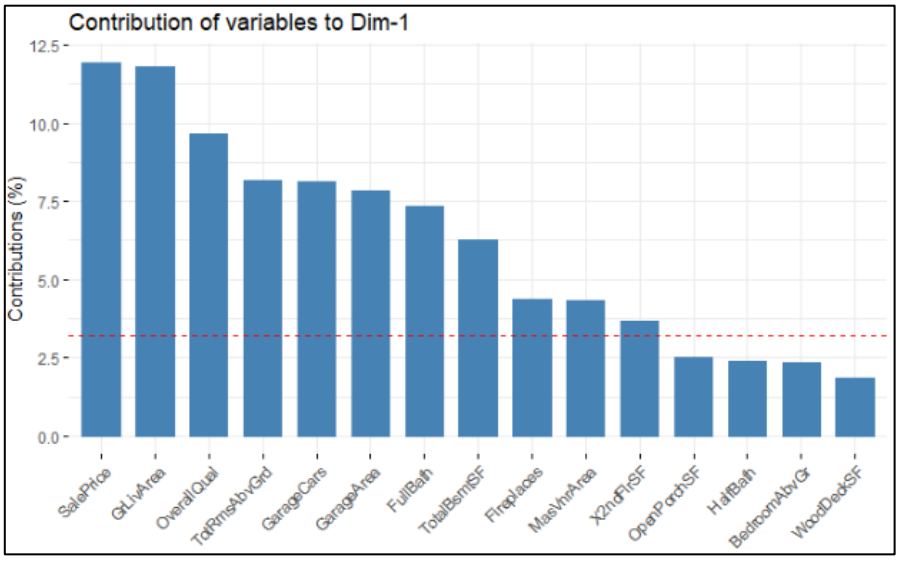


También se hizo el gráfico de la representación de cada variable en cada componente y podemos observar que efectivamente solamente 3 componentes es correcto, ya que después del tercero ya no hay una representación significativa de alguna variable



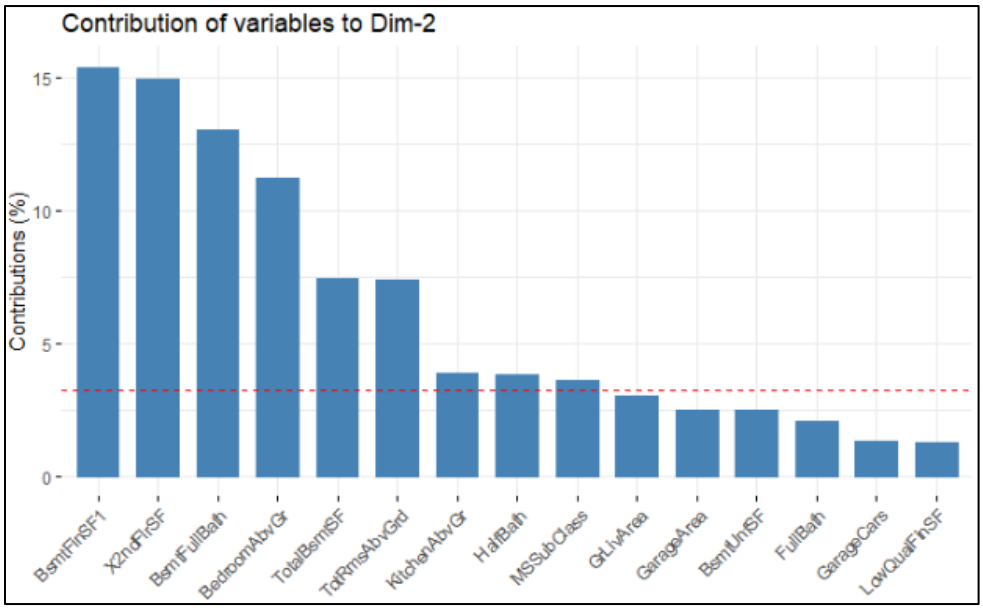
A continuación, se muestra la interpretación de los tres componentes principales:

Componente 1: **Calidad de la casa en general**



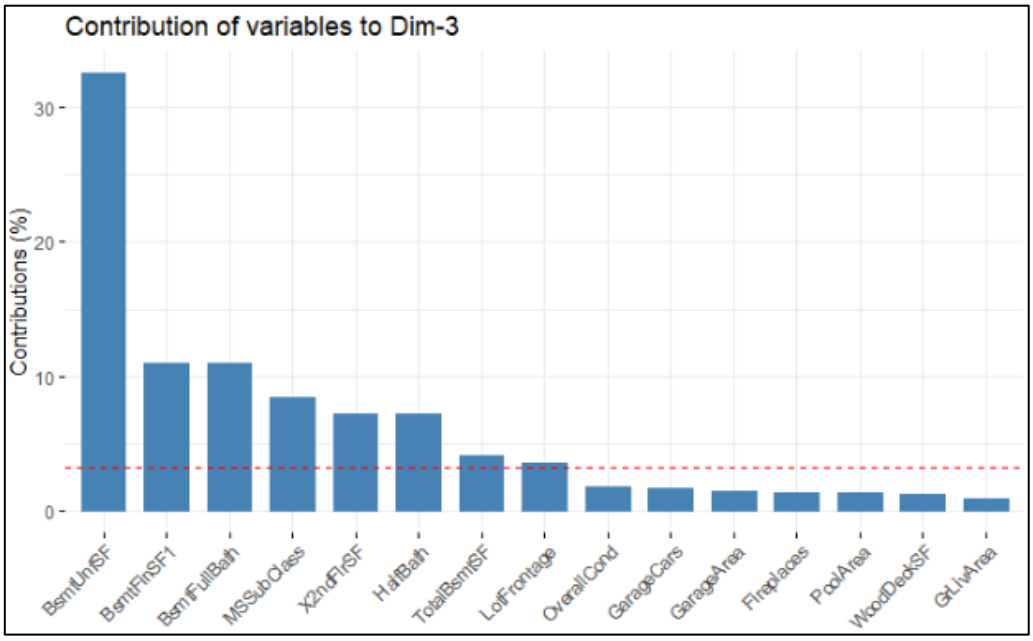
Las variables que mejor están representadas en el componente 1 son: SalePrice, GrLivArea, OverallQual, TotRmsAbvGrd, GarageCars, GarageArea, FullBath, TotalBsmtSF, FirePlaces, MasVnrArea, X2ndFlrSF. Al observar las primeras tres, las cuales son el precio de venta, el area en pies cuadrados de vivienda y la calidad en general. Se podría tomar este componente como la calidad de la casa en general, ya que por lo general una casa grande y que es cara, va a ser una casa de buena calidad.

Componente 2: **Calidad de áreas fuera del piso principal.**



Este componente esta principalmente representado por el finalizado del sótano (BsmtFinSF1), el área del segundo piso (X2ndFlrSF), la cantidad de baños completos del sótano (BsmtFullBath) y la cantidad de cuartos sin incluir los del sótano. No todas las casas tienen sótano o segundo piso, entonces este componente puede ser representado como calidad de áreas fuera del piso principal.

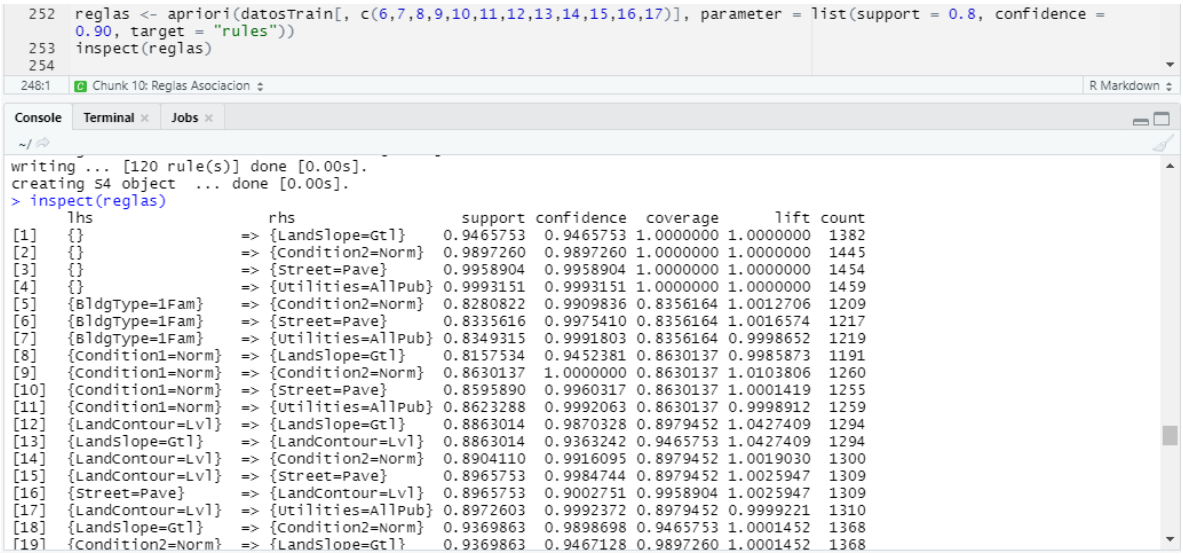
Componente 3: **Calidad del sótano**



El componente 3 está representado principalmente por el área sin finalizar del sótano en pies cuadrados (BsmtUnfSF) y el finalizado del sótano. No queda duda que este componente puede ser representado como calidad del sótano.

7. Obtenga reglas de asociación interesantes del dataset. Discuta sobre el nivel de confianza y soporte.

Se trabajaron solo con los datos cualitativos, así como se vio en clase. El nivel de confianza y soporte se colocaron más altos que en la clase para poder tener las reglas más interesantes. Se tuvo que trabajar por columnas (no todas en el mismo set ya que eran demasiadas columnas y R no lo soportaba). Como el dataset es tan extenso se crean una cantidad grande de reglas.

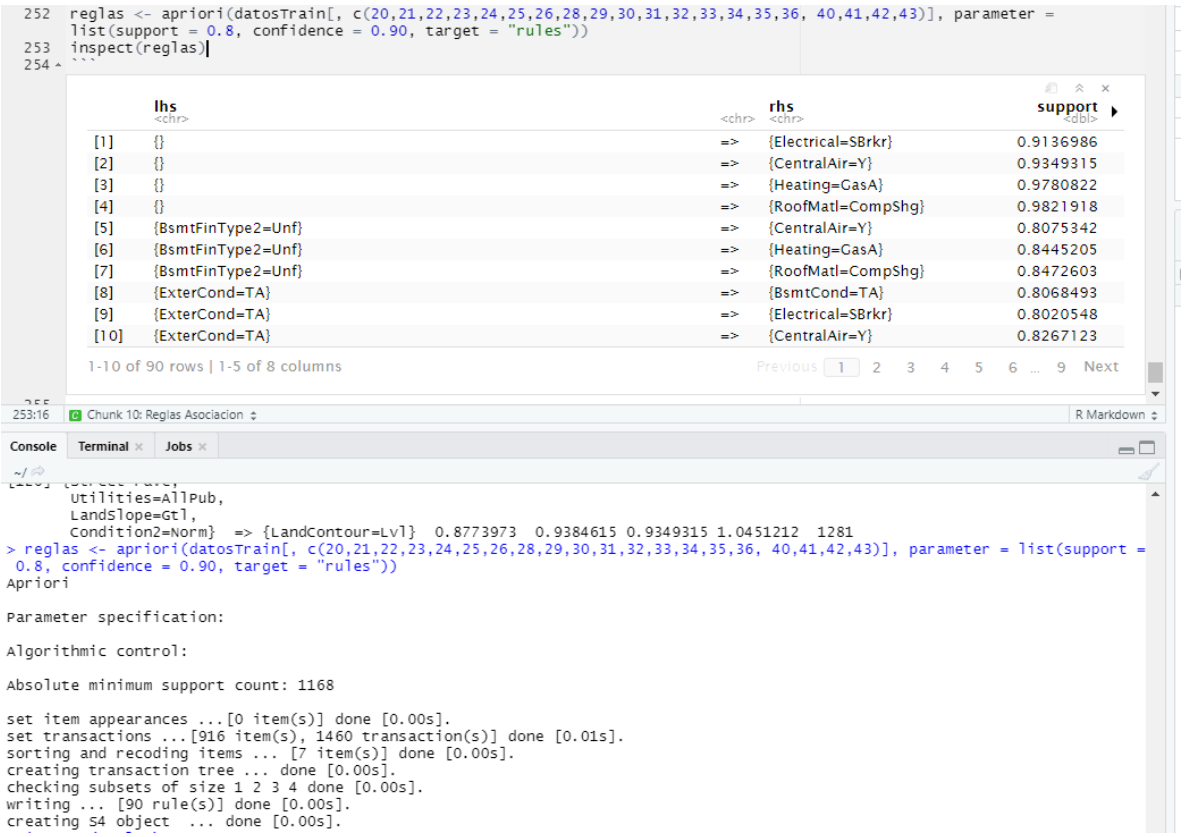


```
252  reglas <- apriori(datosTrain[, c(6,7,8,9,10,11,12,13,14,15,16,17)], parameter = list(support = 0.8, confidence =
     0.90, target = "rules"))
253  inspect(reglas)
254
```

248:1 Chunk 10: Reglas Asociacion — R Markdown

Console | Terminal | Jobs

```
writing ... [120 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
> inspect(reglas)
     lhs                    rhs                    support confidence  coverage      lift count
[1]  {}                  => {LandSlope=Gtl}     0.9465753  0.9465753 1.0000000 1.0000000  1382
[2]  {}                  => {Condition2=Norm}   0.9897260  0.9897260 1.0000000 1.0000000  1445
[3]  {}                  => {Street=Pave}       0.9958904  0.9958904 1.0000000 1.0000000  1454
[4]  {}                  => {Utilities=AllPub}  0.9993151  0.9993151 1.0000000 1.0000000  1459
[5]  {BldgType=1Fam}     => {Condition2=Norm}   0.8280822  0.9909836 0.8356164 1.0012706  1209
[6]  {BldgType=1Fam}     => {Street=Pave}       0.8335616  0.9975410 0.8356164 1.0016574  1217
[7]  {BldgType=1Fam}     => {Utilities=AllPub}  0.8349315  0.9991803 0.8356164 0.9998652  1219
[8]  {Condition1=Norm}   => {LandSlope=Gtl}     0.8157534  0.9452381 0.8630137 0.9985873  1191
[9]  {Condition1=Norm}   => {Condition2=Norm}   0.8630137  1.0000000 0.8630137 1.0103806  1260
[10] {Condition1=Norm}   => {Street=Pave}       0.8595890  0.9960317 0.8630137 1.0001419  1255
[11] {Condition1=Norm}   => {Utilities=AllPub}  0.8623288  0.9992063 0.8630137 0.9998912  1259
[12] {LandContour=Lvl}   => {LandSlope=Gtl}     0.8863014  0.9870328 0.8979452 1.0427409  1294
[13] {LandSlope=Gtl}     => {LandContour=Lvl}   0.8863014  0.9363242 0.9465753 1.0427409  1294
[14] {LandContour=Lvl}   => {Condition2=Norm}   0.8904110  0.9916095 0.8979452 1.0019030  1300
[15] {LandContour=Lvl}   => {Street=Pave}       0.8965753  0.9984744 0.8979452 1.0025947  1309
[16] {Street=Pave}       => {LandContour=Lvl}   0.8965753  0.9002751 0.9958904 1.0025947  1309
[17] {LandContour=Lvl}   => {Utilities=AllPub}  0.8972603  0.9992372 0.8979452 0.9999221  1310
[18] {LandSlope=Gtl}     => {Condition2=Norm}   0.9369863  0.9898698 0.9465753 1.0001452  1368
[19] {Condition2=Norm}   => {LandSlope=Gtl}     0.9369863  0.9467128 0.9897260 1.0001452  1368
```



```
252  reglas <- apriori(datosTrain[, c(20,21,22,23,24,25,26,28,29,30,31,32,33,34,35,36, 40,41,42,43)], parameter =
     list(support = 0.8, confidence = 0.90, target = "rules"))
253  inspect(reglas)
254
```

	lhs <chr>		rhs <chr>	support <dbl>
[1]	{}	=>	{Electrical=SBrkr}	0.9136986
[2]	{}	=>	{CentralAir=Y}	0.9349315
[3]	{}	=>	{Heating=GasA}	0.9780822
[4]	{}	=>	{RoofMatl=CompShg}	0.9821918
[5]	{BsmtFinType2=Unf}	=>	{CentralAir=Y}	0.8075342
[6]	{BsmtFinType2=Unf}	=>	{Heating=GasA}	0.8445205
[7]	{BsmtFinType2=Unf}	=>	{RoofMatl=CompShg}	0.8472603
[8]	{ExterCond=TA}	=>	{BsmtCond=TA}	0.8068493
[9]	{ExterCond=TA}	=>	{Electrical=SBrkr}	0.8020548
[10]	{ExterCond=TA}	=>	{CentralAir=Y}	0.8267123

1-10 of 90 rows | 1-5 of 8 columns — Previous 1 2 3 4 5 6 ... 9 Next

253:16 Chunk 10: Reglas Asociacion — R Markdown

Console | Terminal | Jobs

```
       Utilities=AllPub,
       LandSlope=Gtl,
       Condition2=Norm}  => {LandContour=Lvl}  0.8773973  0.9384615 0.9349315 1.0451212  1281
> reglas <- apriori(datosTrain[, c(20,21,22,23,24,25,26,28,29,30,31,32,33,34,35,36, 40,41,42,43)], parameter = list(support =
 0.8, confidence = 0.90, target = "rules"))
Apriori

Parameter specification:

Algorithmic control:

Absolute minimum support count: 1168

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[916 item(s), 1460 transaction(s)] done [0.01s].
sorting and recoding items ... [7 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [90 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

```
252 reglas <- apriori(datosTrain[, c(54,56,58,59,60,61,64,65,66,73,74,78,79,80)], parameter = list(support = 0.8,
    confidence = 0.90, target = "rules"))
253 inspect(reglas)
254 ```
```

	lhs		rhs	support	confidence	coverage
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
[1]	{}	=>	{PavedDrive=Y}	0.9178082	0.9178082	1.0000000
[2]	{}	=>	{Functional=Typ}	0.9315068	0.9315068	1.0000000
[3]	{}	=>	{GarageCond=TA}	0.9082192	0.9082192	1.0000000
[4]	{SaleType=WD}	=>	{Functional=Typ}	0.8041096	0.9265983	0.8678082
[5]	{GarageQual=TA}	=>	{PavedDrive=Y}	0.8527397	0.9496568	0.8979452
[6]	{PavedDrive=Y}	=>	{GarageQual=TA}	0.8527397	0.9291045	0.9178082
[7]	{GarageQual=TA}	=>	{Functional=Typ}	0.8404110	0.9359268	0.8979452
[8]	{Functional=Typ}	=>	{GarageQual=TA}	0.8404110	0.9022059	0.9315068
[9]	{GarageQual=TA}	=>	{GarageCond=TA}	0.8842466	0.9847445	0.8979452
[10]	{GarageCond=TA}	=>	{GarageQual=TA}	0.8842466	0.9736048	0.9082192

1-10 of 25 rows | 1-7 of 8 columns

Previous 1 2 3 Next

254:1 Chunk 10: Reglas Asociacion R Markdown

Console Terminal Jobs

~/

```
writing ... [90 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
> inspect(reglas)
> reglas <- apriori(datosTrain[, c(54,56,58,59,60,61,64,65,66,73,74,78,79,80)], parameter = list(support = 0.8, confidence =
 0.90, target = "rules"))
Apriori

Parameter specification:

Algorithmic control:

Absolute minimum support count: 1168

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[162 item(s), 1460 transaction(s)] done [0.00s].
sorting and recoding items ... [6 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [25 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```