

Laboratorio No.1

Paul Belches

José Pablo Cifuentes

Oscar Juárez

25 de julio, 2020

Clasificación de las variables

MSSubClass

Identifica el tipo de vivienda involucrada en la venta.

Variable categórica nominal

MSZoning

Identifica la clasificación general de la zona de venta

Variable categórica nominal

LotFrontage

Cantidad de pies de la propiedad conectados a la calle

Variable cuantitativa continua

LotArea

Tamaño del lote en pies cuadrados

Variable cuantitativa continua

Street

Tipo del camino de acceso a la propiedad

Variable categórica nominal

Alley

Tipo del callejón de acceso a la propiedad

Variable categórica nominal

LotShape

Forma general de la propiedad

Variable categórica nominal

LandContour

Horizontalidad de la propiedad

Variable categórica ordinal

Utilities

Tipo de utilidades disponibles

Variable categórica nominal

LotConfig

Configuración del lote

Variable categórica nominal

LandSlope

Pendiente de la propiedad

Variable categórica ordinal

Neighborhood

Ubicación dentro de la ciudad de Ames

Variable categórica nominal

Condition1

Proximidad a condiciones varias

Variable categórica nominal

Condition2

Proximidad a condiciones varias (si existe más de una presente)

Variable categórica nominal

BldgType

Tipo de vivienda

Variable categórica nominal

HouseStyle

Estilo de vivienda

Variable categórica nominal

OverallQual

Calificación promedio de los materiales y acabados de la casa

Variable categórica ordinal

OverallCond

Calificación promedio de la condición de la casa

Variable categórica ordinal

YearBuilt

Fecha original de construcción

Variable categórica ordinal

YearRemodAdd

Fecha de remodelación (misma que en construcción en caso no se hayan efectuado adiciones o remodelaciones)

Variable categórica ordinal

RoofStyle

Tipo de techo

Variable categórica nominal

RoofMatl

Material del techo

Variable categórica nominal

Exterior1st

Revestimiento de la casa

Variable categórica nominal

Exterior2nd

Revestimiento de la casa (en caso exista más de uno)

Variable categórica nominal

MasVnrType

Tipo de mampostería

Variable categórica nominal

MasVnrArea

Área de mampostería en pies cuadrados

Variable cuantitativa continua

ExterQual

Evaluación de la calidad de los materiales del exterior

Variable categórica ordinal

ExterCond

Evaluación de la condición presente de los materiales del exterior

Variable categórica ordinal

Foundation

Tipo de cimentación

Variable categórica nominal

BsmtQual

Evaluación de la altura del sótano

Variable categórica ordinal

BsmtCond

Evaluación general de la condición del sótano

Variable categórica ordinal

BsmtExposure

Evaluación de la pared perimetral

Variable categórica ordinal

BsmtFinType1

Rating of basement finished area

Variable categórica ordinal

BsmtFinSF1

Acabados de tipo 1 en pies cuadrados

Variable cuantitativa continua

BsmtFinType2

Calificación de los acabados del sótano (si existe varios)

Variable categórica ordinal

BsmtFinSF2

Acabados de tipo 2 en pies cuadrados

Variable cuantitativa continua

BsmtUnfSF

Pies cuadrados sin terminar en el área del sótano

Variable cuantitativa continua

TotalBsmtSF

Cantidad total de pies cuadrados en el área del sótano

Variable cuantitativa continua

Heating

Tipo de calefacción

Variable categórica nominal

HeatingQC

Calidad y condición de la calefacción

Variable categórica ordinal

CentralAir

Aire acondicionado central

Variable categórica nominal

Electrical

Sistema eléctrico

Variable categórica nominal

1stFlrSF

Pies cuadrados del primer piso

Variable cuantitativa continua

2ndFlrSF

Pies cuadrados del segundo piso

Variable cuantitativa continua

LowQualFinSF

Cantidad de pies cuadrados con acabados de baja calidad

Variable cuantitativa continua

GrLivArea

Cantidad de pies cuadrados de vivienda sobre la superficie

Variable cuantitativa continua

BsmtFullBath

Cantidad de baños completos en el sótano

Variable cuantitativa discreta

BsmtHalfBath

Cantidad de medios baños en el sótano

Variable cuantitativa discreta

FullBath

Cantidad de baños completos sobre la superficie

Variable cuantitativa discreta

HalfBath

Cantidad de medios baños sobre la superficie

Variable cuantitativa discreta

Bedroom

Cantidad de dormitorios sobre la superficie (no incluye dormitorios en el sótano)

Variable cuantitativa discreta

Kitchen

Cantidad de cocinas sobre la superficie

Variable cuantitativa discreta

KitchenQual

Calidad de la cocina

Variable categórica ordinal

TotRmsAbvGrd

Cantidad de cuartos sobre la superficie (no incluye baños)

Variable cuantitativa discreta

Functional

Funcionalidad de casas (Se asumen típicas a menos que la deducción este justificado)

Variable categórica nominal

Fireplaces

Cantidad de chimeneas

Variable cuantitativa discreta

FireplaceQu

Calidad de la chimenea

Variable categórica ordinal

GarageType

Locación del garaje

Variable categórica nominal

GarageYrBlt

Año en que se construyó el garaje

Variable categórica ordinal

GarageFinish

Acabados interiores del garaje

Variable categórica ordinal

GarageCars

Tamaño del garaje en capacidad de carros

Variable cuantitativa discreta

GarageArea

Tamaño del garaje en pies cuadrados

Variable cuantitativa continua

GarageQual

Calidad del garaje

Variable categórica ordinal

GarageCond

Condiciones del garaje

Variable categórica ordinal

PavedDrive

Pavimento de la entrada

Variable categórica nominal

WoodDeckSF

Área del deck de madera en pies cuadrados

Variable cuantitativa continua

OpenPorchSF

Área abierta del porche en pies cuadrados

Variable cuantitativa continua

EnclosedPorch

Área cerrada del porche en pies cuadrados

Variable cuantitativa continua

3SsnPorch

Área de 3 temporadas del porche en pies cuadrados

Variable cuantitativa continua

ScreenPorch

Área de pantalla del porche en pies cuadrados

Variable cuantitativa continua

PoolArea

Área de piscina en pies cuadrados

Variable cuantitativa continua

PoolQC

Calidad de la piscina

Variable categórica ordinal

Fence

Calidad de la cerca

Variable categórica ordinal

MiscFeature

Otras características no cubiertas en otra categoría

Variable categórica nominal

MiscVal

Valor de las otras características

Variable categórica nominal

MoSold

Mes en que se vendió

Variable categórica ordinal

YrSold

Año en que se vendió

Variable categórica ordinal

SaleType

Tipo de venta

Variable categórica nominal

SaleCondition

Condiciones de venta

Variable categórica nominal

Análisis de variables cuantitativas

Análisis de correlación

Se seleccionan las variables que son numéricas. Asignando 0 en los lugares en donde se encuentra un Na.

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
datosNumericos <- select(dataSet, LotArea, MasVnrArea, BsmtFinSF1, BsmtFinSF2, BsmtUnfSF, TotalBsmtSF, X1stFlrSF, X2ndFlrSF, LowQualFinSF, GrLivArea, BsmtFullBath, BsmtHalfBath, FullBath, HalfBath, TotRmsAbvGrd, Fireplaces, GarageCars)
```

```
datosNumericos[is.na(datosNumericos)] <- 0
```

```
datosNumericos2 <- datosNumericos
```

```
View(datosNumericos2)
```

```
datosNumericos$LotArea <- as.numeric(datosNumericos$LotArea)
```

```
datosNumericos$MasVnrArea <- as.numeric(datosNumericos$MasVnrArea)
```

```
## Warning: NAs introduced by coercion
```

```
datosNumericos$BsmtFinSF1 <- as.numeric(datosNumericos$BsmtFinSF1)
```

```
datosNumericos$BsmtFinSF2 <- as.numeric(datosNumericos$BsmtFinSF2)
```

```
datosNumericos$BsmtUnfSF <- as.numeric(datosNumericos$BsmtUnfSF)
```

```
datosNumericos$TotalBsmtSF <- as.numeric(datosNumericos$TotalBsmtSF)
```

```
datosNumericos$X1stFlrSF <- as.numeric(datosNumericos$X1stFlrSF)
```

```
datosNumericos$X2ndFlrSF <- as.numeric(datosNumericos$X2ndFlrSF)
```

```
datosNumericos$LowQualFinSF <- as.numeric(datosNumericos$LowQualFinSF)
```

```
datosNumericos$GrLivArea <- as.numeric(datosNumericos$GrLivArea)
```

```
datosNumericos$BsmtFullBath <- as.numeric(datosNumericos$BsmtFullBath)
```

```
datosNumericos$BsmtHalfBath <- as.numeric(datosNumericos$BsmtHalfBath)
```

```
datosNumericos$FullBath <- as.numeric(datosNumericos$FullBath)
```

```
datosNumericos$HalfBath <- as.numeric(datosNumericos$HalfBath)
```

```
datosNumericos$TotRmsAbvGrd <- as.numeric(datosNumericos$TotRmsAbvGrd)
```

```
datosNumericos$Fireplaces <- as.numeric(datosNumericos$Fireplaces)
```

```
datosNumericos$GarageCars <- as.numeric(datosNumericos$GarageCars)
```

```

datosNumericos$GarageArea <- as.numeric(datosNumericos$GarageArea)
datosNumericos$WoodDeckSF <- as.numeric(datosNumericos$WoodDeckSF)
datosNumericos$OpenPorchSF <- as.numeric(datosNumericos$OpenPorchSF)
datosNumericos$EnclosedPorch <- as.numeric(datosNumericos$EnclosedPorch)
datosNumericos$X3SsnPorch <- as.numeric(datosNumericos$X3SsnPorch)
datosNumericos$ScreenPorch <- as.numeric(datosNumericos$ScreenPorch)
datosNumericos$PoolArea <- as.numeric(datosNumericos$PoolArea)
View(datosNumericos)

```

```

correlacion <- cor(datosNumericos)

```

```

library(corrplot)

```

```

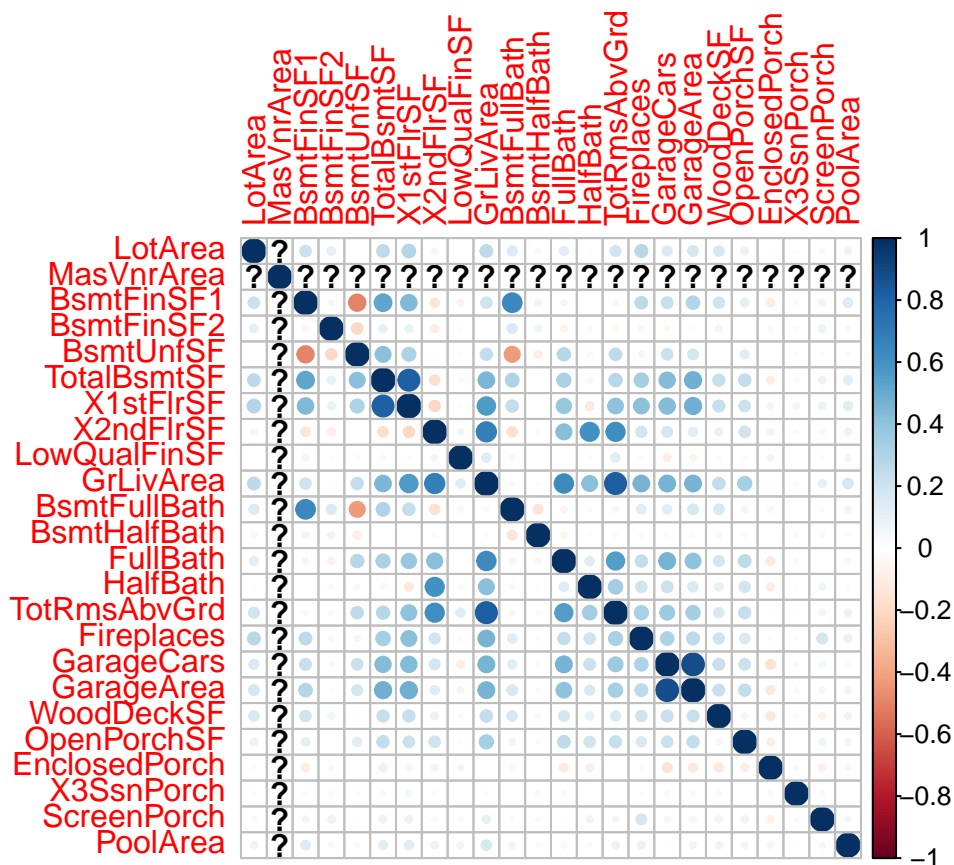
## corrplot 0.84 loaded

```

```

corrplot(correlacion)

```

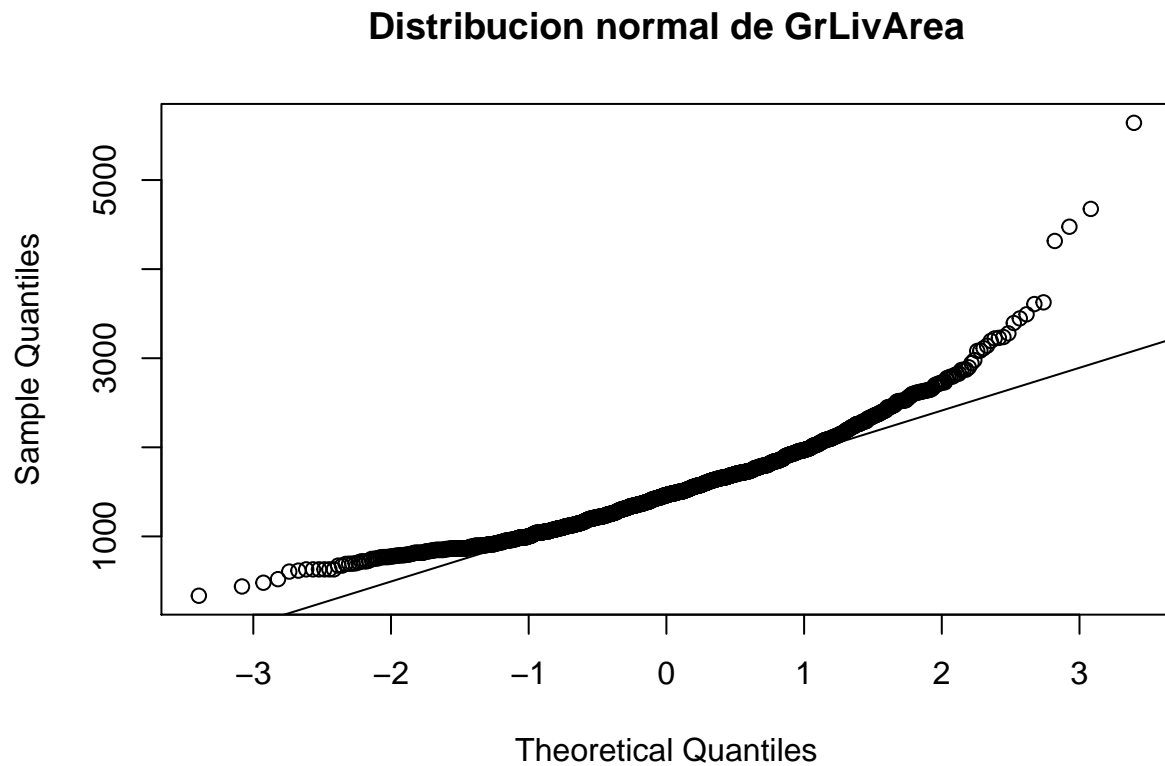


Como podemos observar existe una correlacion alta entre la cantidad de pies cuadrados del primer piso y el tamaño del sótano, la capacidad en carros del garaje y el área del garaje , y la cantidad de cuartos sobre la superficie y la cantidad de mtros cuadrados sobre la superficie. De alguna manera podemos decir que esto es de esperarse gracias a que estas son variables que describen una misma sección de la casa apartir de dos enfoques diferentes. Este mismo argumento se puede utilizar para las variables inversamente proporcionales. Por ejemplo en el caso de la cantidad de metros cuadrados terminados y sin terminar en el sotano, encontramos esta relación inversamente proporcional tan alta se debe a que, ambas describen el estado del sótano de maneras distintas.

Análisis de normalidad

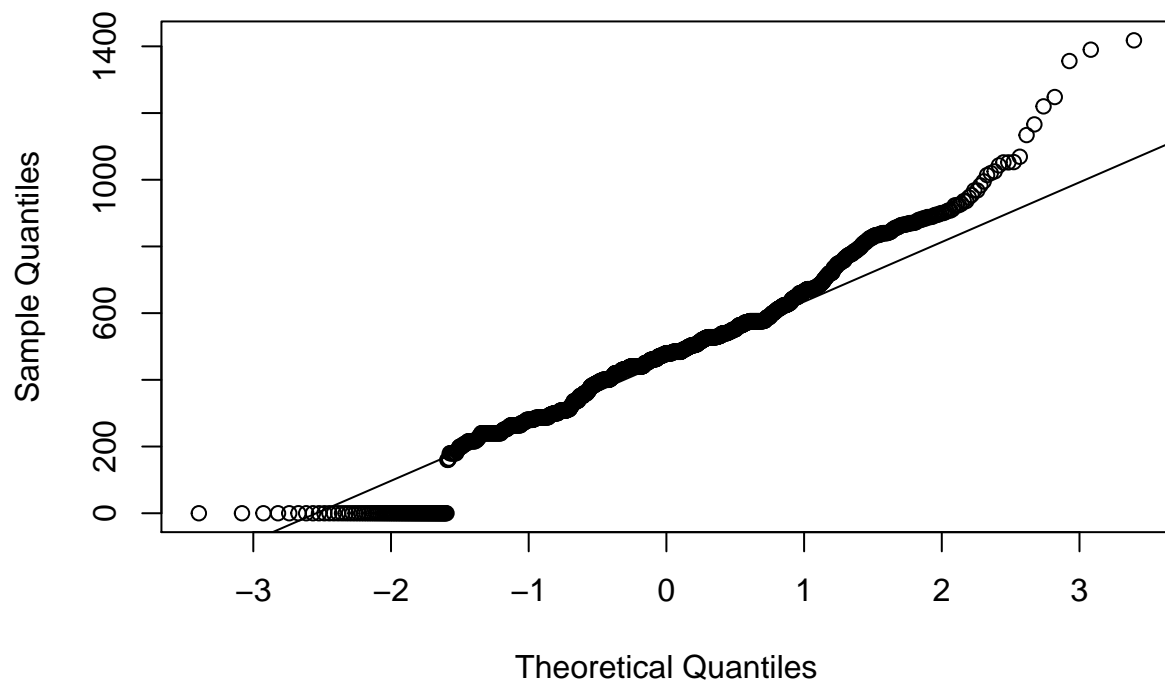
A continuación, se realizan gráficas Q-Q (quantile-quantile) para conocer que tan normales son nuestras variables de tipo cuantitativo. Lo ideal es que los puntos se acerquen a una recta diagonal.

```
qqnorm(na.omit(as.numeric(dataSet$GrLivArea)),main = "Distribucion normal de GrLivArea")
qqline(na.omit(as.numeric(dataSet$GrLivArea)))
```



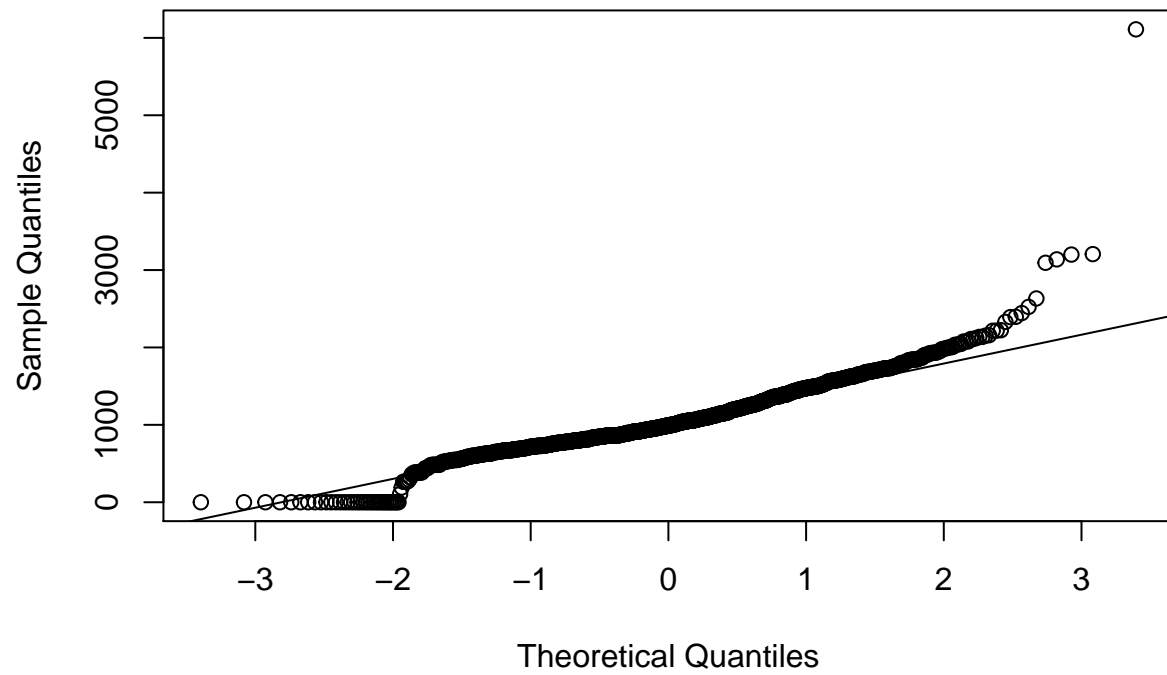
```
qqnorm(na.omit(as.numeric(dataSet$GarageArea)),main = "Distribucion normal de GarageArea")
qqline(na.omit(as.numeric(dataSet$GarageArea)))
```

Distribucion normal de GarageArea



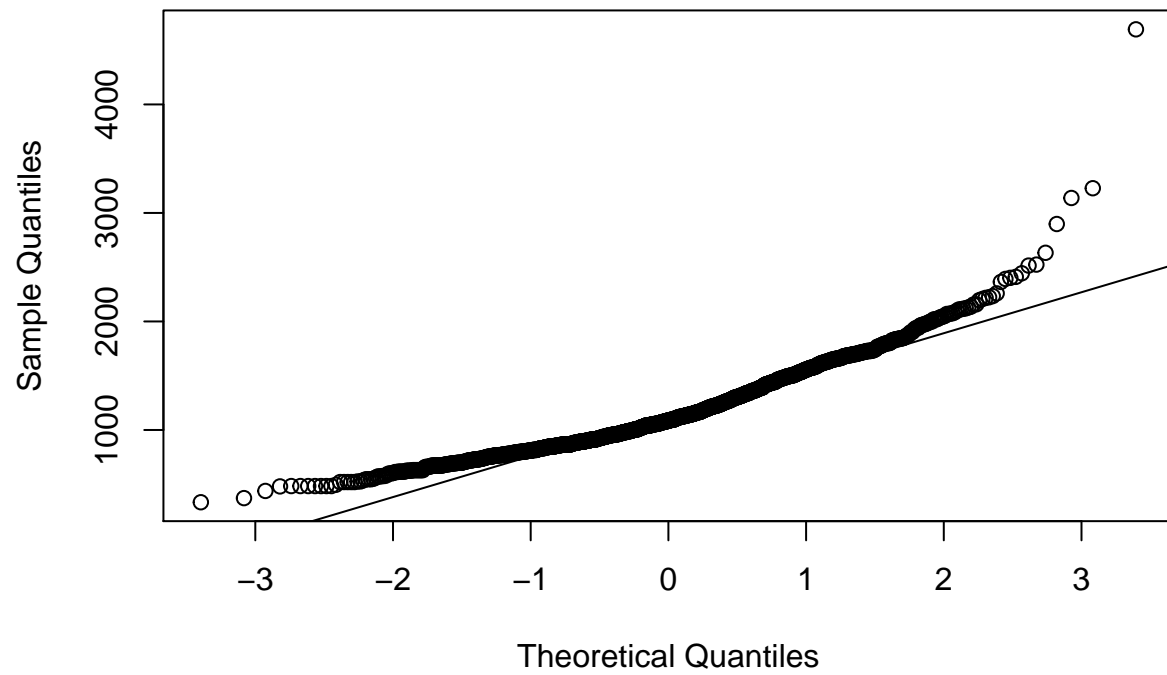
```
qqnorm(na.omit(as.numeric(dataSet$TotalBsmtSF)),main = "Distribucion normal de TotalBsmtSF")
qqline(na.omit(as.numeric(dataSet$TotalBsmtSF)))
```

Distribucion normal de TotalBsmtSF



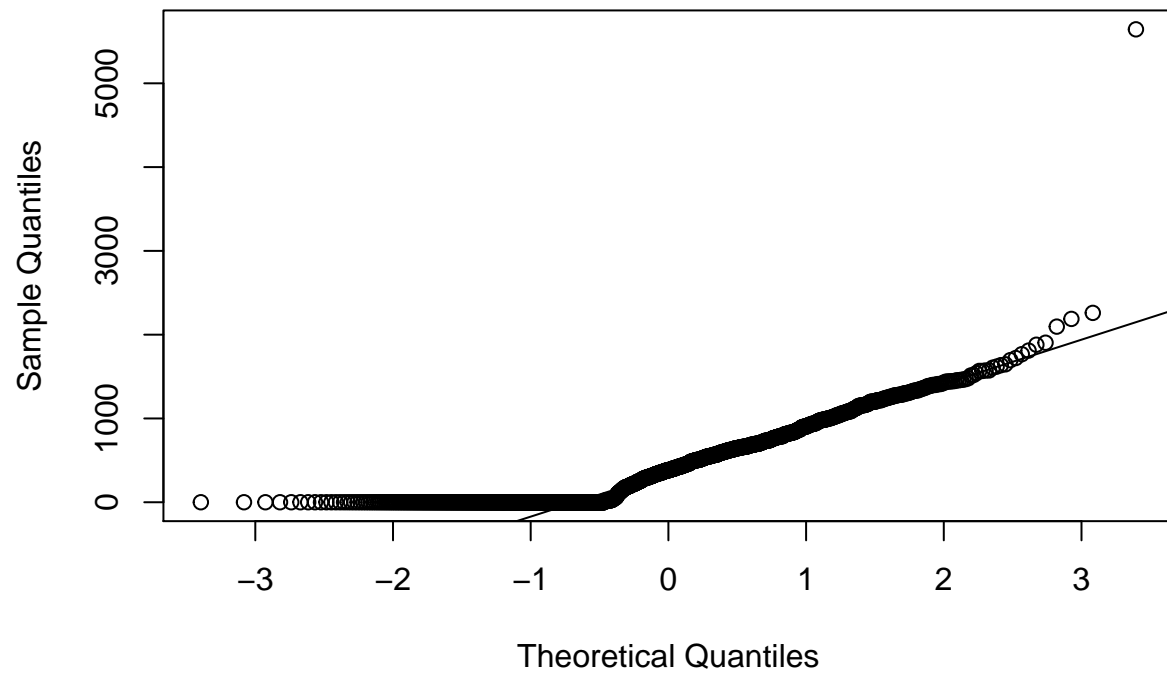
```
qqnorm(na.omit(as.numeric(dataSet$X1stFlrSF)),main = "Distribucion normal de X1stFlrSF")  
qqline(na.omit(as.numeric(dataSet$X1stFlrSF)))
```

Distribucion normal de X1stFlrSF



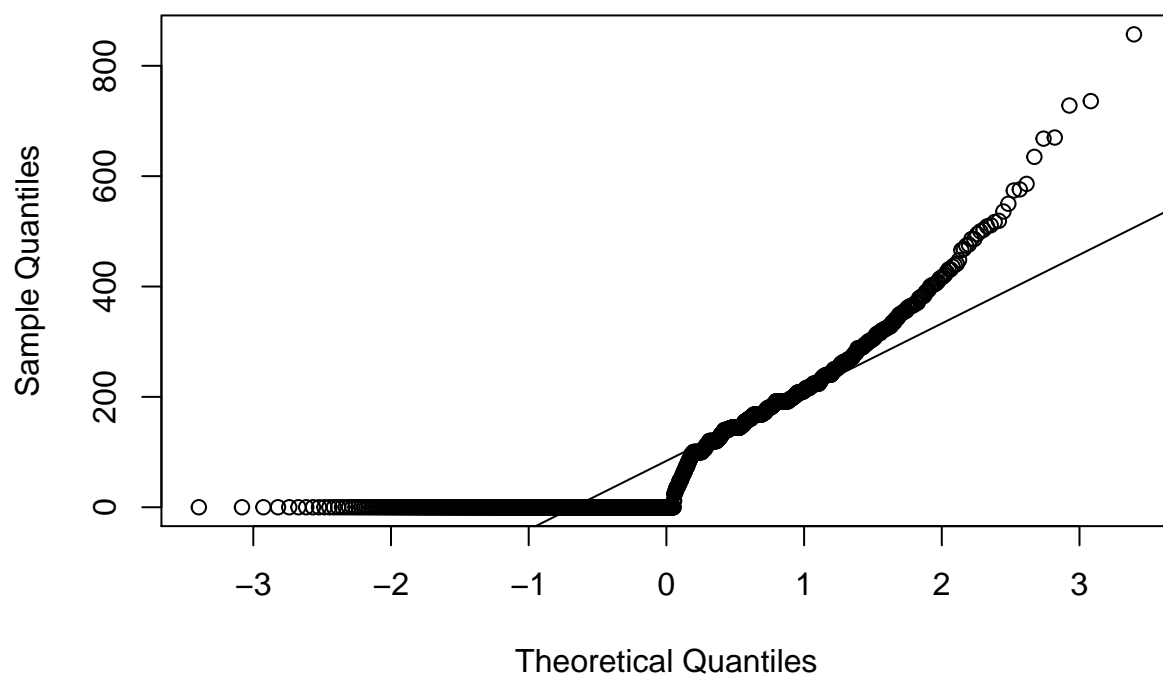
```
qqnorm(na.omit(as.numeric(dataSet$BsmtFinSF1)),main = "Distribucion normal de BsmtFinSF1")  
qqline(na.omit(as.numeric(dataSet$BsmtFinSF1)))
```

Distribucion normal de BsmtFinSF1



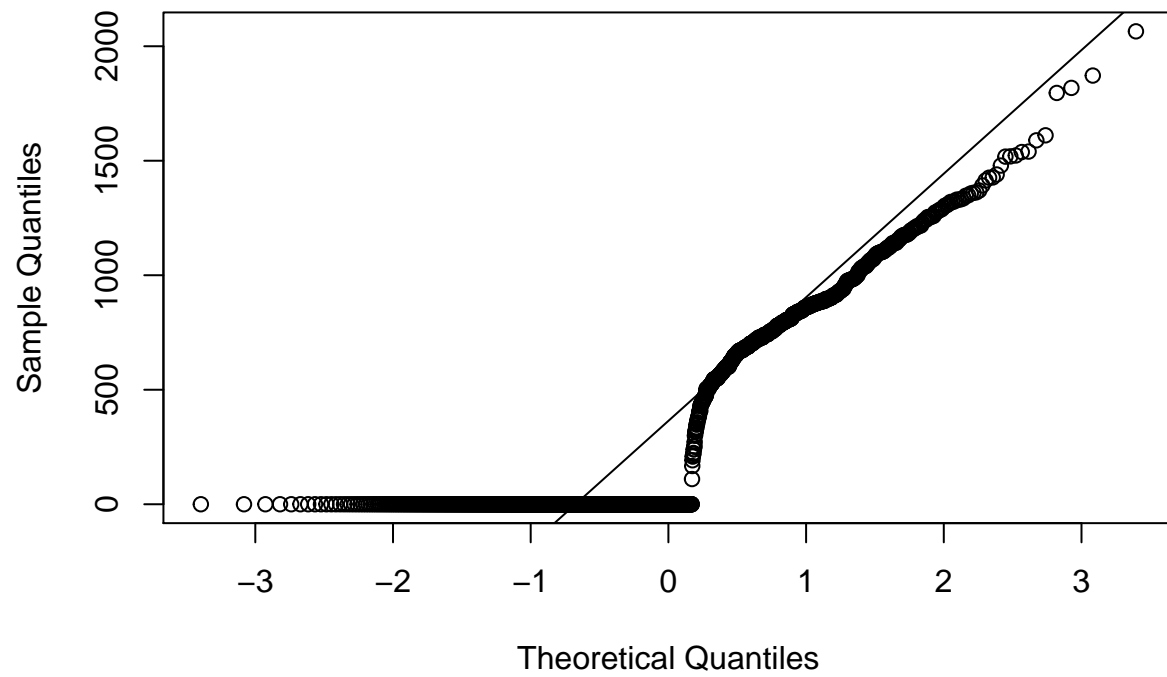
```
qqnorm(na.omit(as.numeric(dataSet$WoodDeckSF)),main = "Distribucion normal de WoodDeckSF")
qqline(na.omit(as.numeric(dataSet$WoodDeckSF)))
```


Distribucion normal de WoodDeckSF



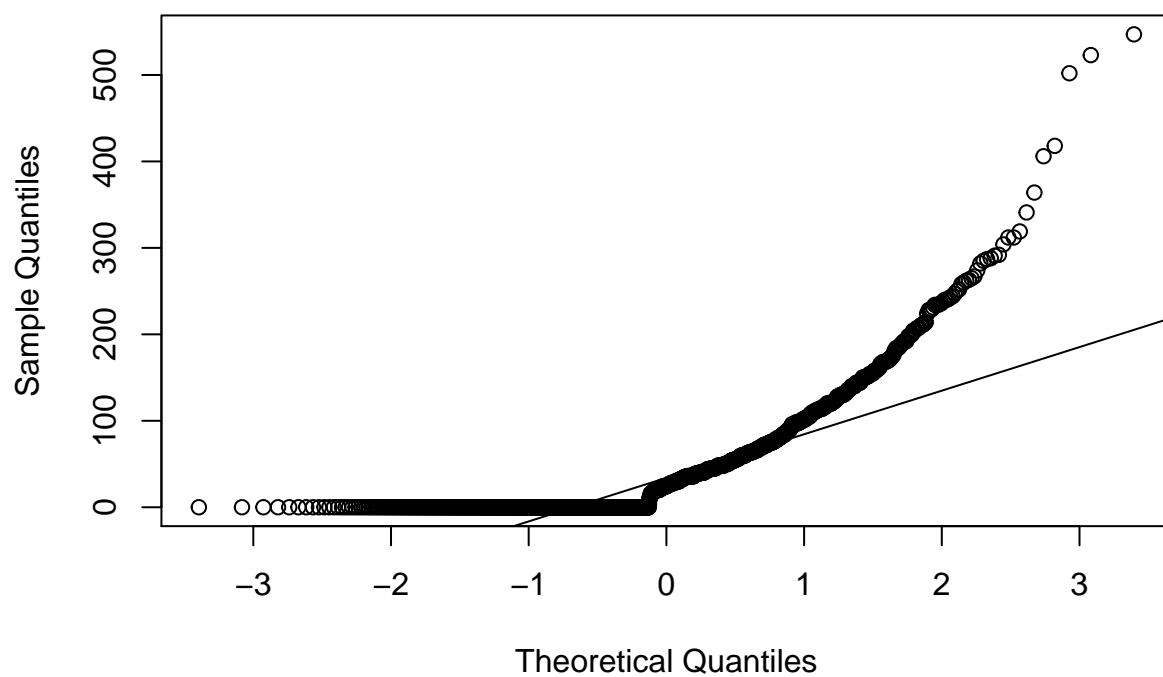
```
qqnorm(na.omit(as.numeric(dataSet$X2ndFlrSF)),main = "Distribucion normal de X2ndFlrSF")
qqline(na.omit(as.numeric(dataSet$X2ndFlrSF)))
```

Distribucion normal de X2ndFlrSF



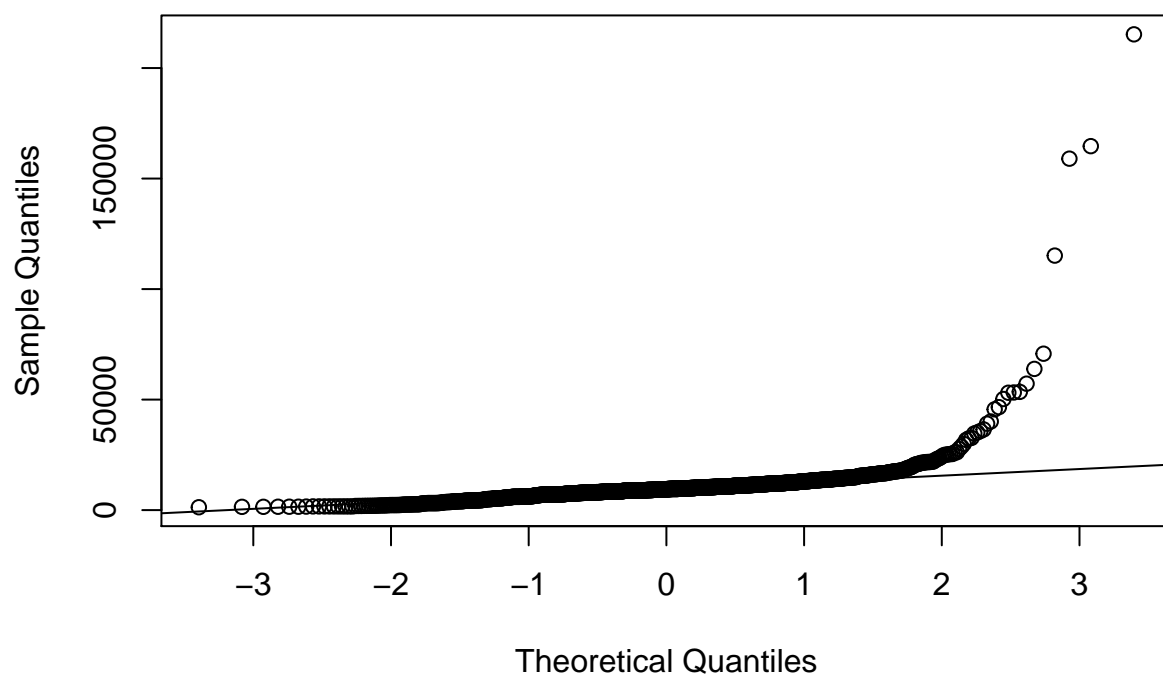
```
qqnorm(na.omit(as.numeric(dataSet$OpenPorchSF)),main = "Distribucion normal de OpenPorchSF")
qqline(na.omit(as.numeric(dataSet$OpenPorchSF)))
```

Distribucion normal de OpenPorchSF



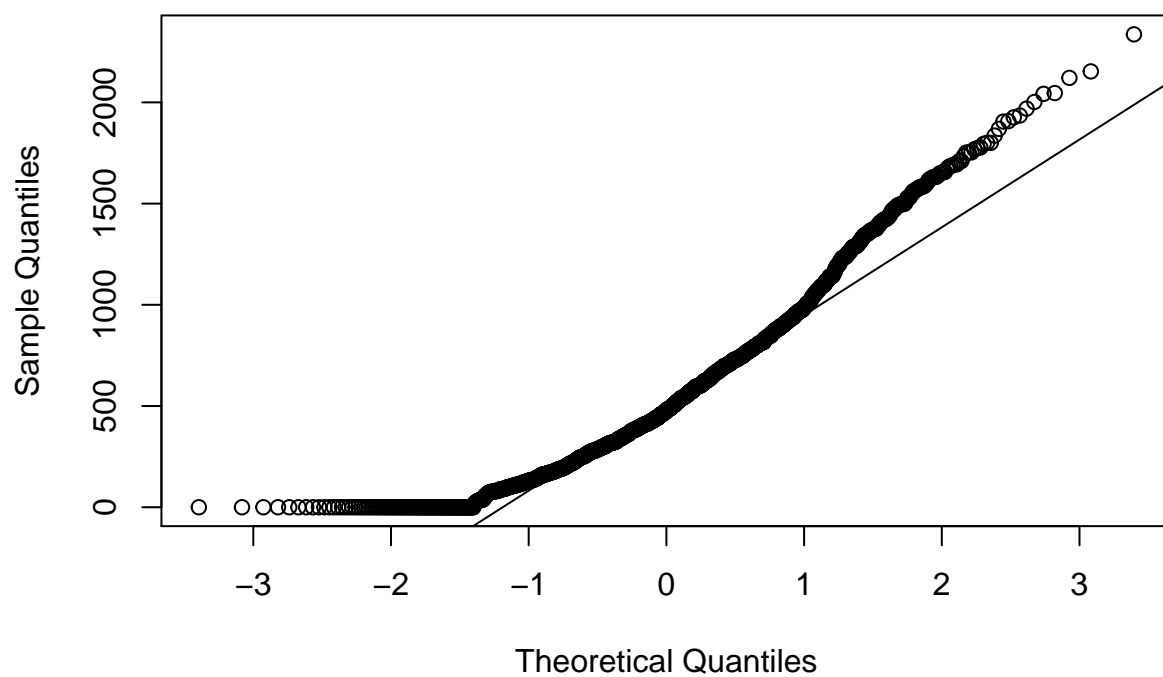
```
qqnorm(na.omit(as.numeric(dataSet$LotArea)),main = "Distribucion normal de LotArea")  
qqline(na.omit(as.numeric(dataSet$LotArea)))
```

Distribucion normal de LotArea



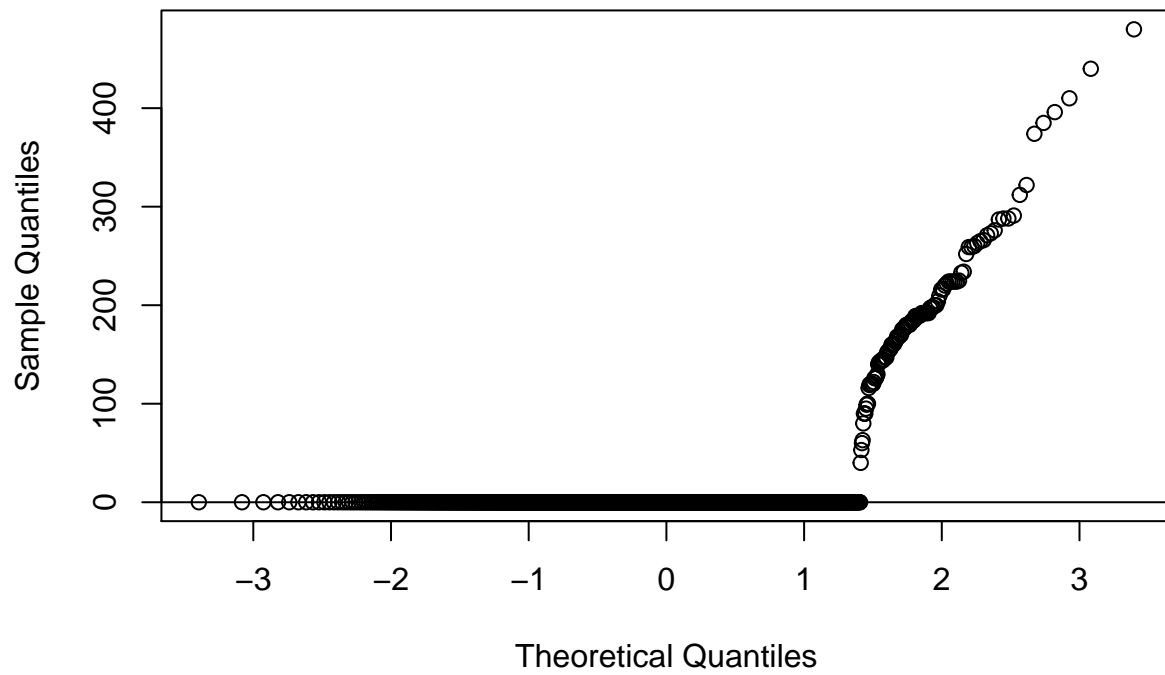
```
qqnorm(na.omit(as.numeric(dataSet$BsmtUnfSF)),main = "Distribucion normal de BsmtUnfSF")  
qqline(na.omit(as.numeric(dataSet$BsmtUnfSF)))
```

Distribucion normal de BsmtUnfSF



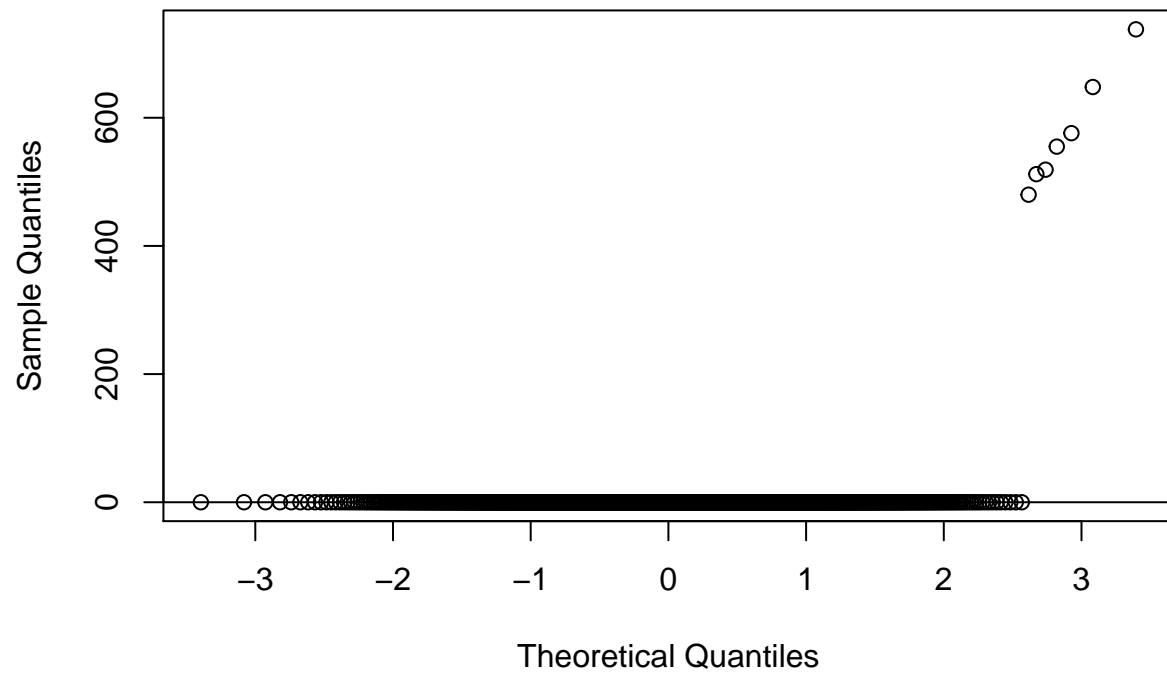
```
qqnorm(na.omit(as.numeric(dataSet$ScreenPorch)),main = "Distribucion normal de ScreenPorch")
qqline(na.omit(as.numeric(dataSet$ScreenPorch)))
```

Distribucion normal de ScreenPorch



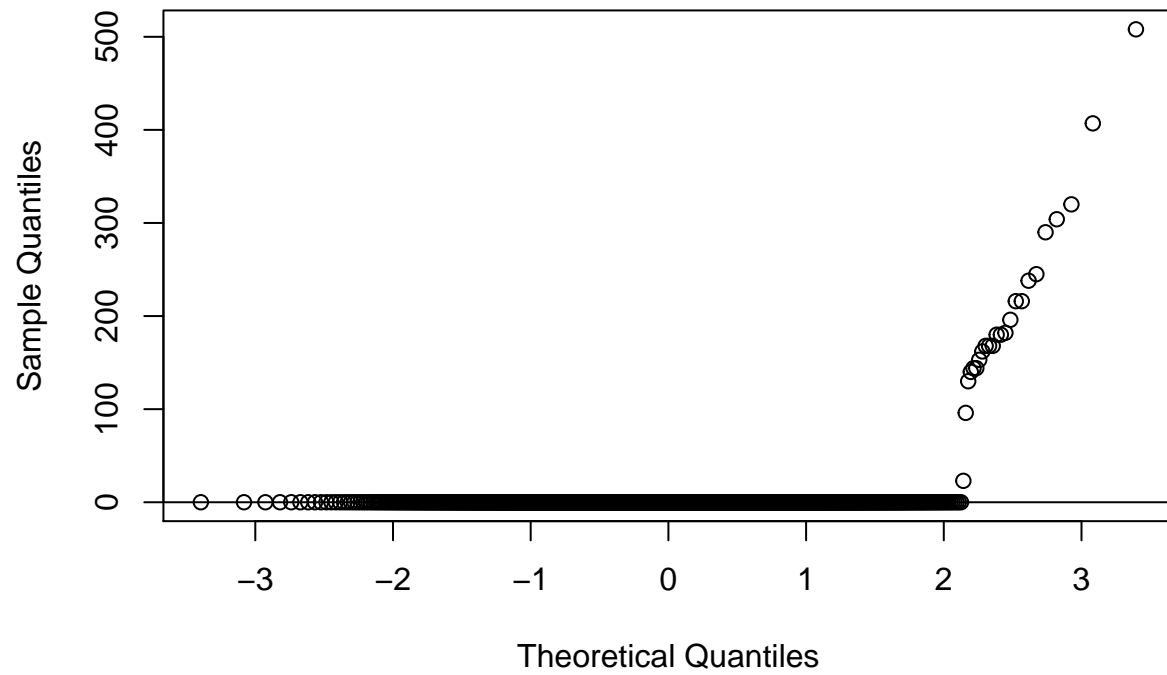
```
qqnorm(na.omit(as.numeric(dataSet$PoolArea)),main = "Distribucion normal de PoolArea")  
qqline(na.omit(as.numeric(dataSet$PoolArea)))
```

Distribucion normal de PoolArea



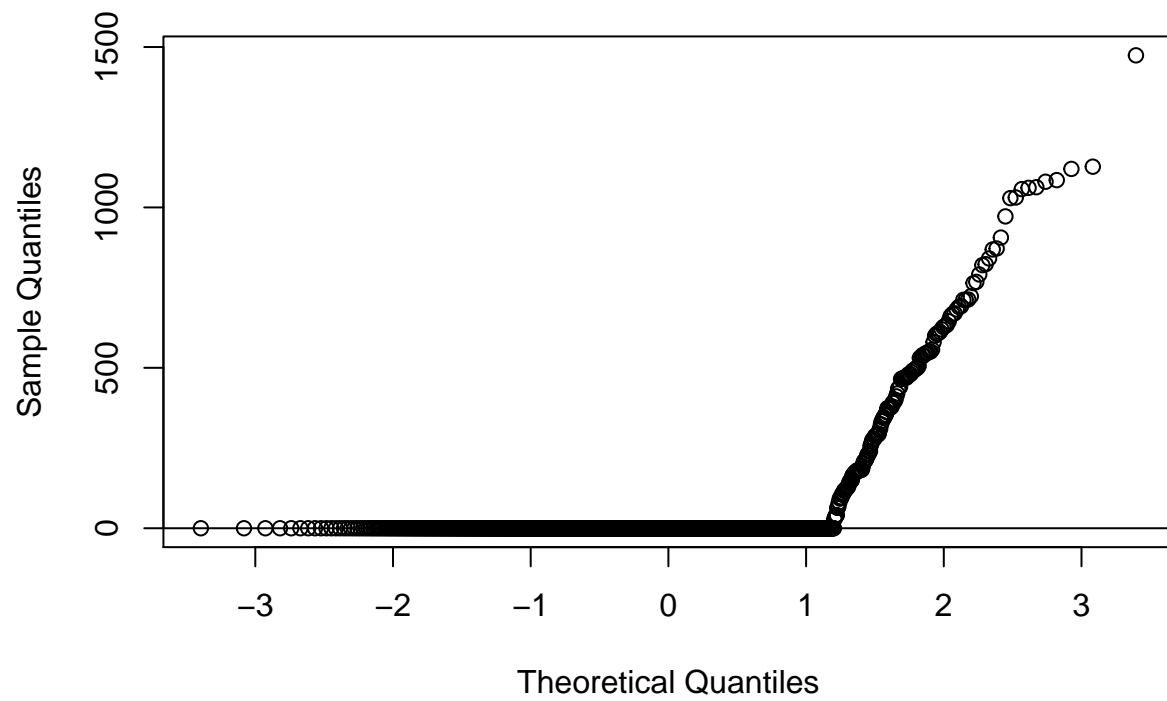
```
qqnorm(na.omit(as.numeric(dataSet$X3SsnPorch)),main = "Distribucion normal de X3SsnPorch")  
qqline(na.omit(as.numeric(dataSet$X3SsnPorch)))
```

Distribucion normal de X3SsnPorch



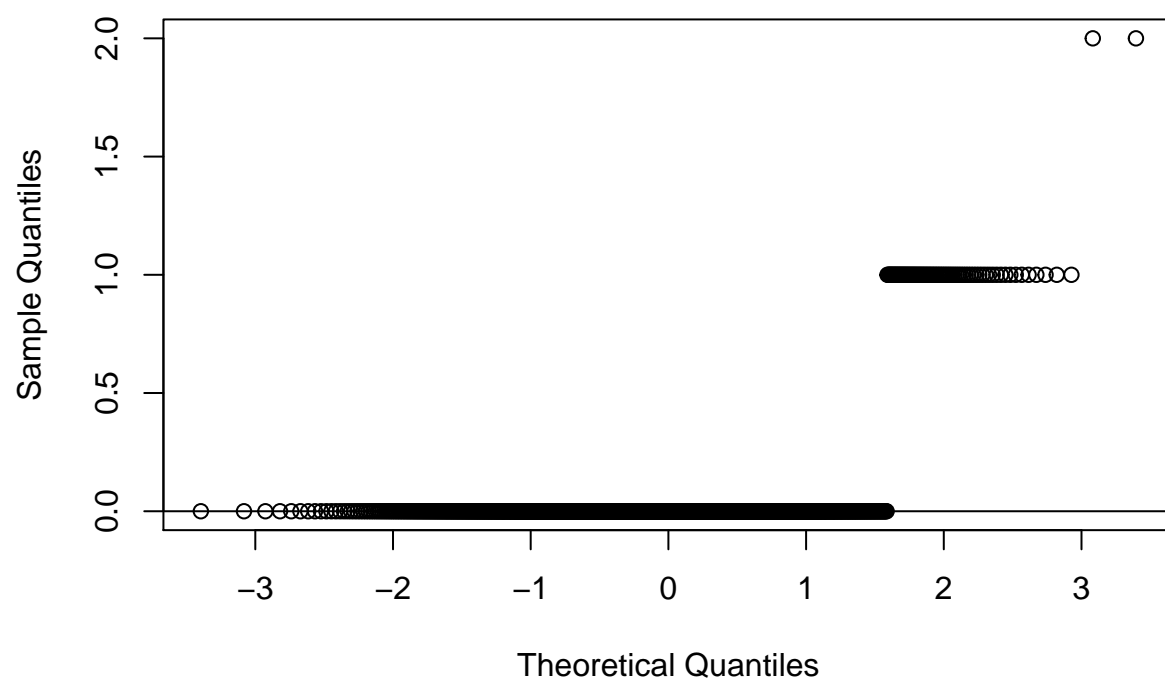
```
qqnorm(na.omit(as.numeric(dataSet$BsmtFinSF2)),main = "Distribucion normal de BsmtFinSF2")  
qqline(na.omit(as.numeric(dataSet$BsmtFinSF2)))
```


Distribucion normal de BsmtFinSF2



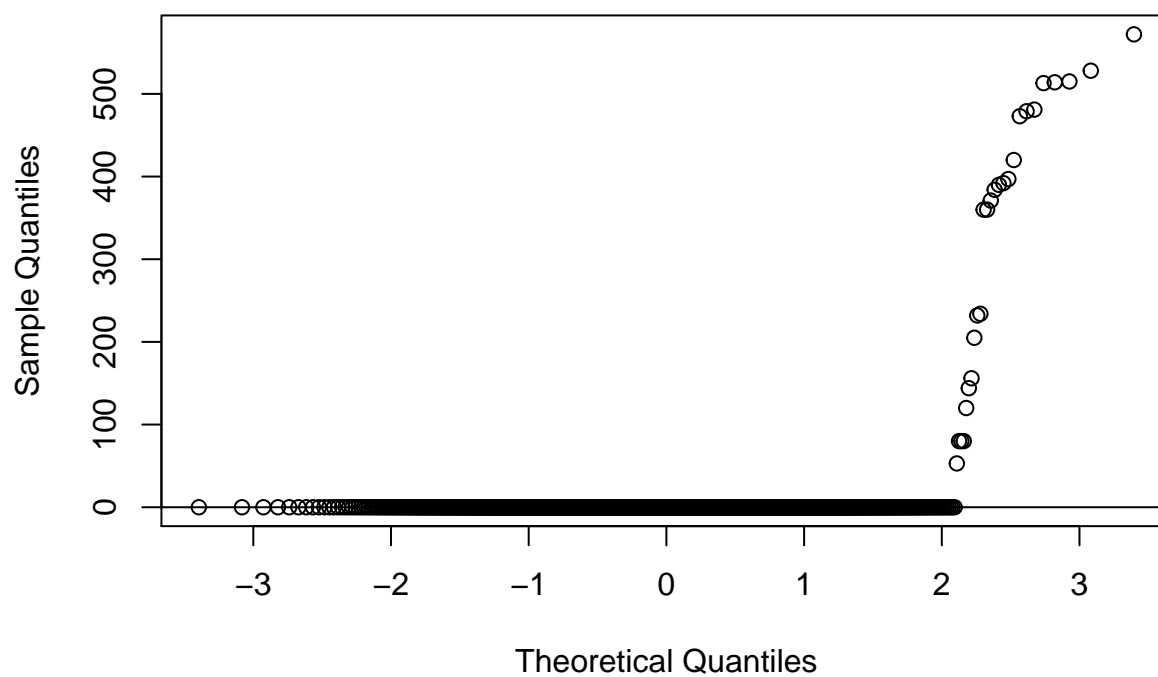
```
qqnorm(na.omit(as.numeric(dataSet$BsmtHalfBath)),main = "Distribucion normal de BsmtHalfBath")
qqline(na.omit(as.numeric(dataSet$BsmtHalfBath)))
```

Distribucion normal de BsmthalfBath



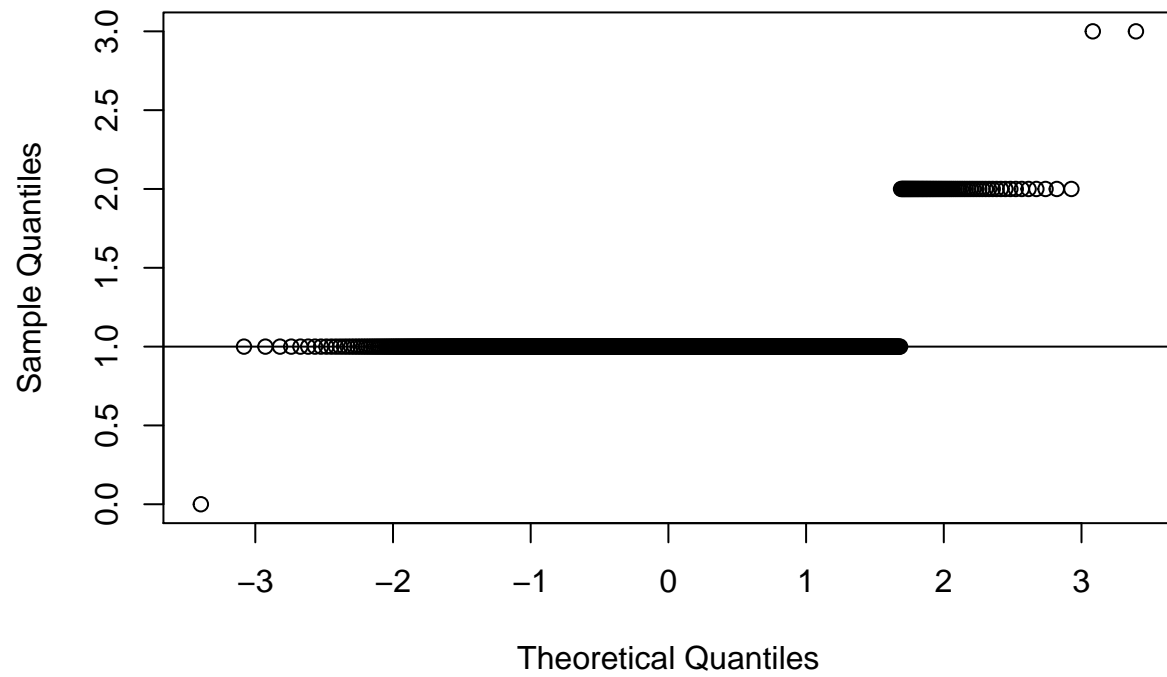
```
qqnorm(na.omit(as.numeric(dataSet$LowQualFinSF)),main = "Distribucion normal de LowQualFinSF")
qqline(na.omit(as.numeric(dataSet$LowQualFinSF)))
```

Distribucion normal de LowQualFinSF



```
qqnorm(na.omit(as.numeric(dataSet$KitchenAbvGr)),main = "Distribucion normal de KitchenAbvGr")  
qqline(na.omit(as.numeric(dataSet$KitchenAbvGr)))
```

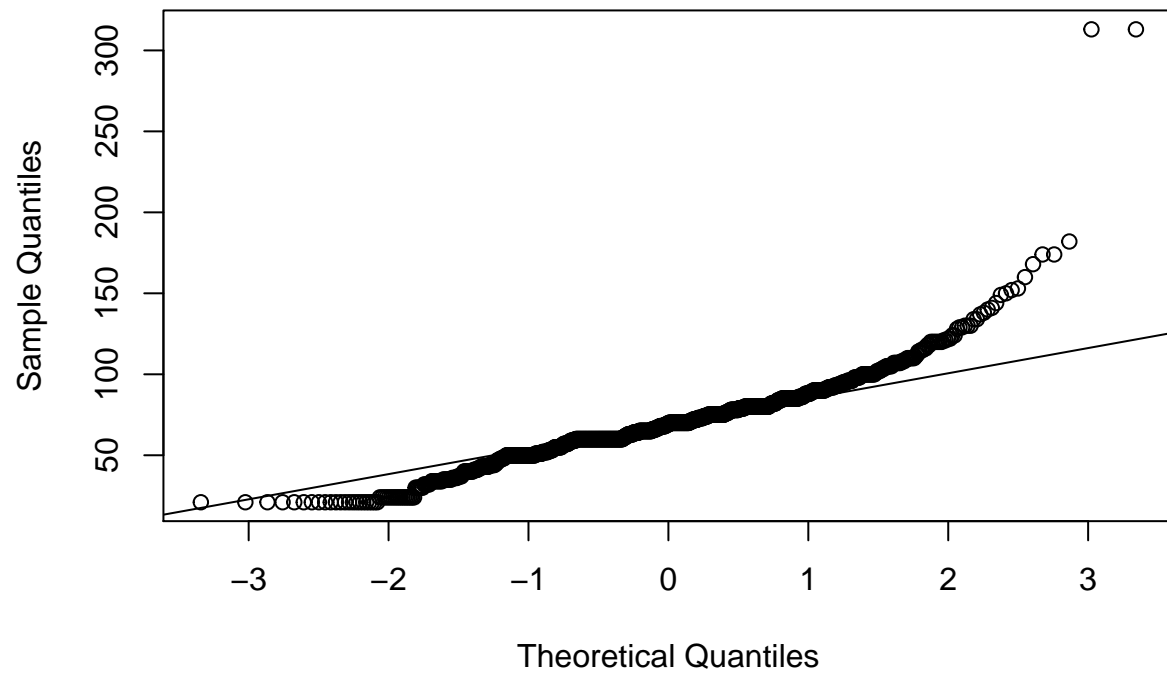
Distribucion normal de KitchenAbvGr



```
## Warning in na.omit(as.numeric(dataSet$LotFrontage)): NAs introduced by coercion
```

```
## Warning in na.omit(as.numeric(dataSet$LotFrontage)): NAs introduced by coercion
```

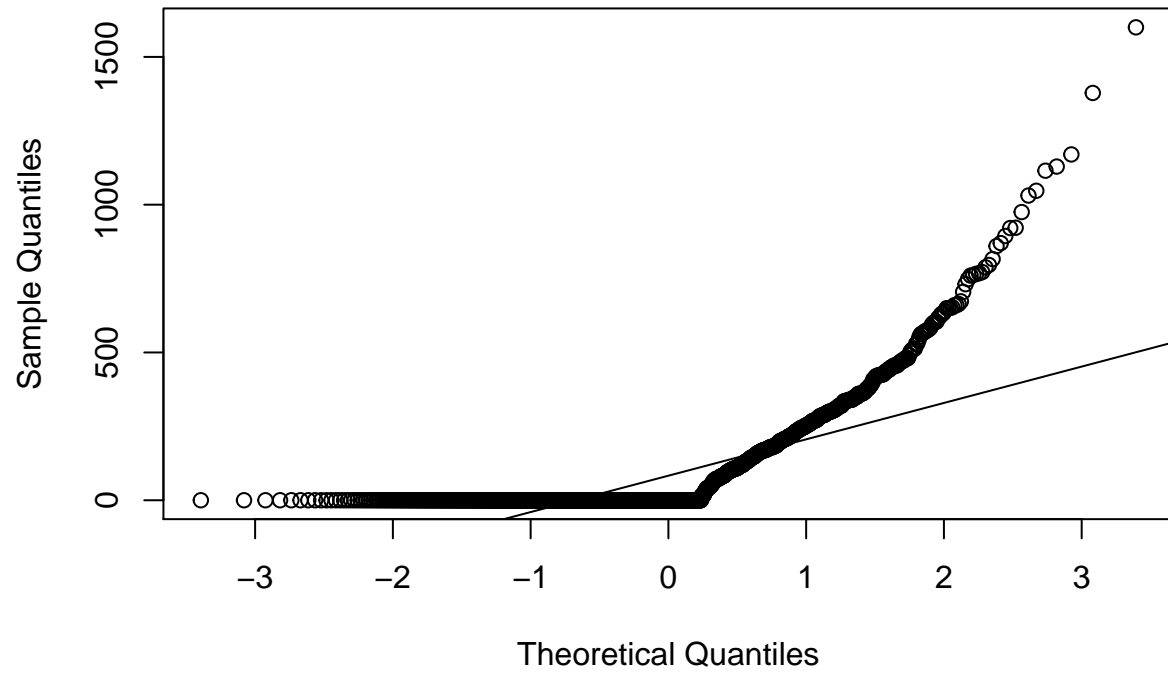
Distribucion normal de LotFrontage



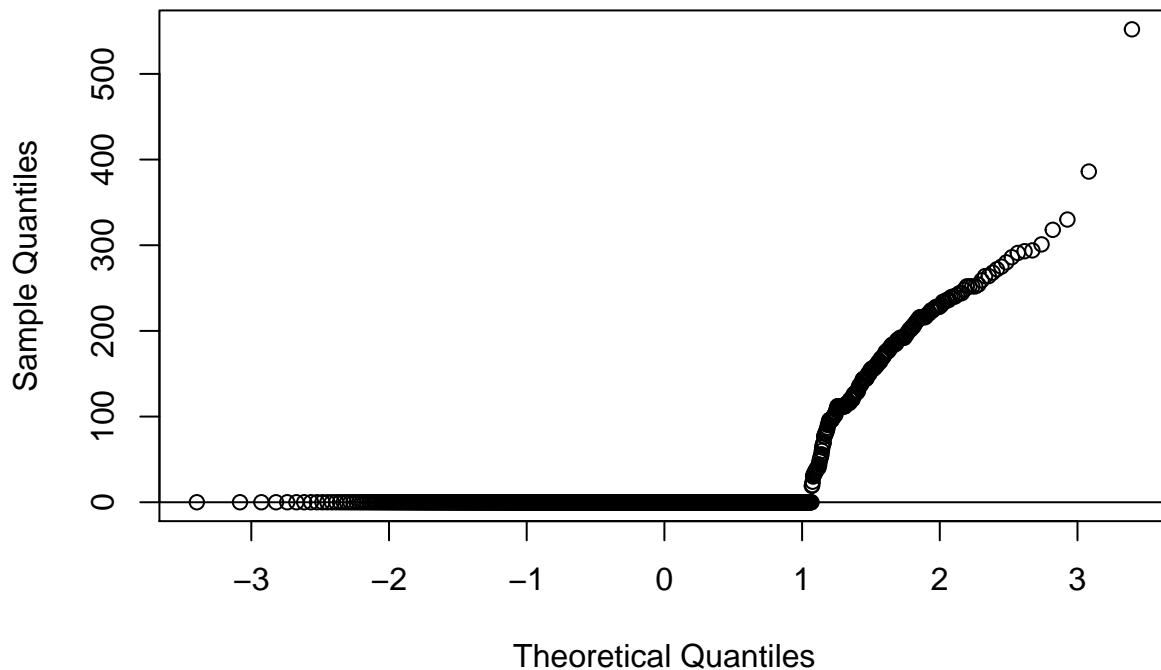
```
## Warning in na.omit(as.numeric(dataSet$MasVnrArea)): NAs introduced by coercion
```

```
## Warning in na.omit(as.numeric(dataSet$MasVnrArea)): NAs introduced by coercion
```

Distribucion normal de MasVnrArea



Distribucion normal de EnclosedPorch



Como podemos observar la mayoría de las variables siguen el comportamiento de la recta diagonal especialmente en el centro, en los extremos de la gráfica se empiezan a desviar los puntos lo que quiere decir que tenemos datos atípicos. Lo que puede significar es que muy pocas casas tengan ciertas cualidades, por ejemplo es lógico que la mayoría de casas van a ser estándar pero muy pocas van a ser de lujo o muy pero muy sencillas.

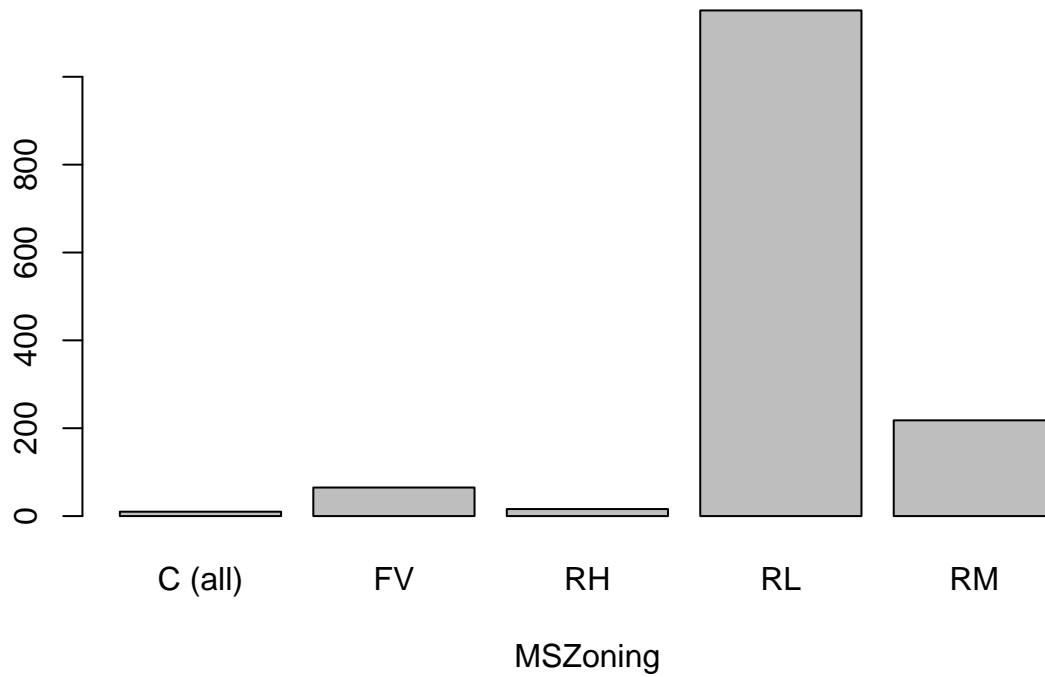
Análisis de variables cualitativas

Tablas de frecuencia y gráficas de distribución

A continuación, se muestran gráficos de barras para las variables cualitativas, en el caso de que sean muchos posibles valores de la variable se muestran tablas de frecuencia. Esto se hace para conocer un poco el comportamiento de los datos, poder identificar los valores más predominantes en cada variable, así como también poder identificar si hay valores NAs.

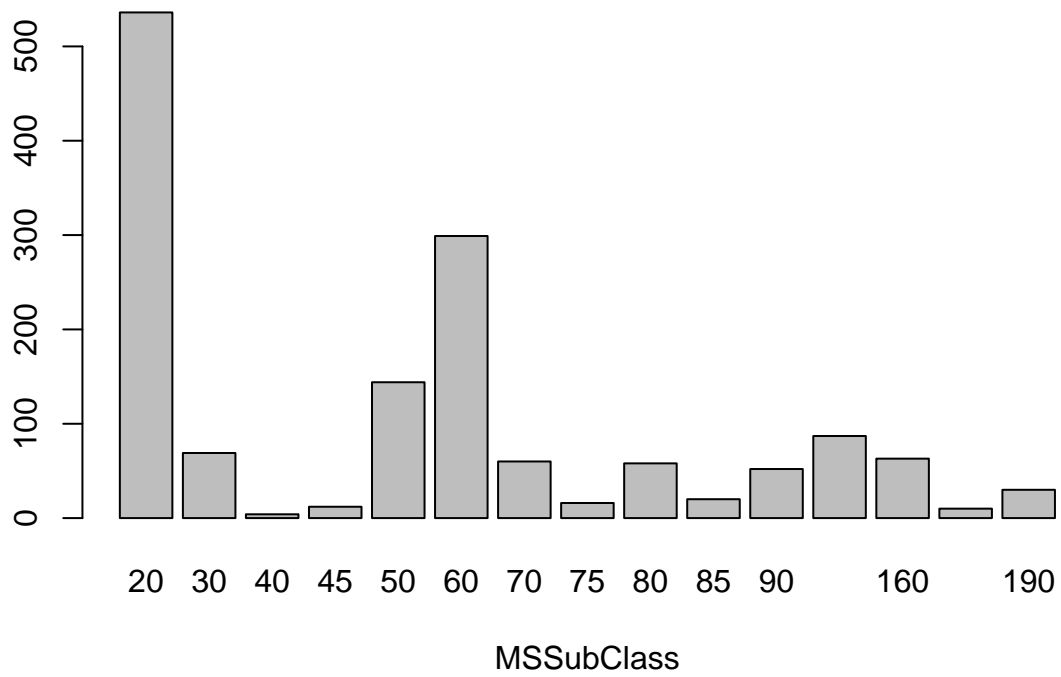
```
barplot(table(dataSet$MSZoning),main = "Distribución MSZoning",  
        xlab="MSZoning")
```

Distribución MSZoning



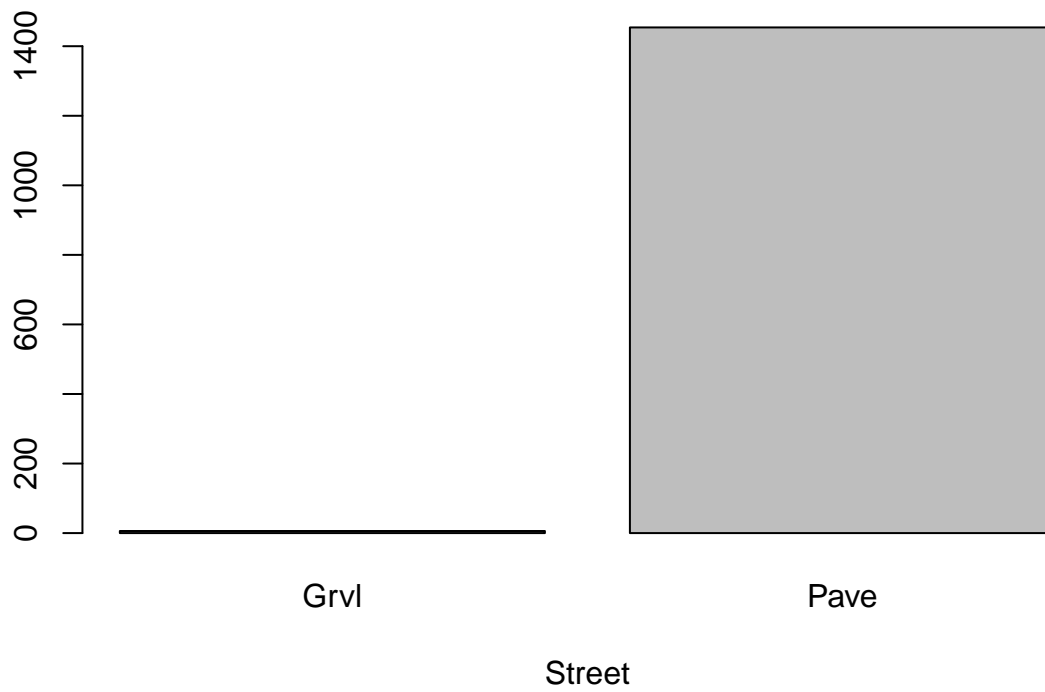
```
barplot(table(dataSet$MSSubClass),main = "Distribución MSSubClass",  
        xlab="MSSubClass")
```


Distribución MSSubClass



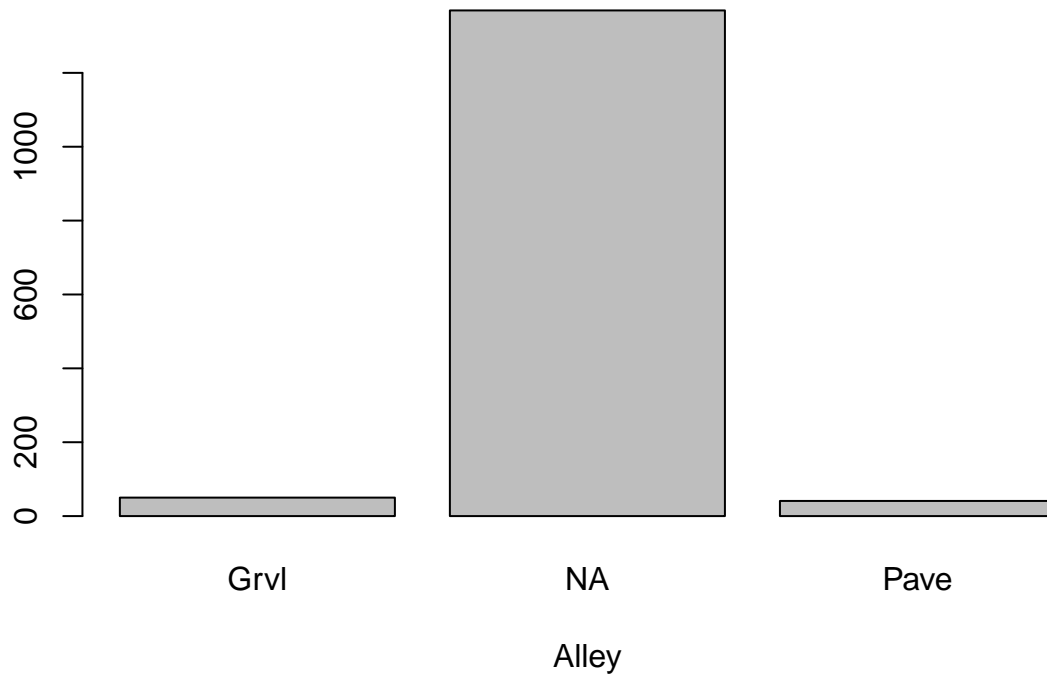
```
barplot(table(dataSet$Street),main = "Distribución Street",  
        xlab="Street")
```

Distribución Street

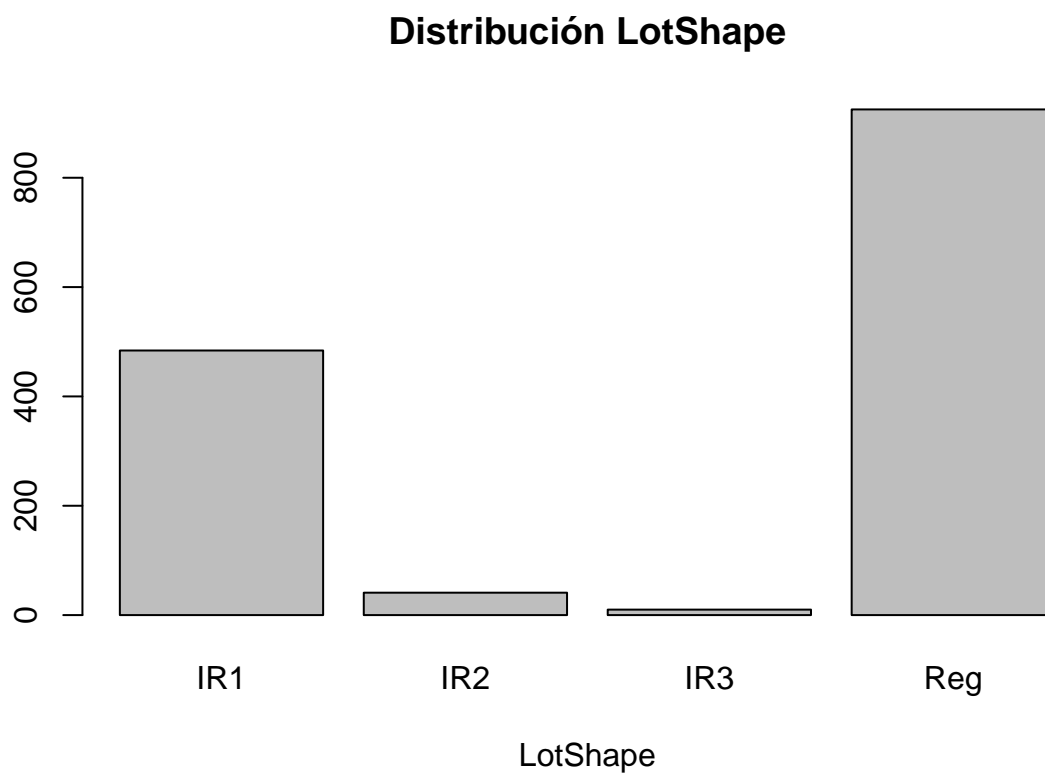


```
barplot(table(dataSet$Alley),main = "Distribución Alley",  
        xlab="Alley")
```

Distribución Alley

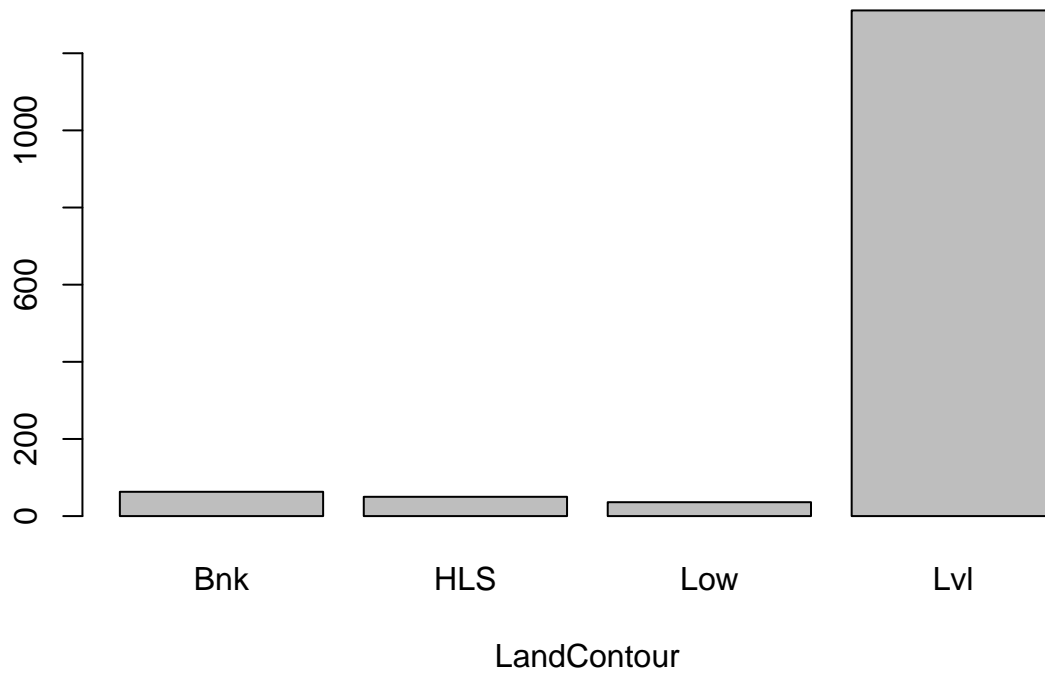


```
barplot(table(dataSet$LotShape),main = "Distribución LotShape",  
        xlab="LotShape")
```



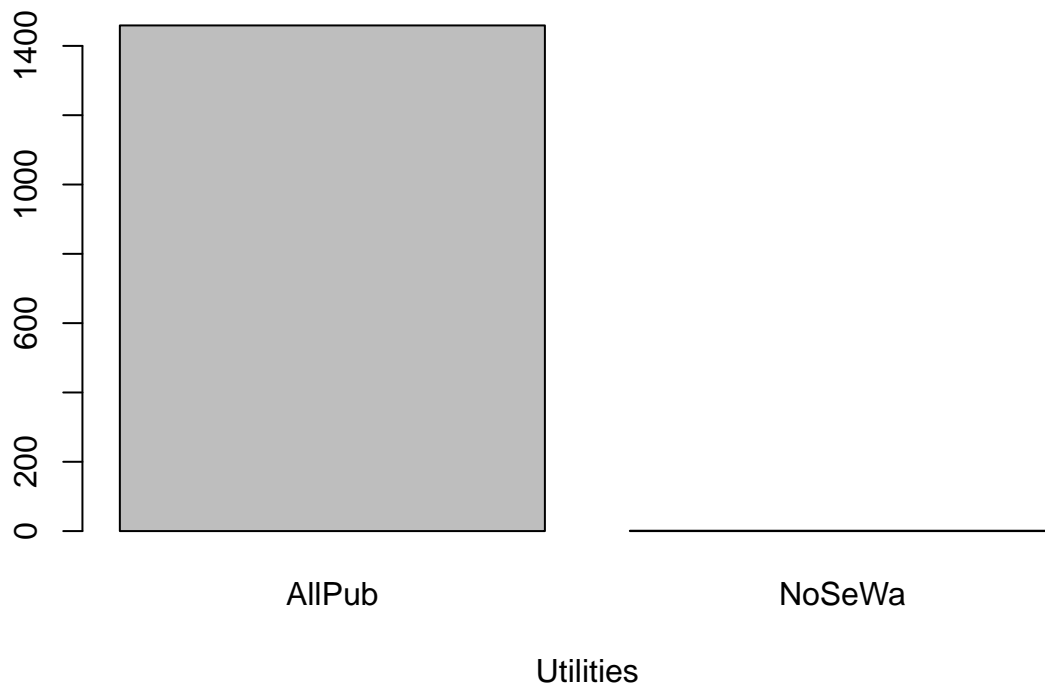
```
barplot(table(dataSet$LandContour),main = "Distribución LandContour",  
        xlab="LandContour")
```

Distribución LandContour

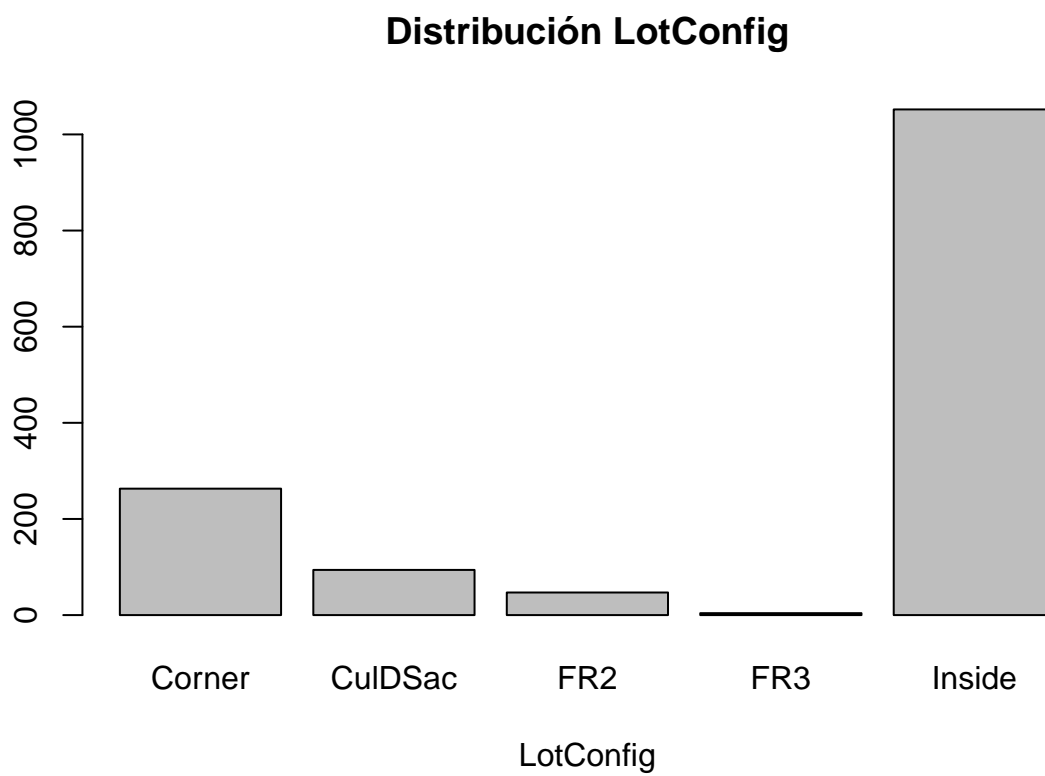


```
barplot(table(dataSet$Utilities),main = "Distribución Utilities",  
        xlab="Utilities")
```

Distribución Utilities



```
barplot(table(dataSet$LotConfig),main = "Distribución LotConfig",  
        xlab="LotConfig")
```



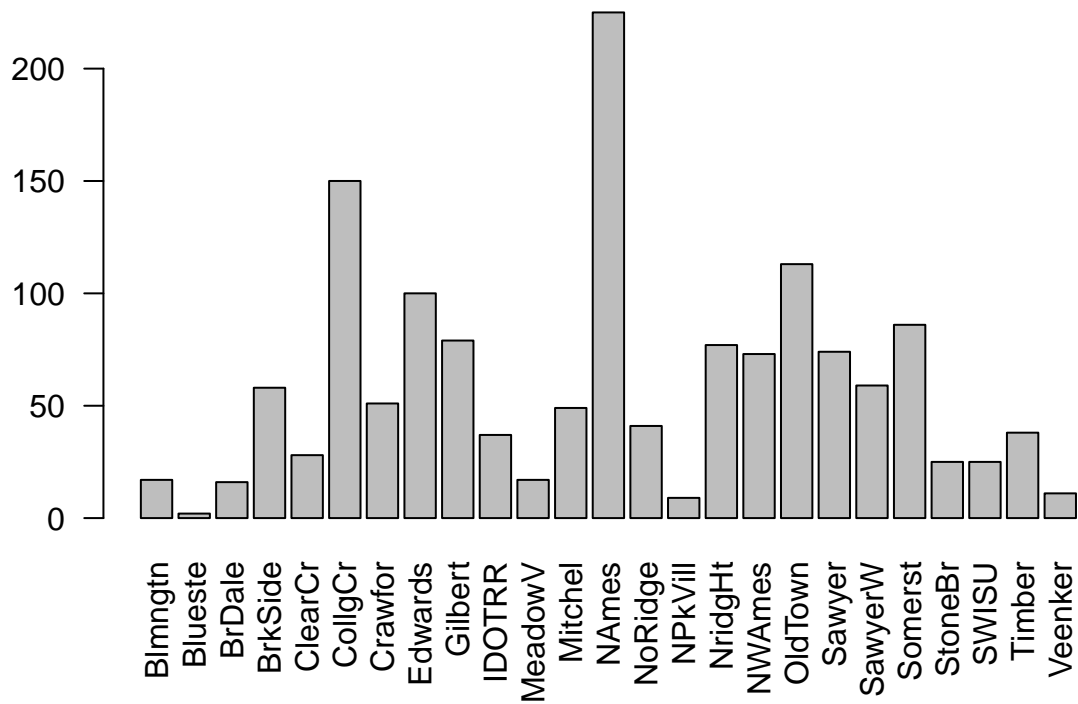
```
barplot(table(dataSet$LandSlope),main = "Distribución LandSlope",  
        xlab="LandSlope")
```

Distribución LandSlope



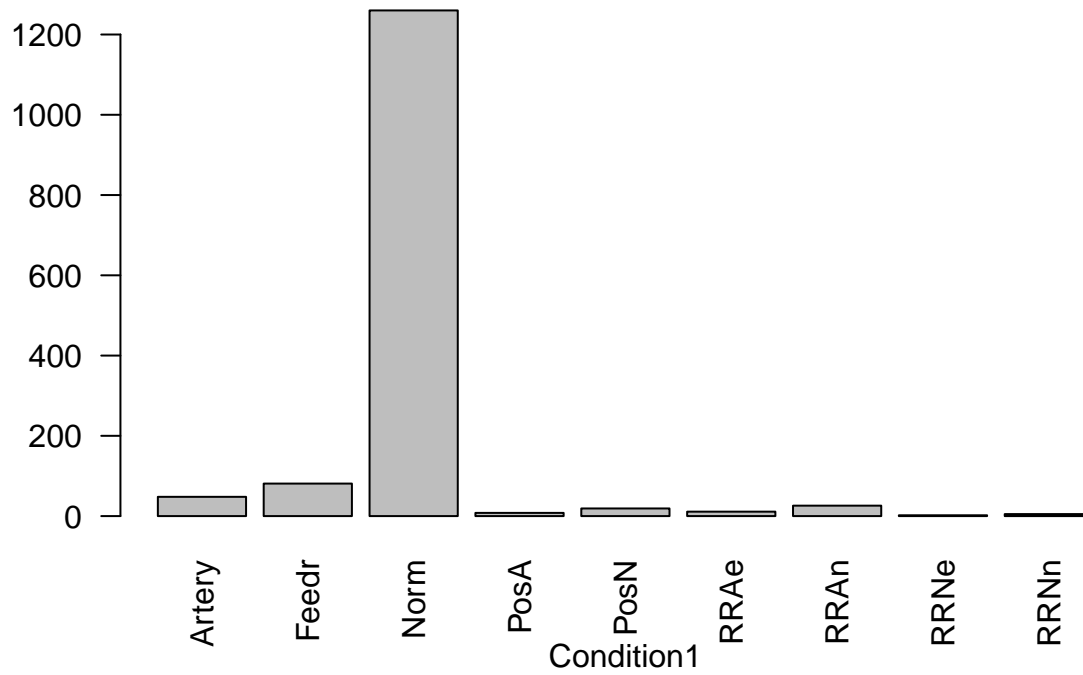
```
barplot(table(dataSet$Neighborhood),main = "Distribución Neighborhood",las=2)
```


Distribución Neighborhood



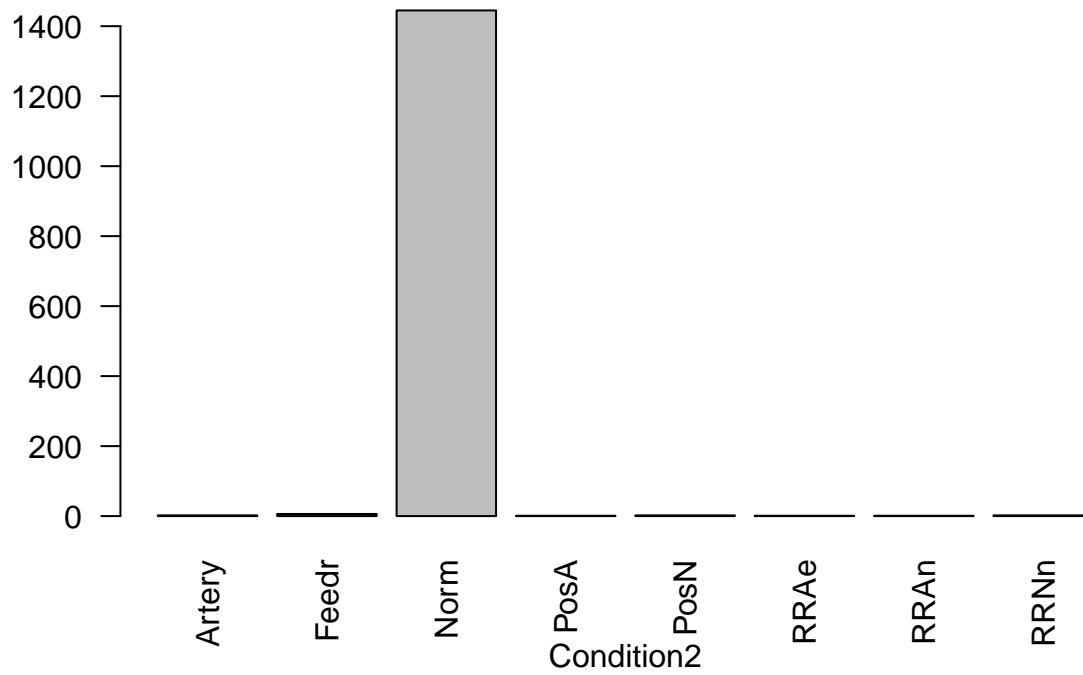
```
barplot(table(dataSet$Condition1),main = "Distribución Condition1",  
        xlab="Condition1",las=2)
```

Distribución Condition1

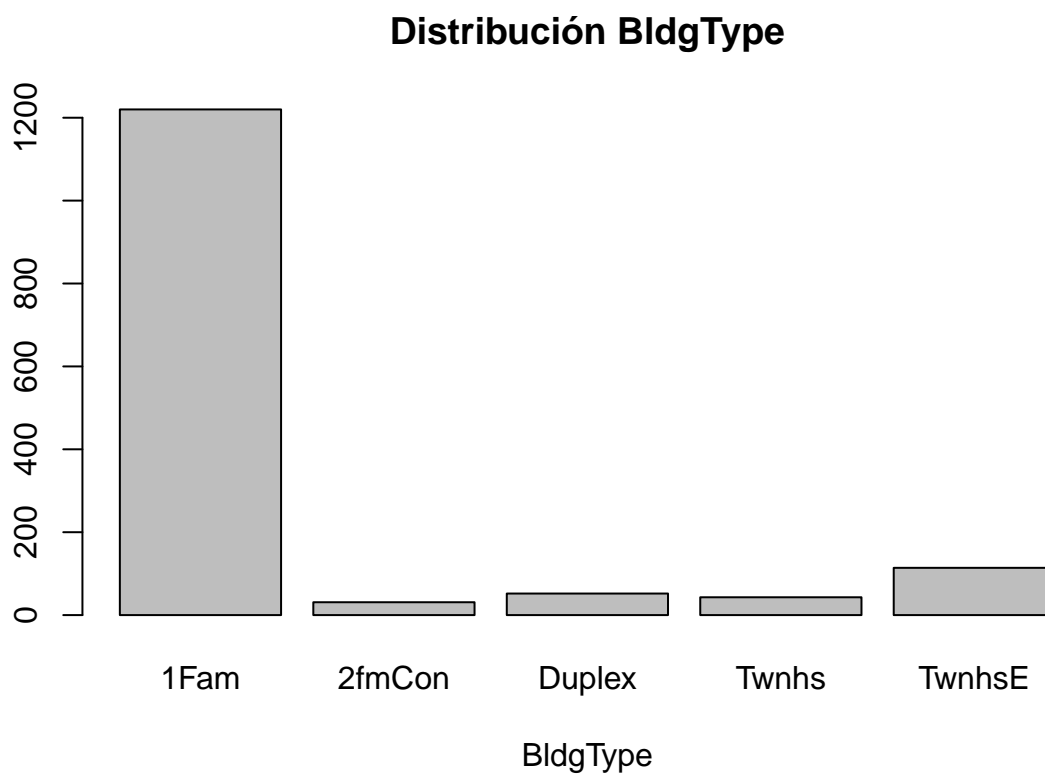


```
barplot(table(dataSet$Condition2),main = "Distribución Condition2",  
        xlab="Condition2",las=2)
```

Distribución Condition2

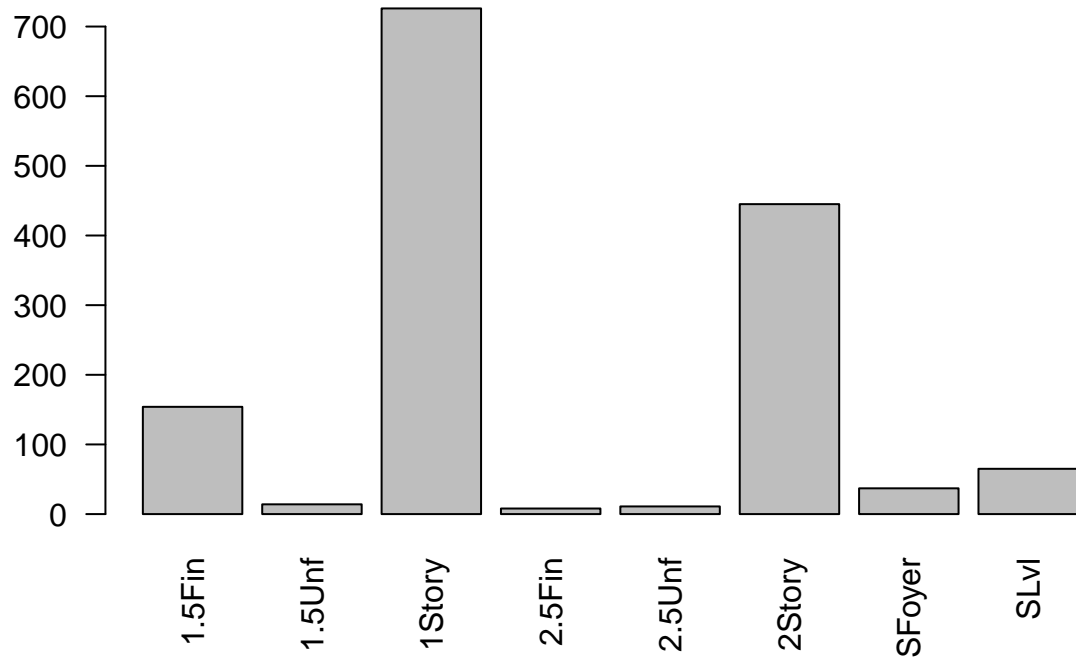


```
barplot(table(dataSet$BldgType),main = "Distribución BldgType",  
        xlab="BldgType")
```



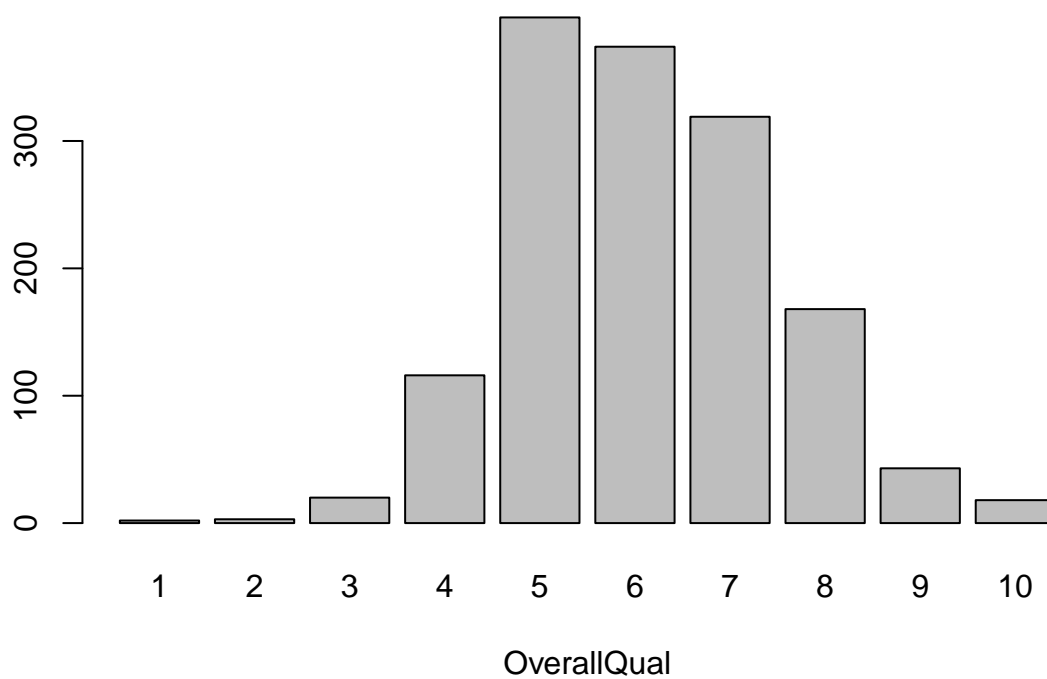
```
barplot(table(dataSet$HouseStyle),main = "Distribución HouseStyle",  
        las=2)
```

Distribución HouseStyle



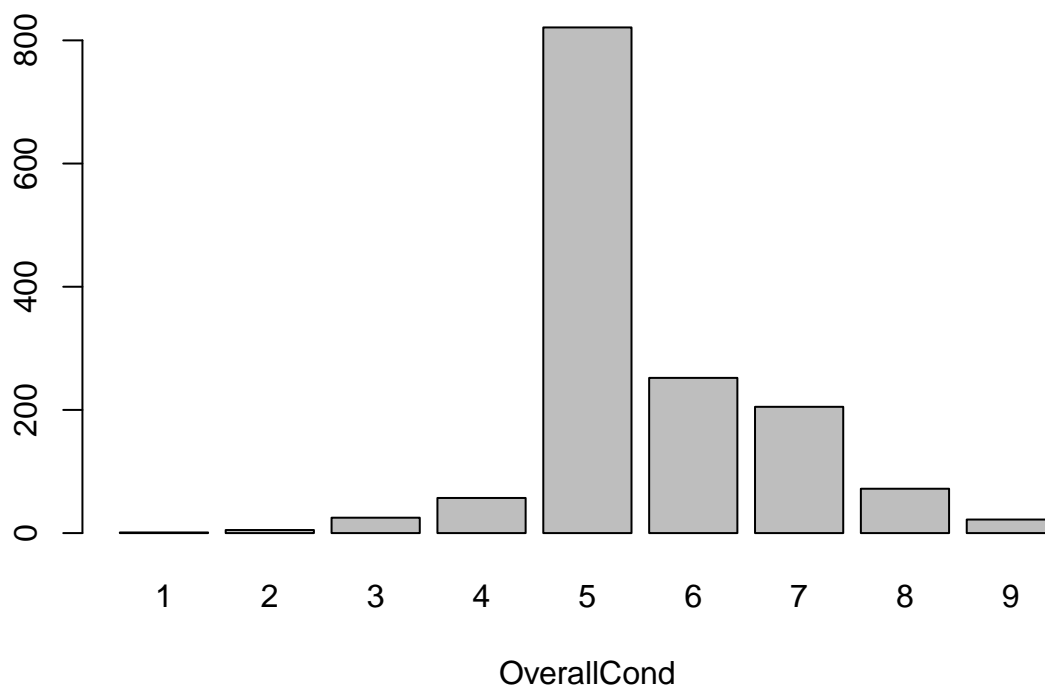
```
barplot(table(dataSet$OverallQual),main = "Distribución OverallQual",  
        xlab="OverallQual")
```

Distribución OverallQual



```
barplot(table(dataSet$OverallQual),main = "Distribución OverallQual",  
        xlab="OverallQual")
```

Distribución OverallCond



```
knitr::kable(
  table(dataSet$YearBuilt), caption = 'Tabla de frecuencia YearBuilt'
)
```

Table 1: Tabla de frecuencia YearBuilt

Var1	Freq
1872	1
1875	1
1880	4
1882	1
1885	2
1890	2
1892	2
1893	1
1898	1
1900	10
1904	1
1905	1
1906	1
1908	2
1910	17
1911	1
1912	3
1913	1

Var1	Freq
1914	7
1915	10
1916	8
1917	1
1918	7
1919	3
1920	30
1921	6
1922	8
1923	7
1924	7
1925	16
1926	9
1927	3
1928	7
1929	4
1930	9
1931	6
1932	4
1934	3
1935	6
1936	9
1937	5
1938	4
1939	8
1940	18
1941	15
1942	2
1945	6
1946	7
1947	5
1948	14
1949	12
1950	20
1951	6
1952	5
1953	12
1954	24
1955	16
1956	14
1957	20
1958	24
1959	26
1960	17
1961	14
1962	19
1963	16
1964	15
1965	24
1966	18
1967	16
1968	22

Var1	Freq
1969	14
1970	24
1971	22
1972	23
1973	11
1974	10
1975	8
1976	33
1977	32
1978	16
1979	9
1980	10
1981	5
1982	6
1983	4
1984	9
1985	5
1986	5
1987	3
1988	11
1989	3
1990	12
1991	5
1992	13
1993	17
1994	19
1995	18
1996	15
1997	14
1998	25
1999	25
2000	24
2001	20
2002	23
2003	45
2004	54
2005	64
2006	67
2007	49
2008	23
2009	18
2010	1

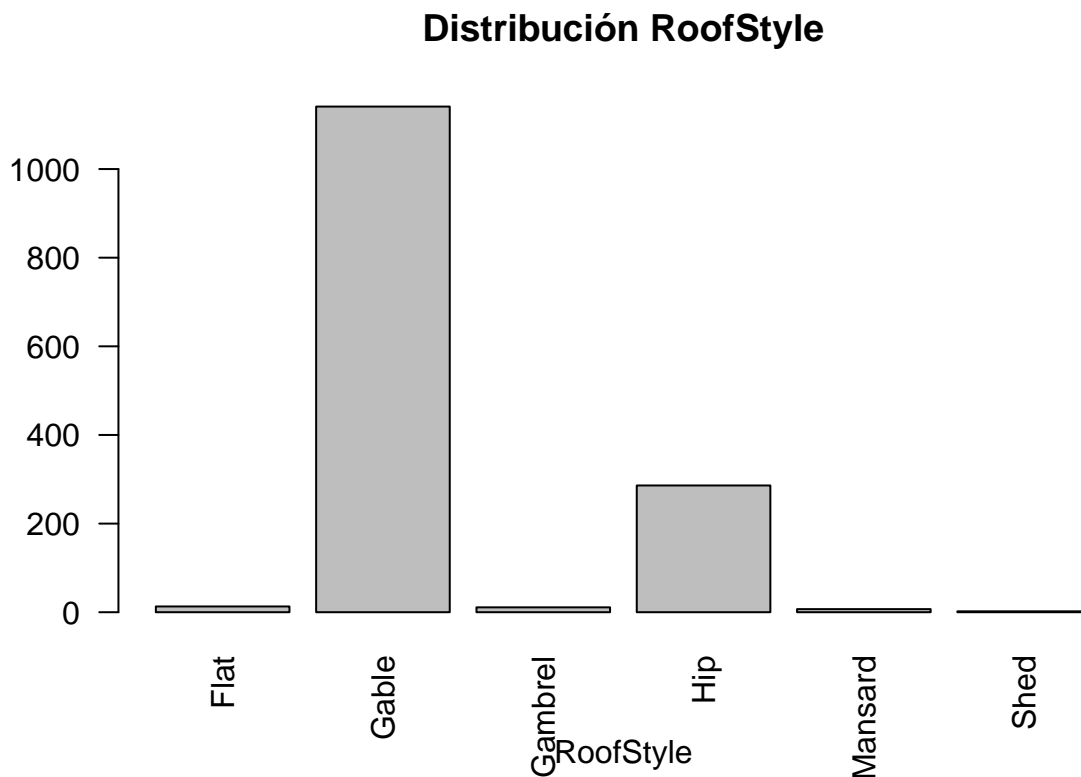
```
knitr::kable(
  table(dataSet$YearRemodAdd), caption = 'Tabla de frecuencia YearRemodAdd'
)
```

Table 2: Tabla de frecuencia YearRemodAdd

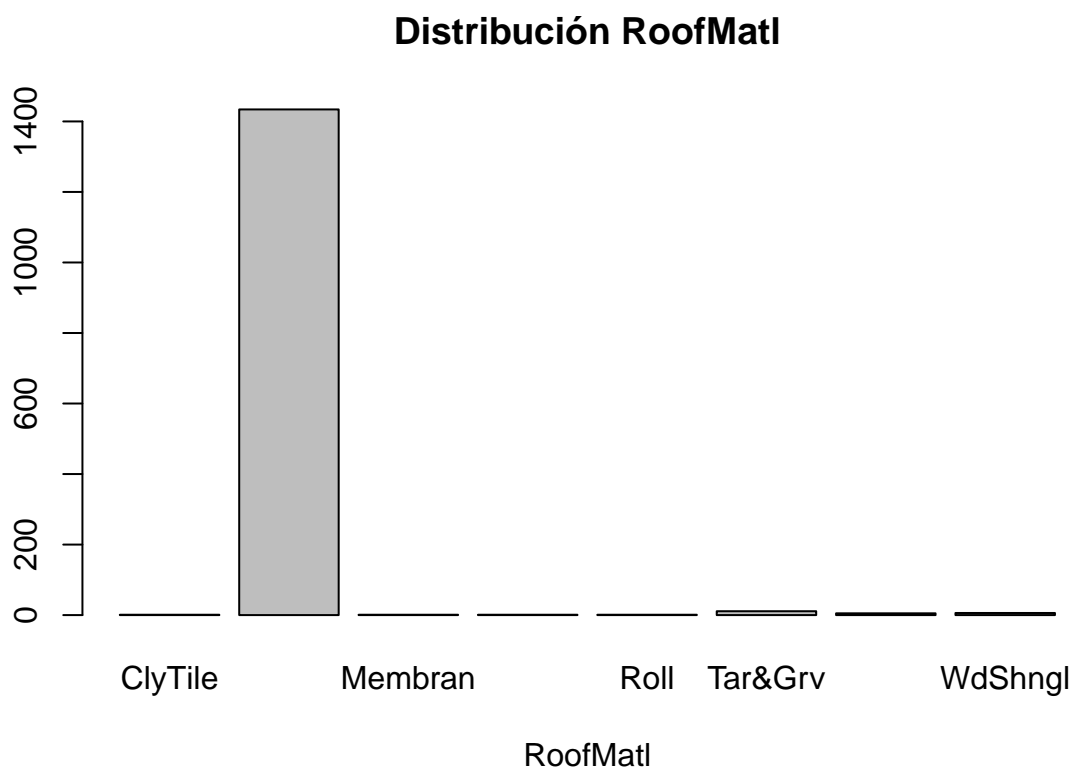
Var1	Freq
1950	178
1951	4
1952	5
1953	10
1954	14
1955	9
1956	10
1957	9
1958	15
1959	18
1960	12
1961	8
1962	14
1963	13
1964	11
1965	19
1966	15
1967	12
1968	17
1969	14
1970	26
1971	18
1972	20
1973	11
1974	7
1975	10
1976	30
1977	25
1978	16
1979	10
1980	12
1981	8
1982	7
1983	5
1984	7
1985	9
1986	5
1987	10
1988	9
1989	11
1990	15
1991	14
1992	17
1993	19
1994	22
1995	31
1996	36
1997	25
1998	36
1999	30
2000	55

Var1	Freq
2001	21
2002	48
2003	51
2004	62
2005	73
2006	97
2007	76
2008	40
2009	23
2010	6

```
barplot(table(dataSet$RoofStyle),main = "Distribución RoofStyle",
        xlab="RoofStyle",las=2)
```

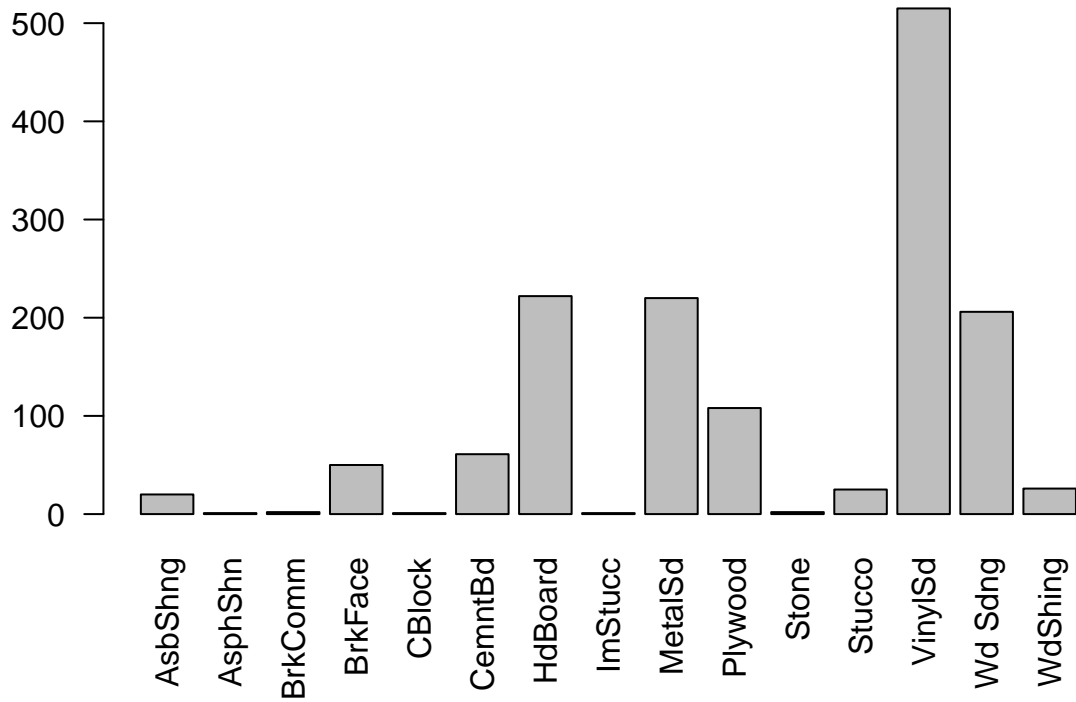


```
barplot(table(dataSet$RoofMat1),main = "Distribución RoofMat1",
        xlab="RoofMat1")
```



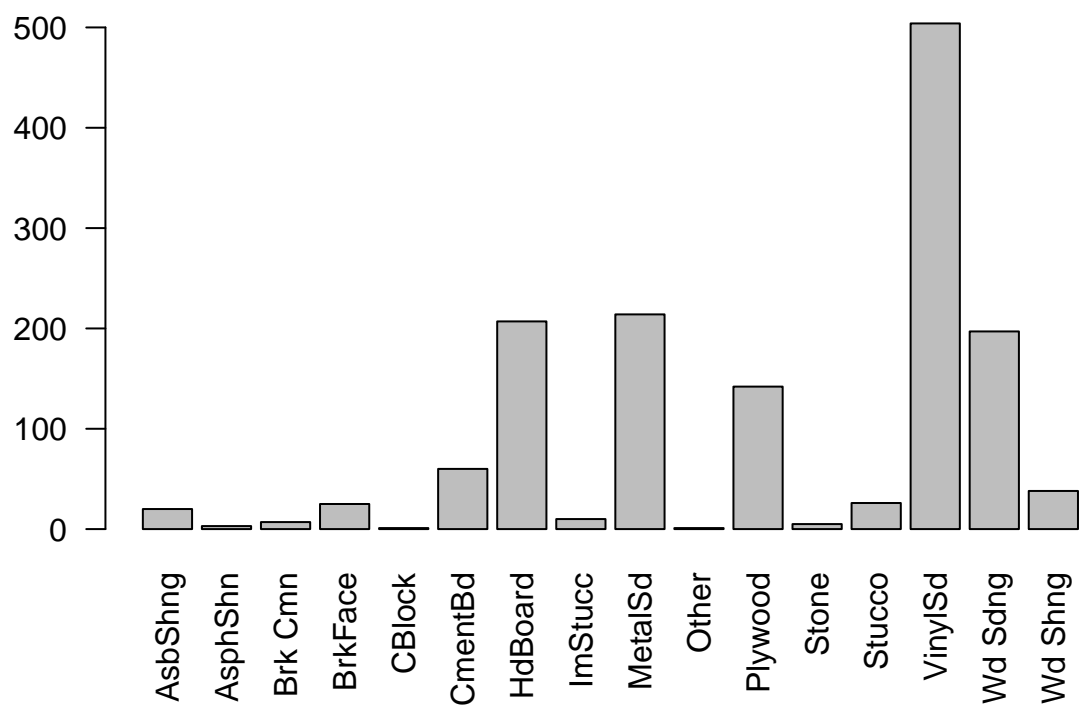
```
barplot(table(dataSet$Exterior1st),main = "Distribución Exterior1st",  
        las=2)
```

Distribución Exterior1st



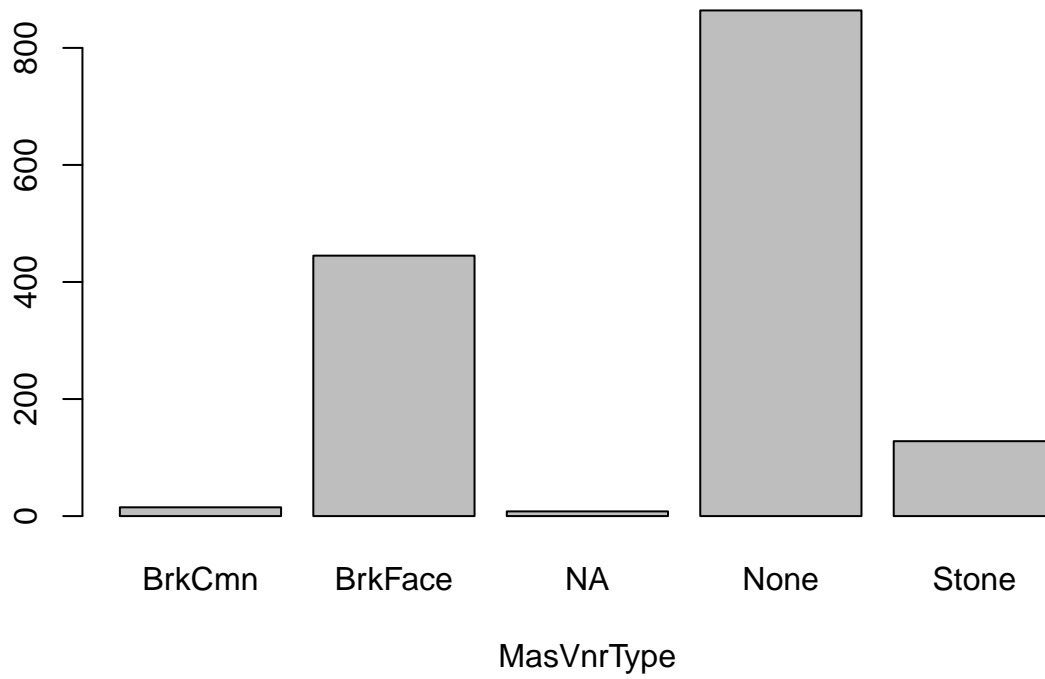
```
barplot(table(dataSet$Exterior2nd),main = "Distribución Exterior2nd",  
        las=2)
```

Distribución Exterior2nd



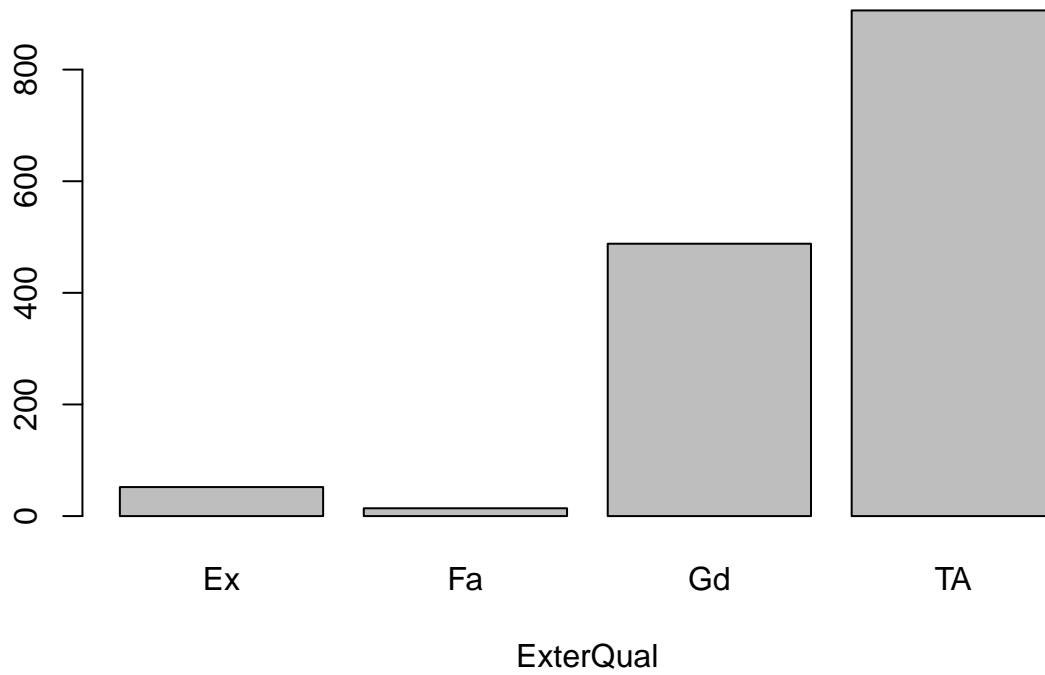
```
barplot(table(dataSet$MasVnrType),main = "Distribución MasVnrType",  
        xlab="MasVnrType")
```

Distribución MasVnrType



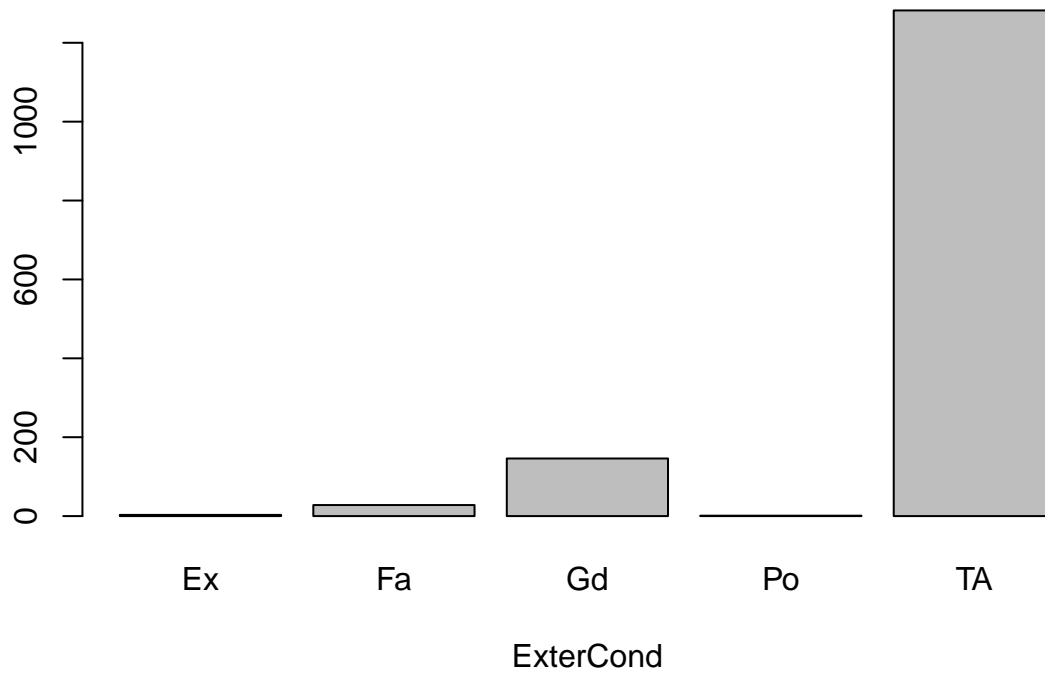
```
barplot(table(dataSet$ExterQual),main = "Distribución ExterQual",  
        xlab="ExterQual")
```

Distribución ExterQual

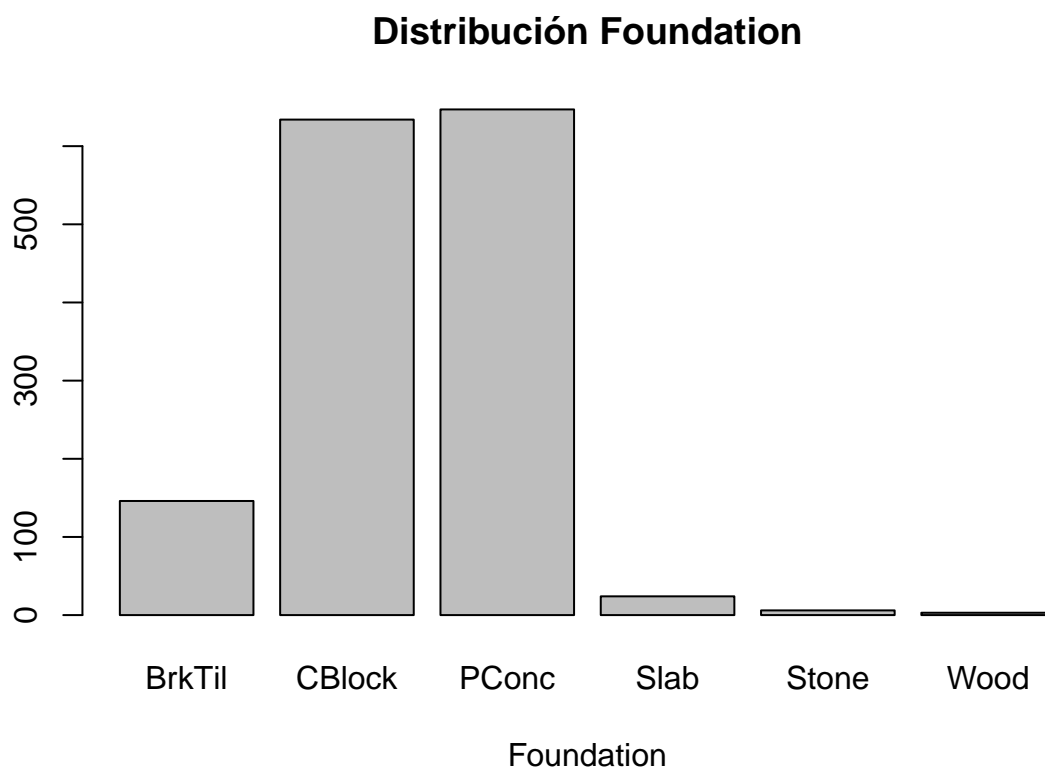


```
barplot(table(dataSet$ExterCond),main = "Distribución ExterCond",  
        xlab="ExterCond")
```

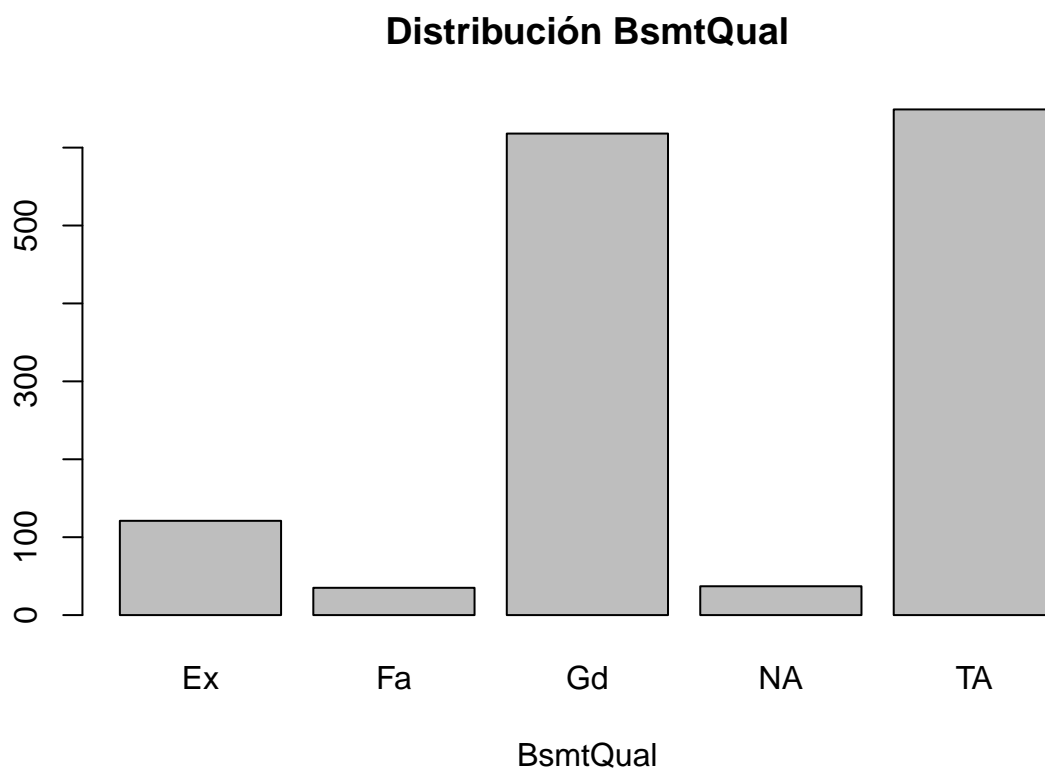

Distribución ExterCond



```
barplot(table(dataSet$Foundation),main = "Distribución Foundation",  
        xlab="Foundation")
```

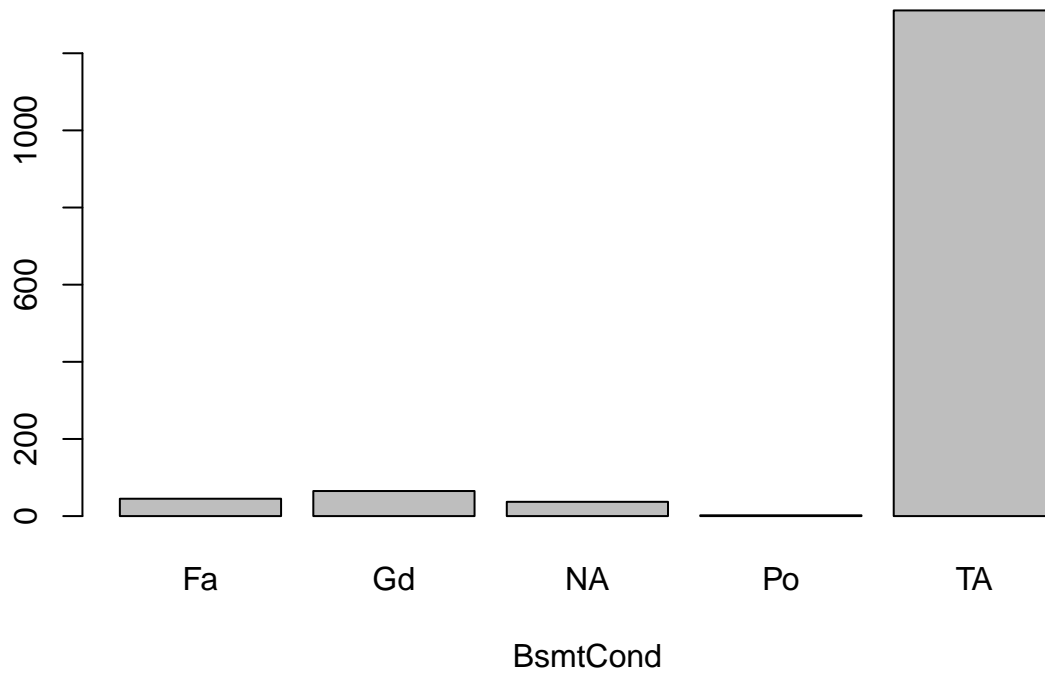


```
barplot(table(dataSet$BsmtQual),main = "Distribución BsmtQual",  
        xlab="BsmtQual")
```



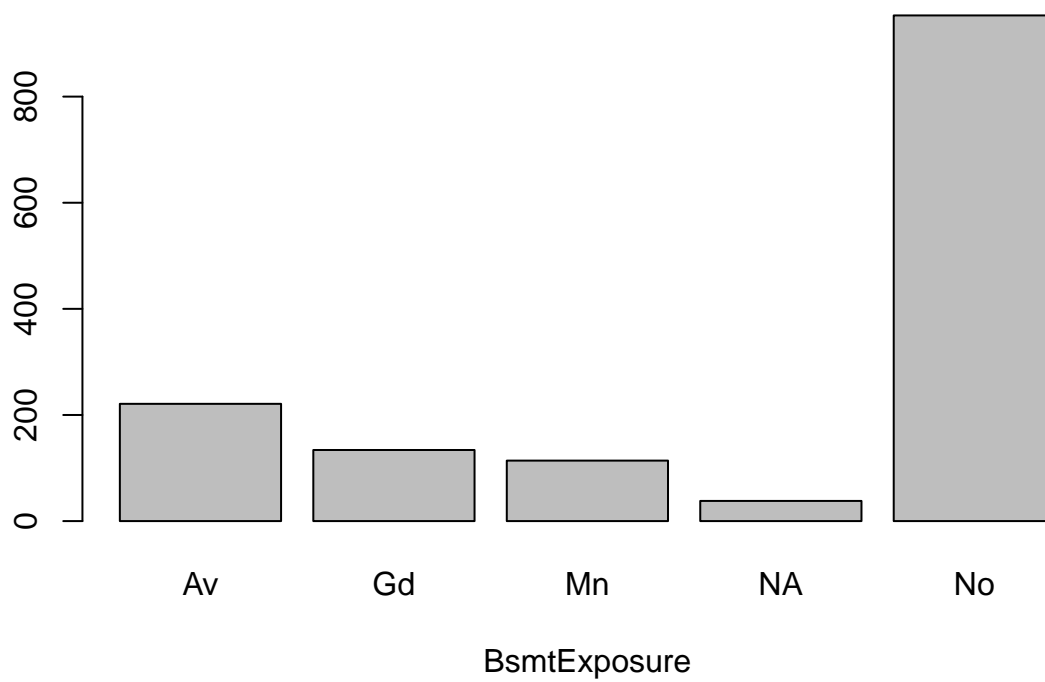
```
barplot(table(dataSet$BsmtCond),main = "Distribución BsmtCond",  
        xlab="BsmtCond")
```

Distribución BsmtCond



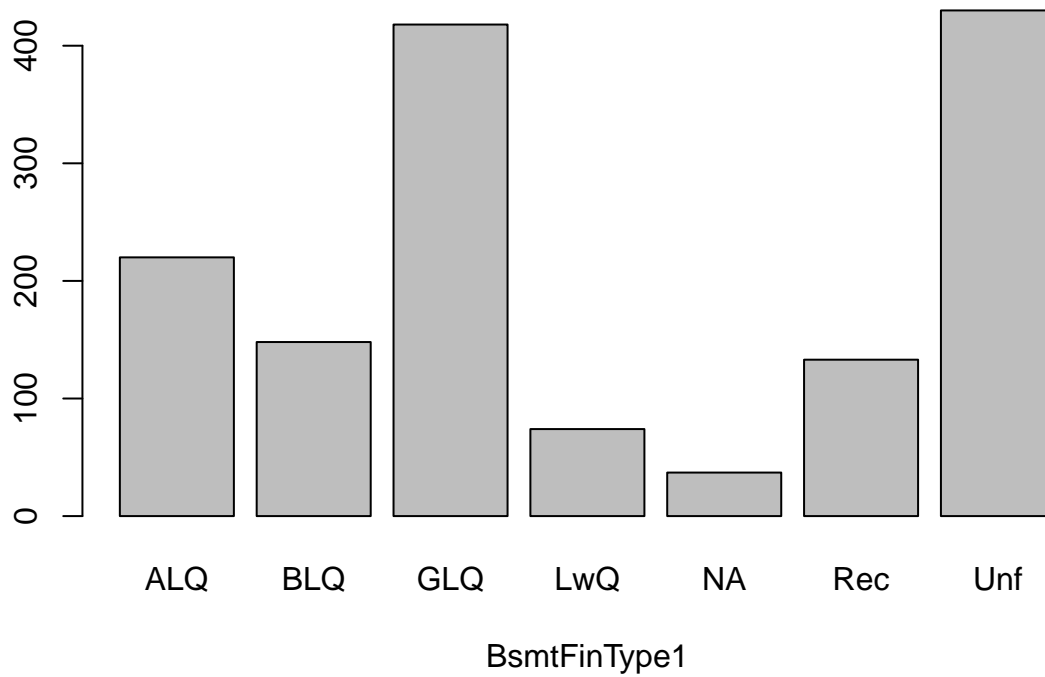
```
barplot(table(dataSet$BsmtExposure),main = "Distribución BsmtExposure",  
        xlab="BsmtExposure")
```

Distribución BsmtExposure



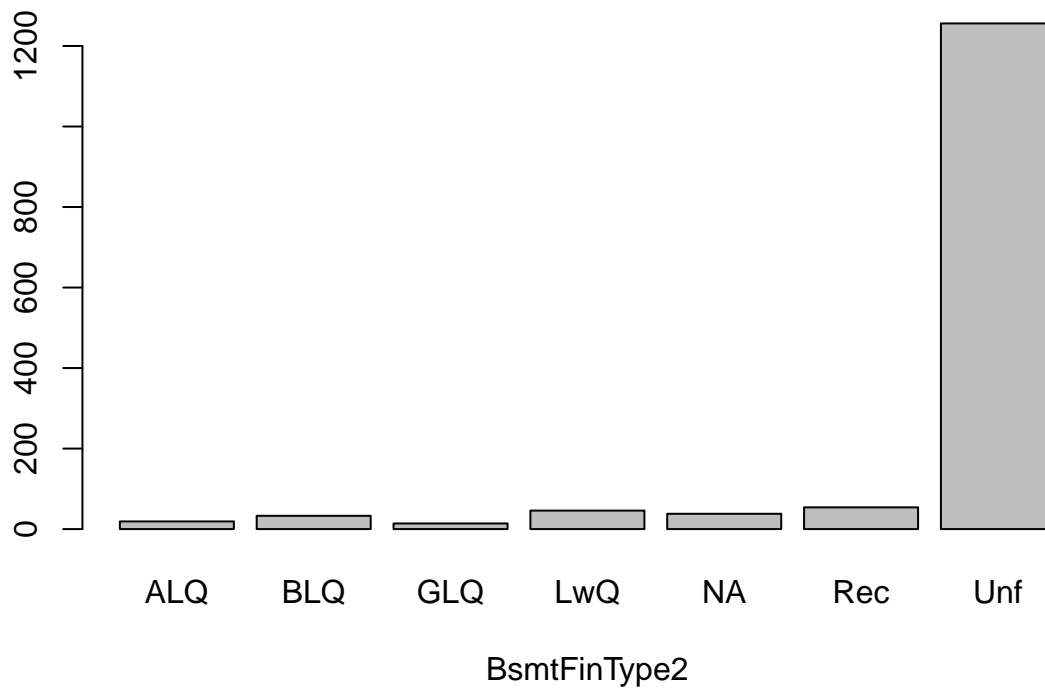
```
barplot(table(dataSet$BsmtFinType1),main = "Distribución BsmtFinType1",  
        xlab="BsmtFinType1")
```

Distribución BsmtFinType1

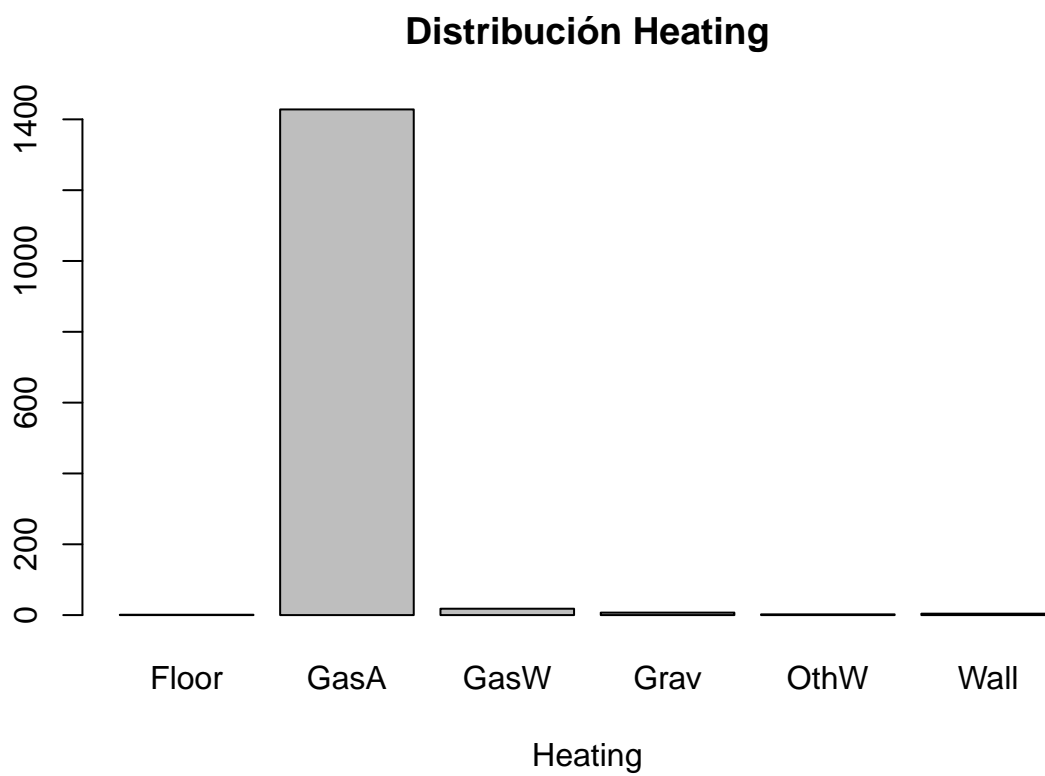


```
barplot(table(dataSet$BsmtFinType2),main = "Distribución BsmtFinType2",  
        xlab="BsmtFinType2")
```

Distribución BsmtFinType2

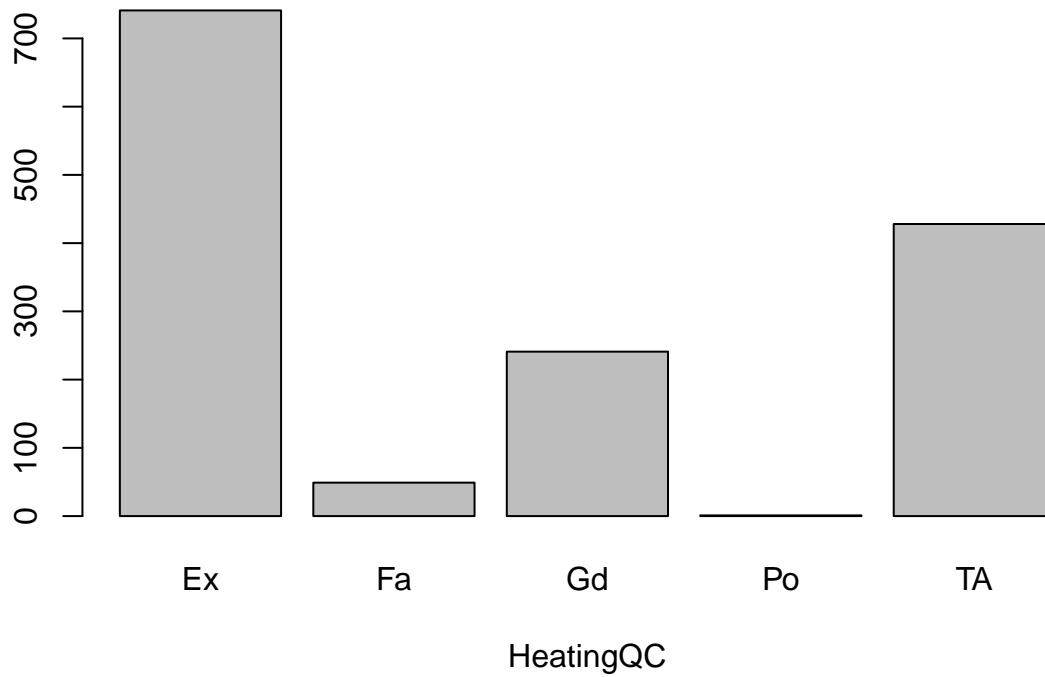


```
barplot(table(dataSet$Heating), main = "Distribución Heating",  
        xlab="Heating")
```



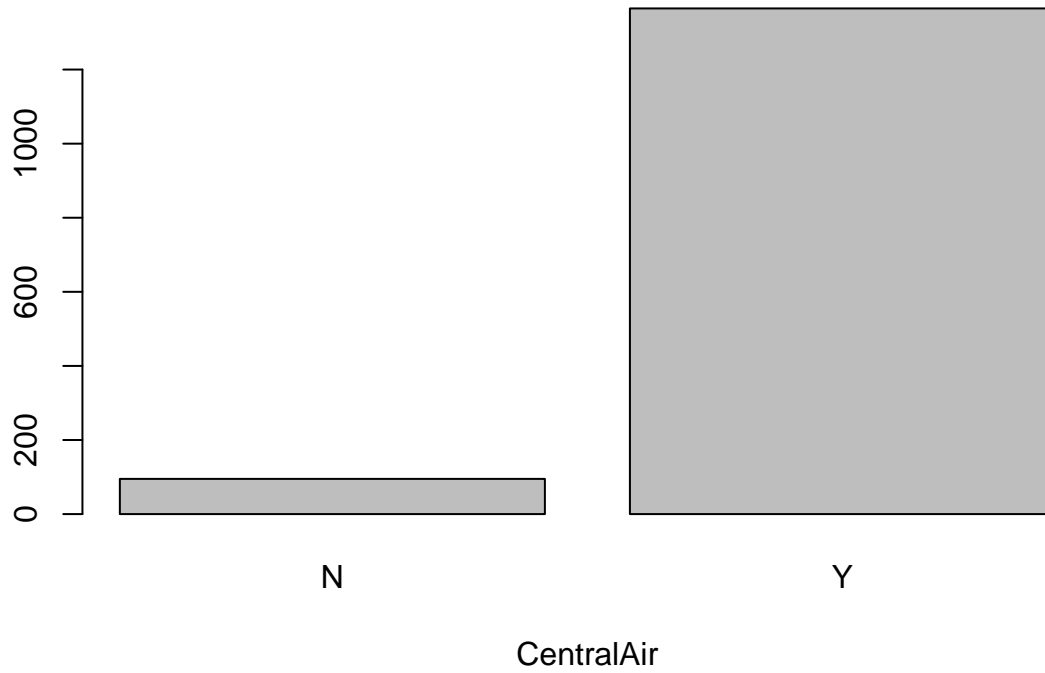
```
barplot(table(dataSet$HeatingQC),main = "Distribución HeatingQC",  
        xlab="HeatingQC")
```


Distribución HeatingQC



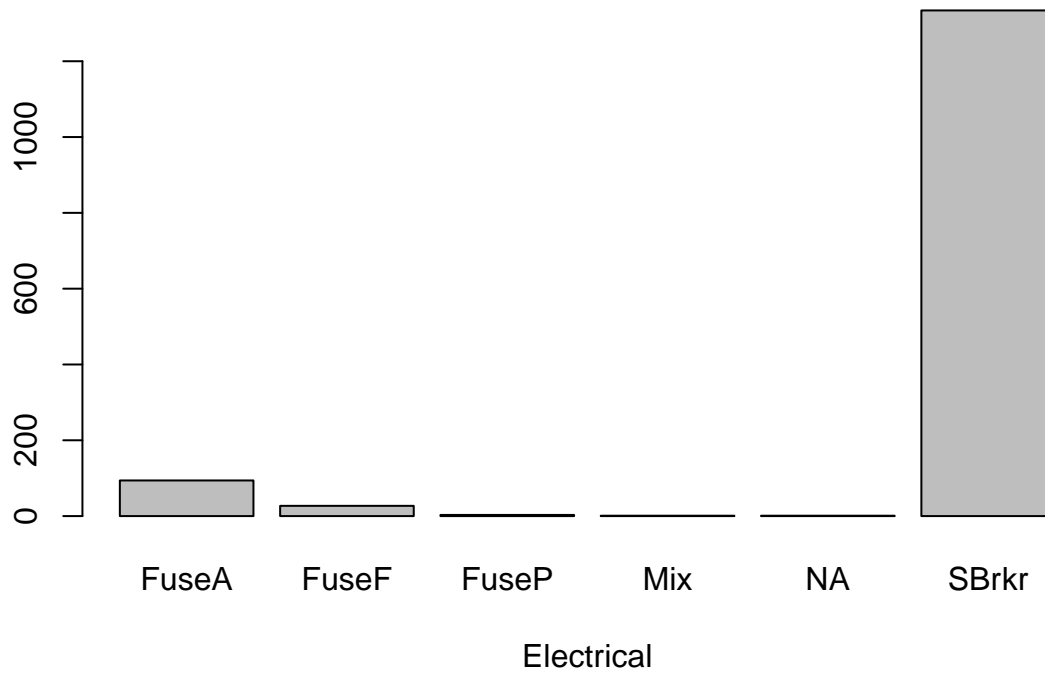
```
barplot(table(dataSet$CentralAir),main = "Distribución CentralAir",  
        xlab="CentralAir")
```

Distribución CentralAir



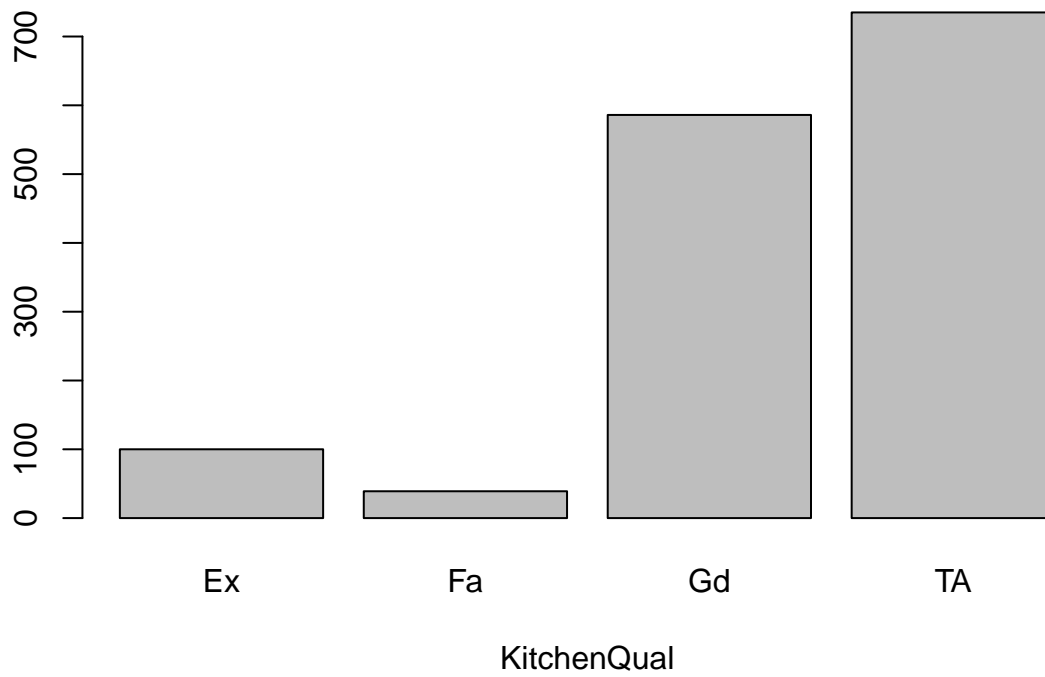
```
barplot(table(dataSet$Electrical),main = "Distribución Electrical",  
        xlab="Electrical")
```

Distribución Electrical



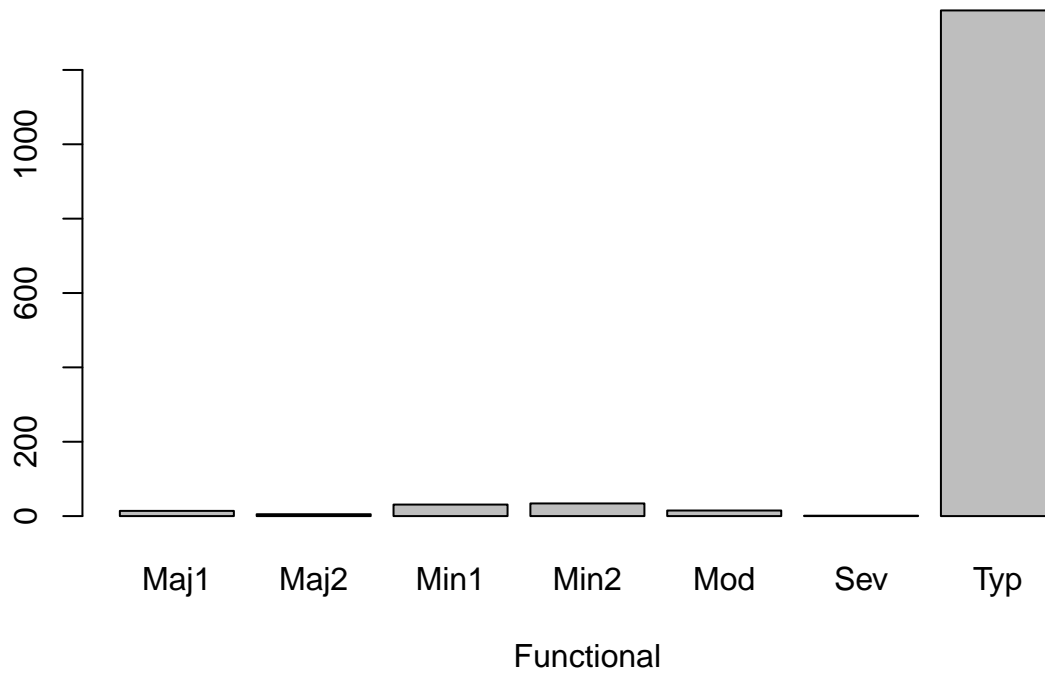
```
barplot(table(dataSet$KitchenQual),main = "Distribución KitchenQual",  
        xlab="KitchenQual")
```

Distribución KitchenQual



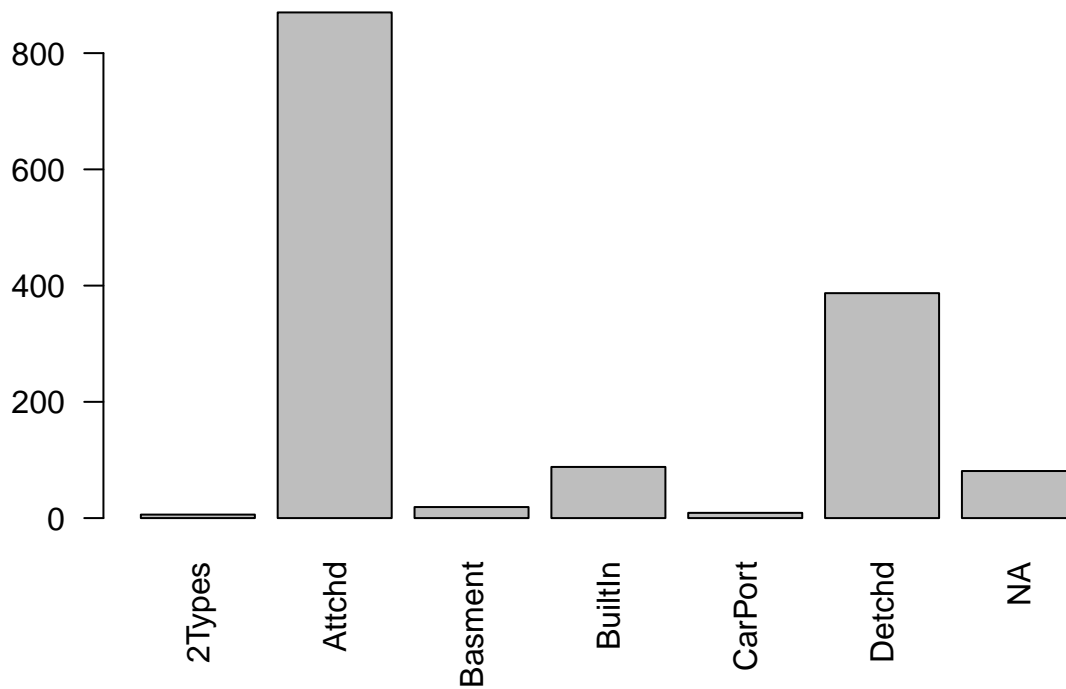
```
barplot(table(dataSet$Functional),main = "Distribución Functional",  
        xlab="Functional")
```

Distribución Funcional



```
barplot(table(dataSet$GarageType),main = "Distribución GarageType",  
        las=2)
```

Distribución GarageType



```
knitr::kable(
  table(dataSet$GarageYrBlt), caption = 'Tabla de frecuencia GarageYrBlt'
)
```

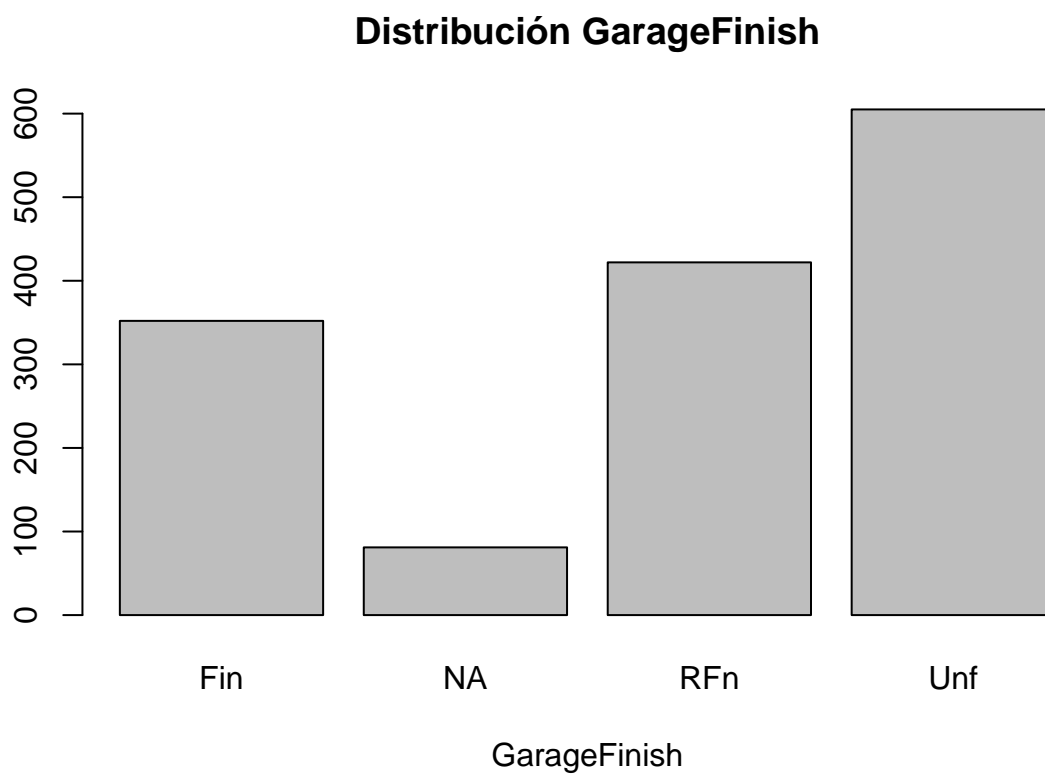
Table 3: Tabla de frecuencia GarageYrBlt

Var1	Freq
1900	1
1906	1
1908	1
1910	3
1914	2
1915	2
1916	5
1918	2
1920	14
1921	3
1922	5
1923	3
1924	3
1925	10
1926	6
1927	1
1928	4
1929	2

Var1	Freq
1930	8
1931	4
1932	3
1933	1
1934	2
1935	4
1936	5
1937	2
1938	3
1939	9
1940	14
1941	10
1942	2
1945	4
1946	4
1947	2
1948	11
1949	8
1950	24
1951	6
1952	3
1953	12
1954	19
1955	13
1956	16
1957	20
1958	21
1959	17
1960	19
1961	13
1962	21
1963	16
1964	18
1965	21
1966	21
1967	15
1968	26
1969	15
1970	20
1971	13
1972	14
1973	14
1974	18
1975	9
1976	29
1977	35
1978	19
1979	15
1980	15
1981	10
1982	4
1983	7

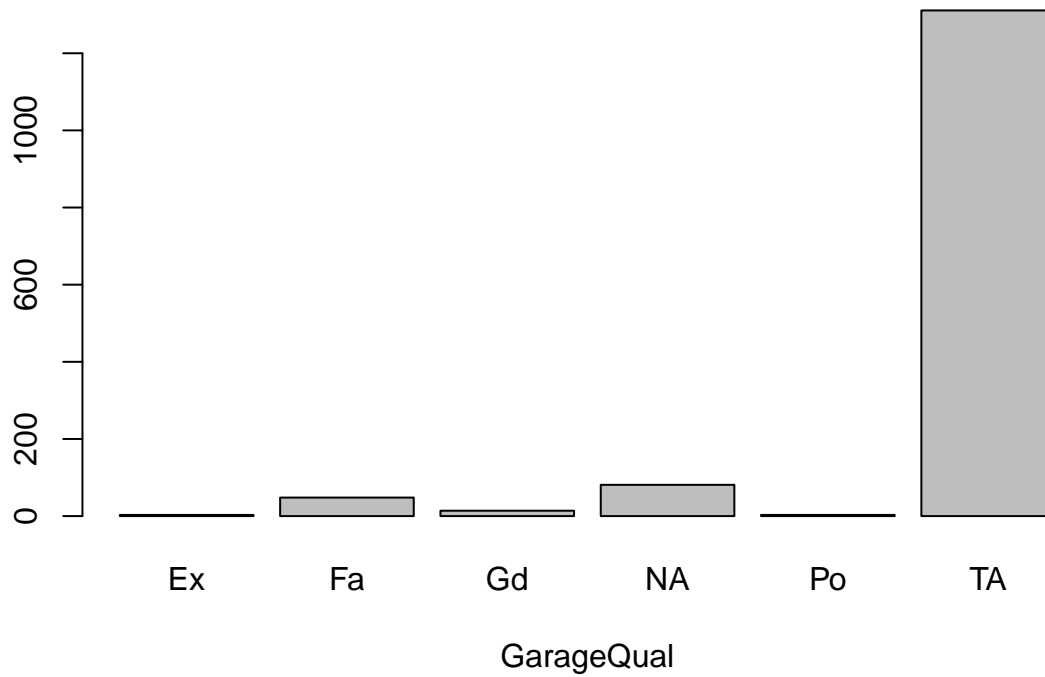
Var1	Freq
1984	8
1985	10
1986	6
1987	11
1988	14
1989	10
1990	16
1991	9
1992	13
1993	22
1994	18
1995	18
1996	20
1997	19
1998	31
1999	30
2000	27
2001	20
2002	26
2003	50
2004	53
2005	65
2006	59
2007	49
2008	29
2009	21
2010	3
NA	81

```
barplot(table(dataSet$GarageFinish),main = "Distribución GarageFinish",
        xlab="GarageFinish")
```

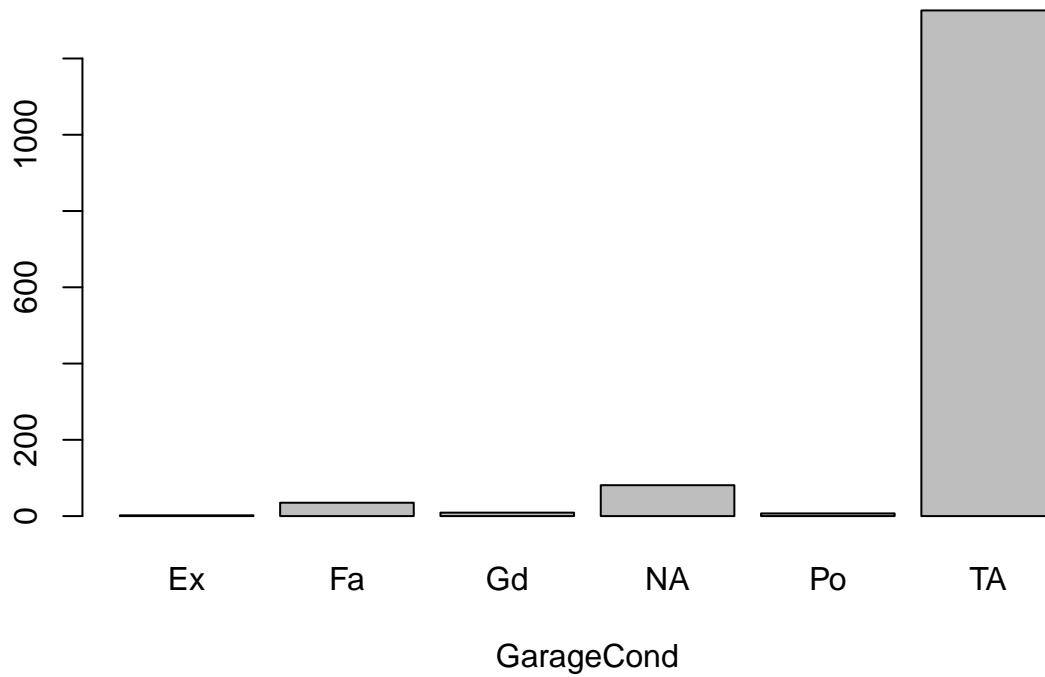
```
barplot(table(dataSet$GarageQual),main = "Distribución GarageQual",  
        xlab="GarageQual")
```

Distribución GarageQual



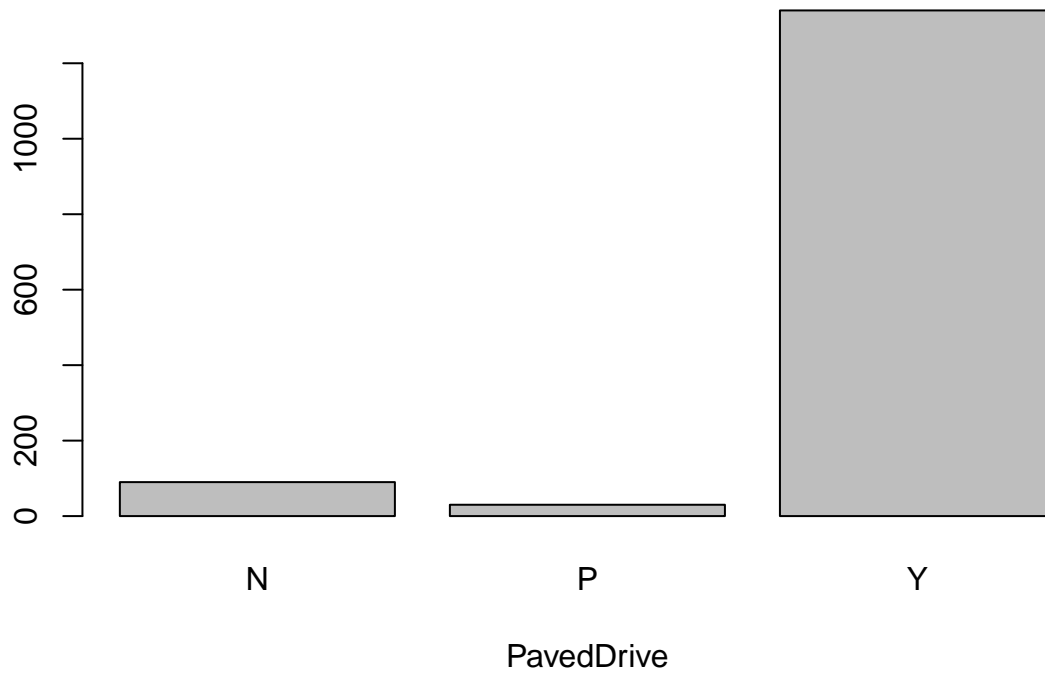
```
barplot(table(dataSet$GarageCond),main = "Distribución GarageCond",  
        xlab="GarageCond")
```

Distribución GarageCond

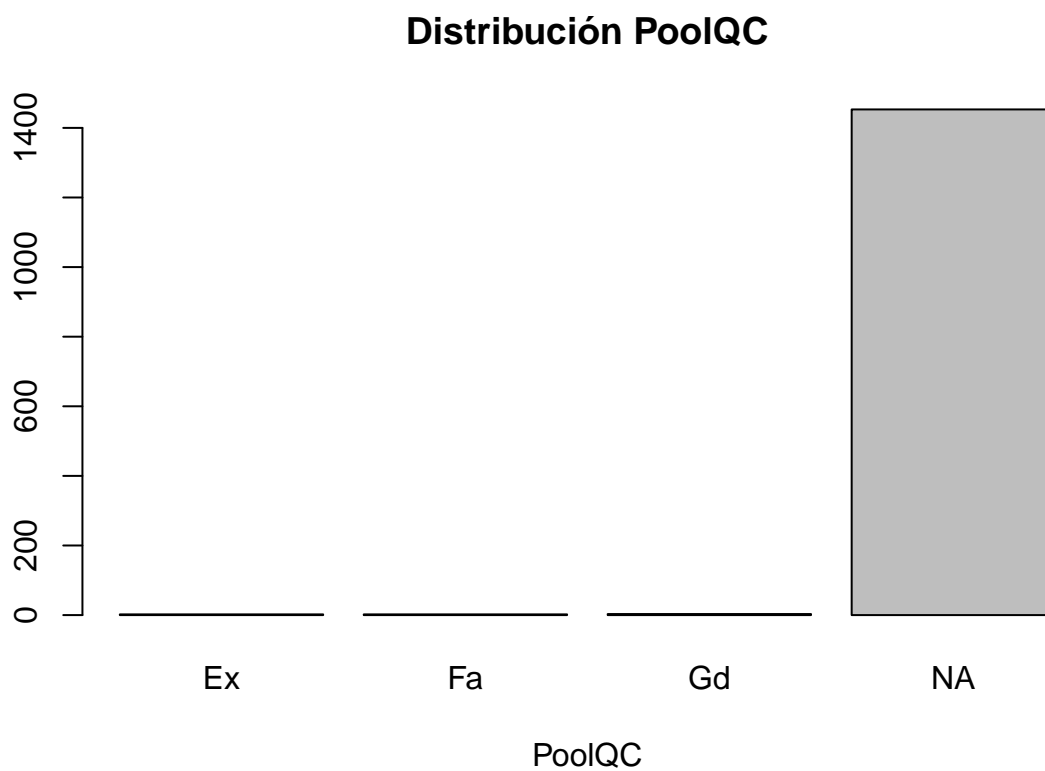


```
barplot(table(dataSet$PavedDrive),main = "Distribución PavedDrive",  
        xlab="PavedDrive")
```

Distribución PavedDrive

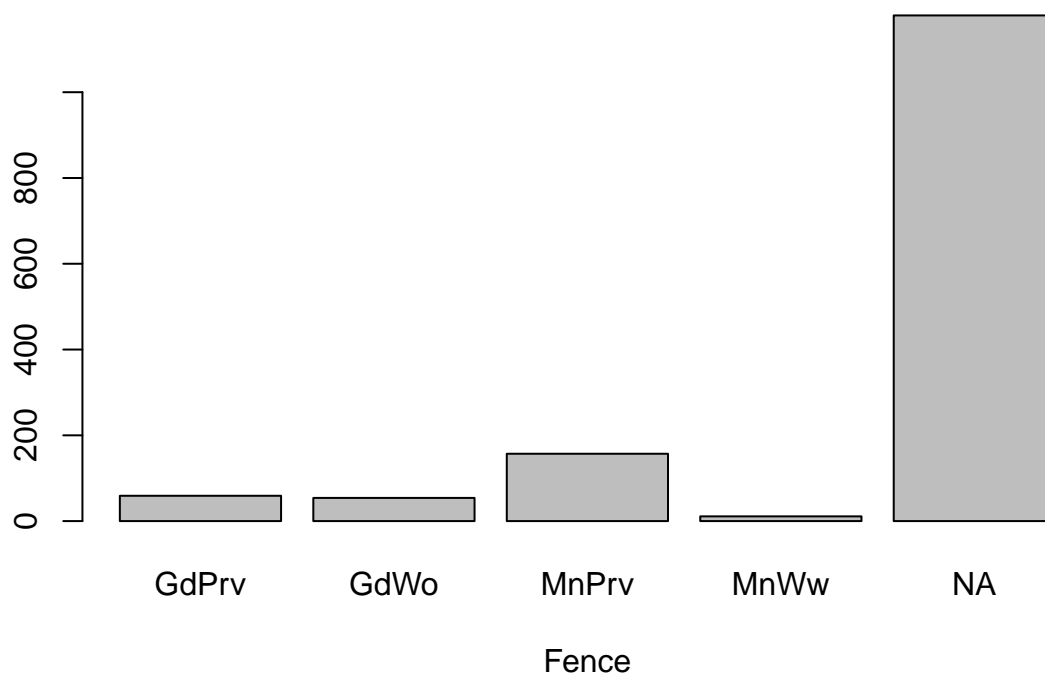


```
barplot(table(dataSet$PoolQC),main = "Distribución PoolQC",  
        xlab="PoolQC")
```

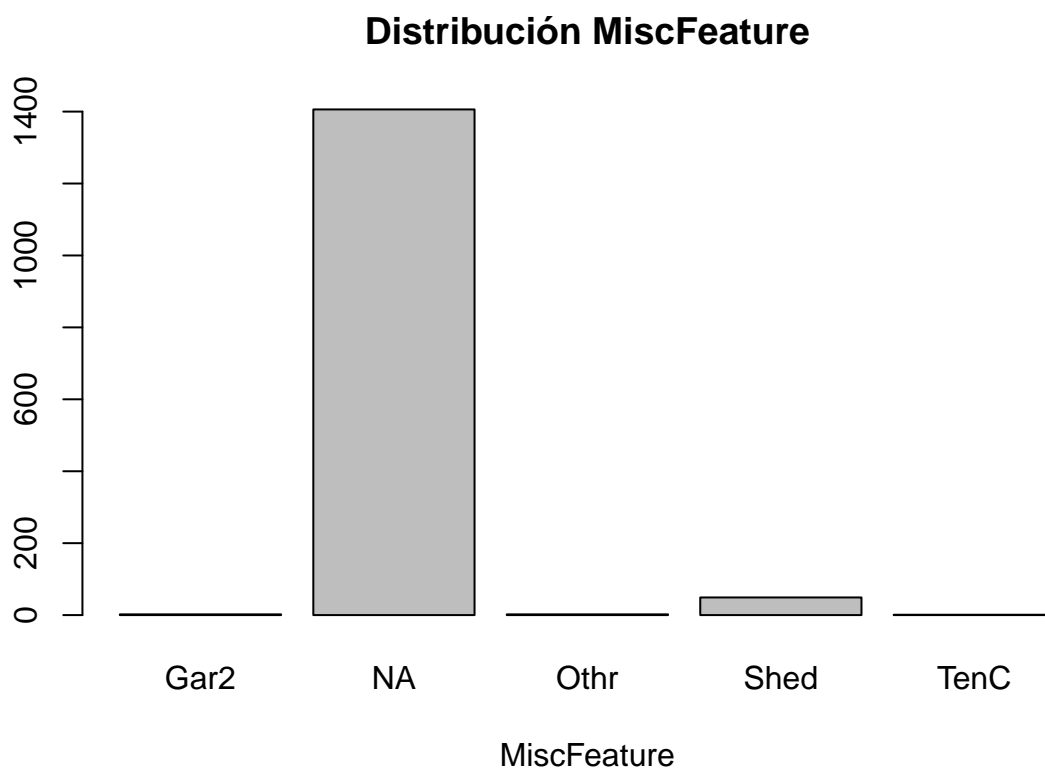


```
barplot(table(dataSet$Fence),main = "Distribución Fence",  
        xlab="Fence")
```

Distribución Fence

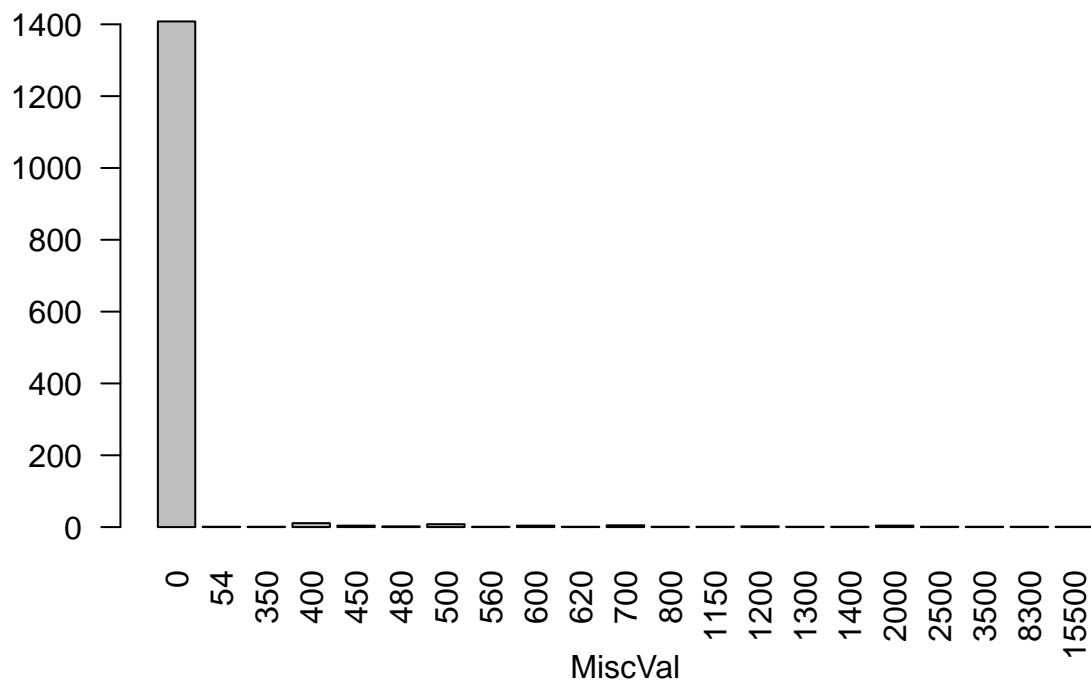


```
barplot(table(dataSet$MiscFeature),main = "Distribución MiscFeature",  
        xlab="MiscFeature")
```

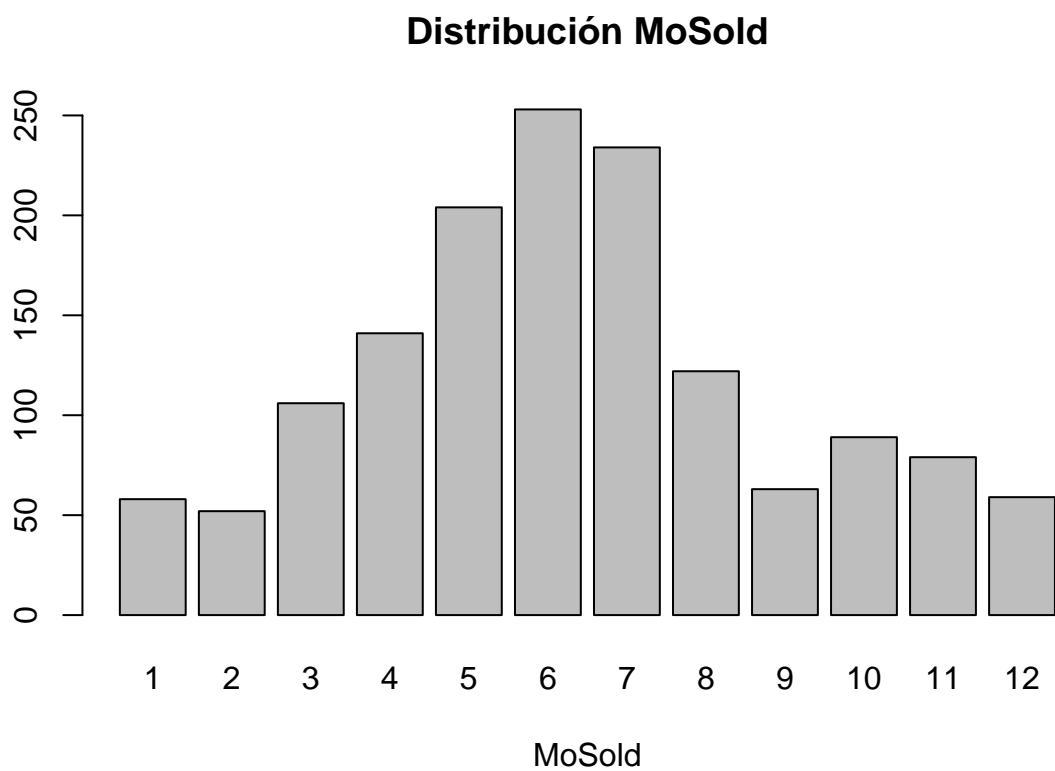


```
barplot(table(dataSet$MiscVal),main = "Distribución MiscVal",  
        xlab="MiscVal",las=2)
```

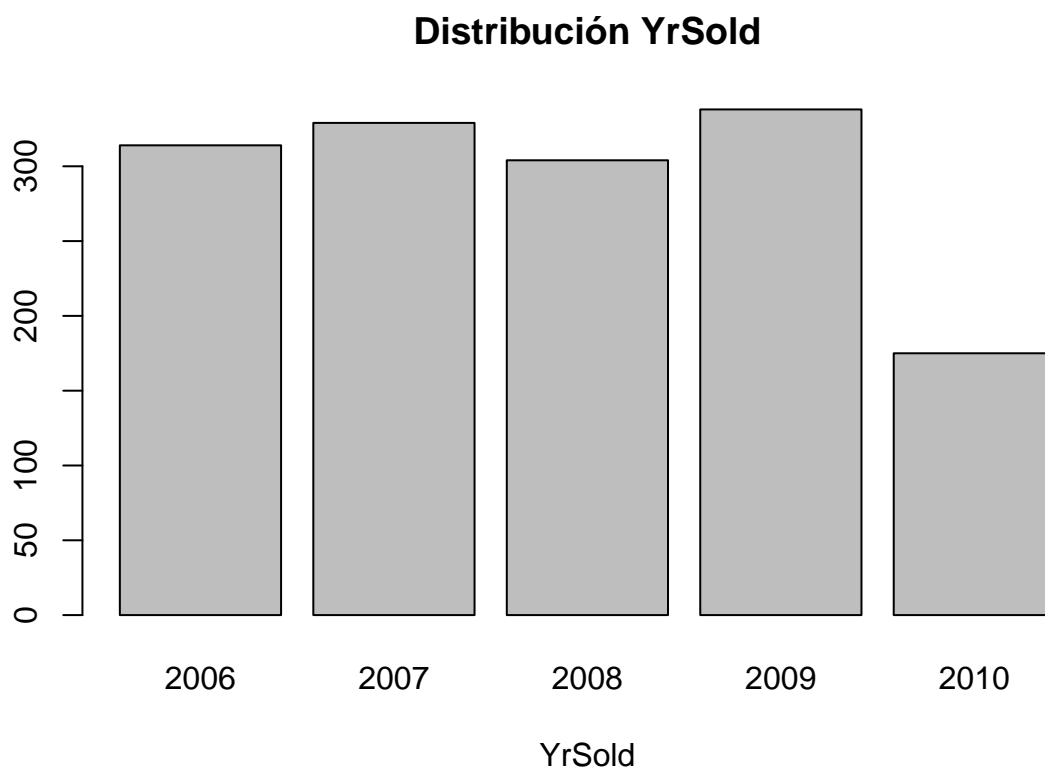
Distribución MiscVal



```
barplot(table(dataSet$MoSold),main = "Distribución MoSold",  
        xlab="MoSold")
```

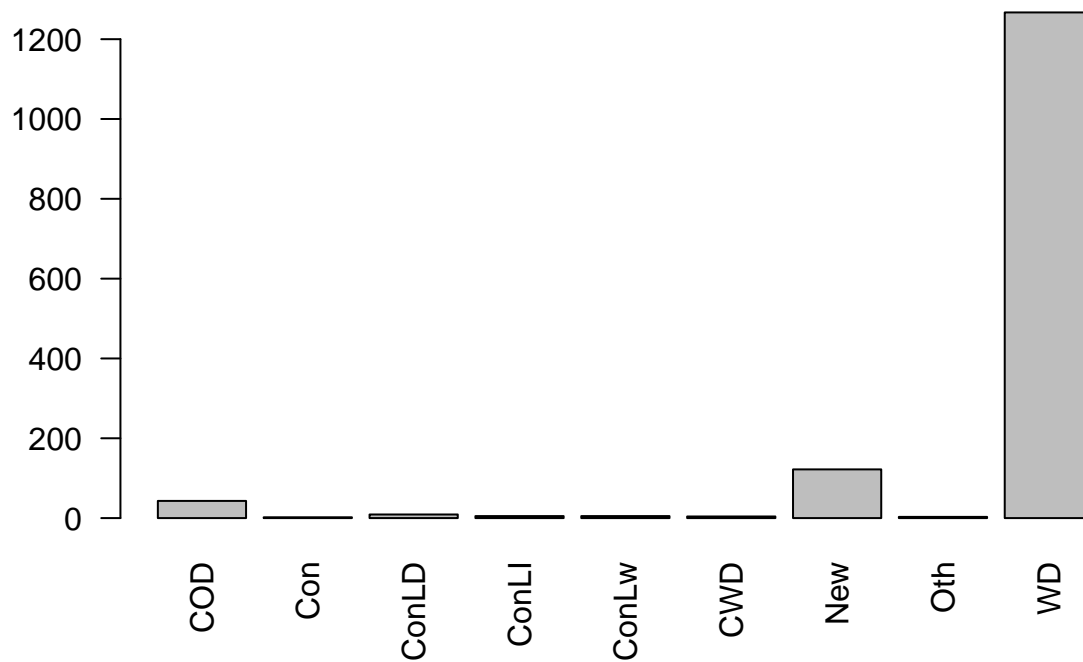



```
barplot(table(dataSet$YrSold),main = "Distribución YrSold",  
        xlab="YrSold")
```

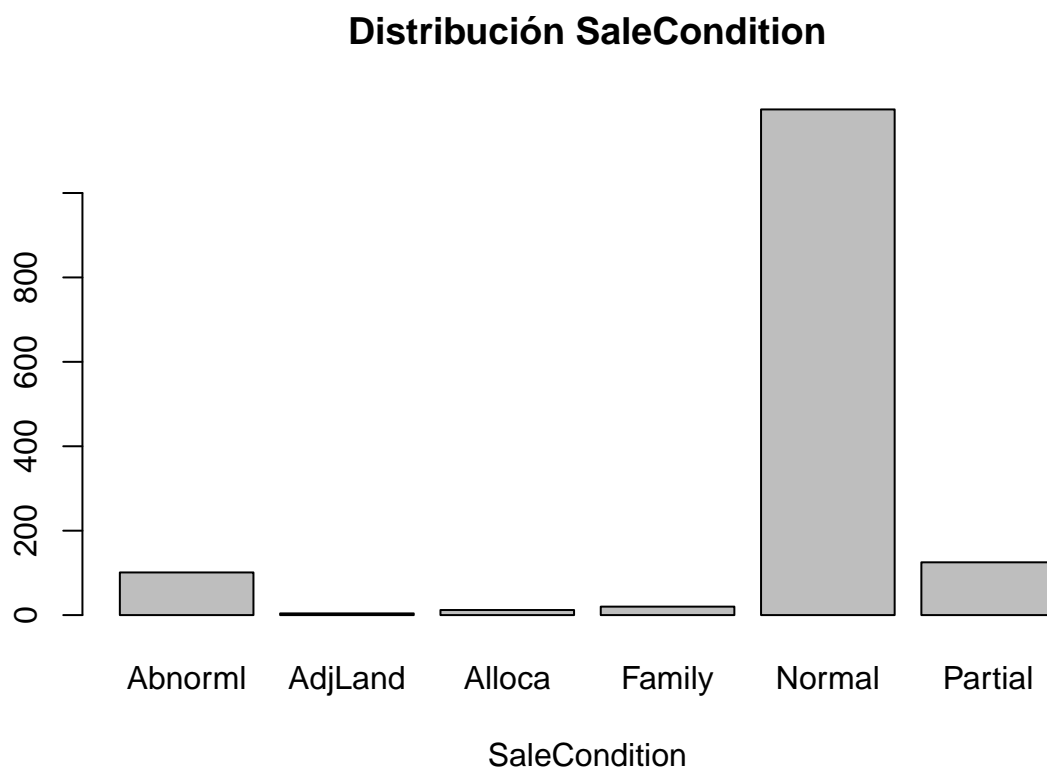


```
barplot(table(dataSet$SaleType),main = "Distribución SaleType",  
las=2)
```

Distribución SaleType



```
barplot(table(dataSet$SaleCondition),main = "Distribución SaleCondition",  
        xlab="SaleCondition")
```



Como podemos observar en la mayoría de las variables no hay valores NAs y si lo hay es mínima. Las únicas variables que tienen una fuerte presencia de estos son las siguientes: calidad de la piscina (PoolQC), calidad de la cerca (Fence), características extras (MiscFeature). Podemos suponer que estas variables tienen ausencia de valores debido a que son características que no todas las casas tienen.