

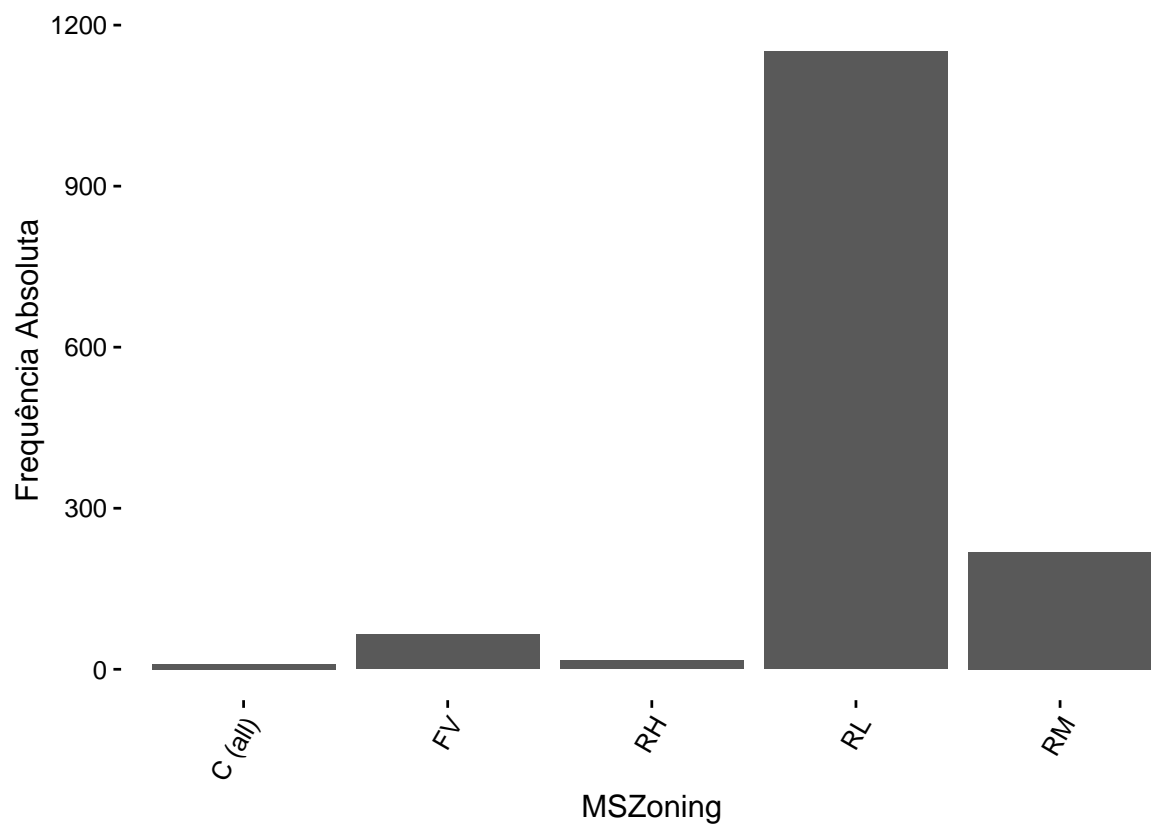
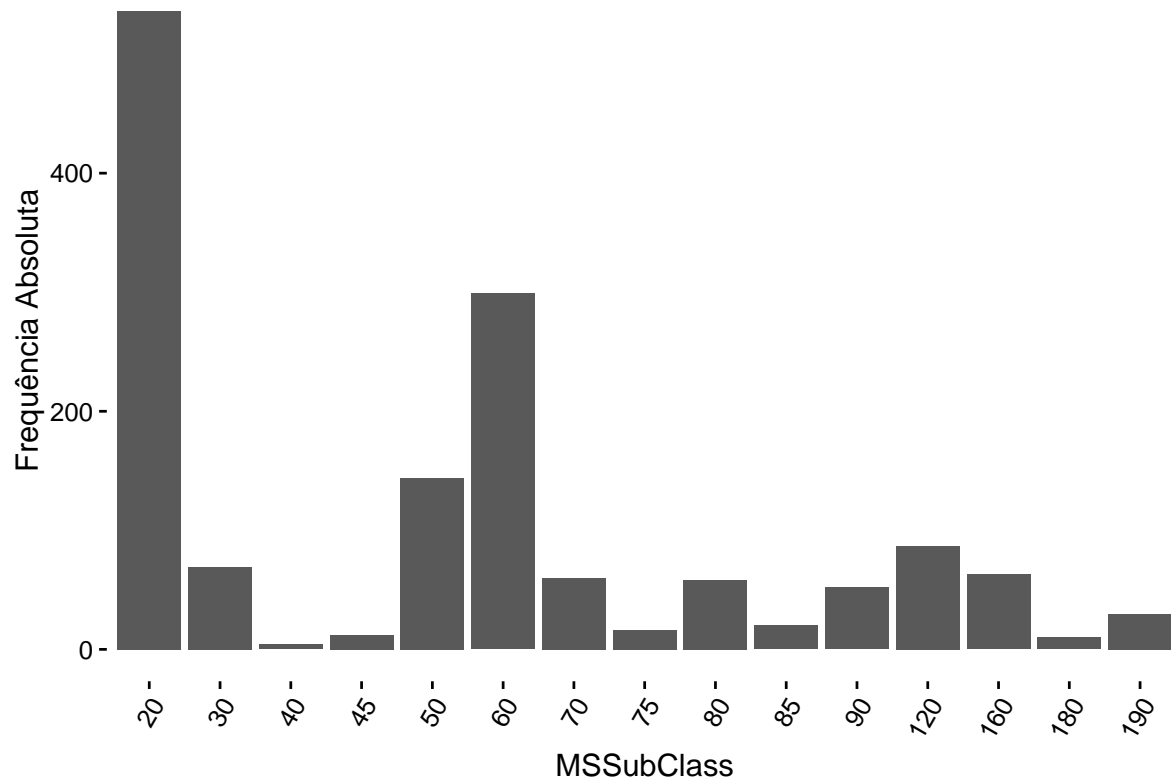
Análise Exploratória

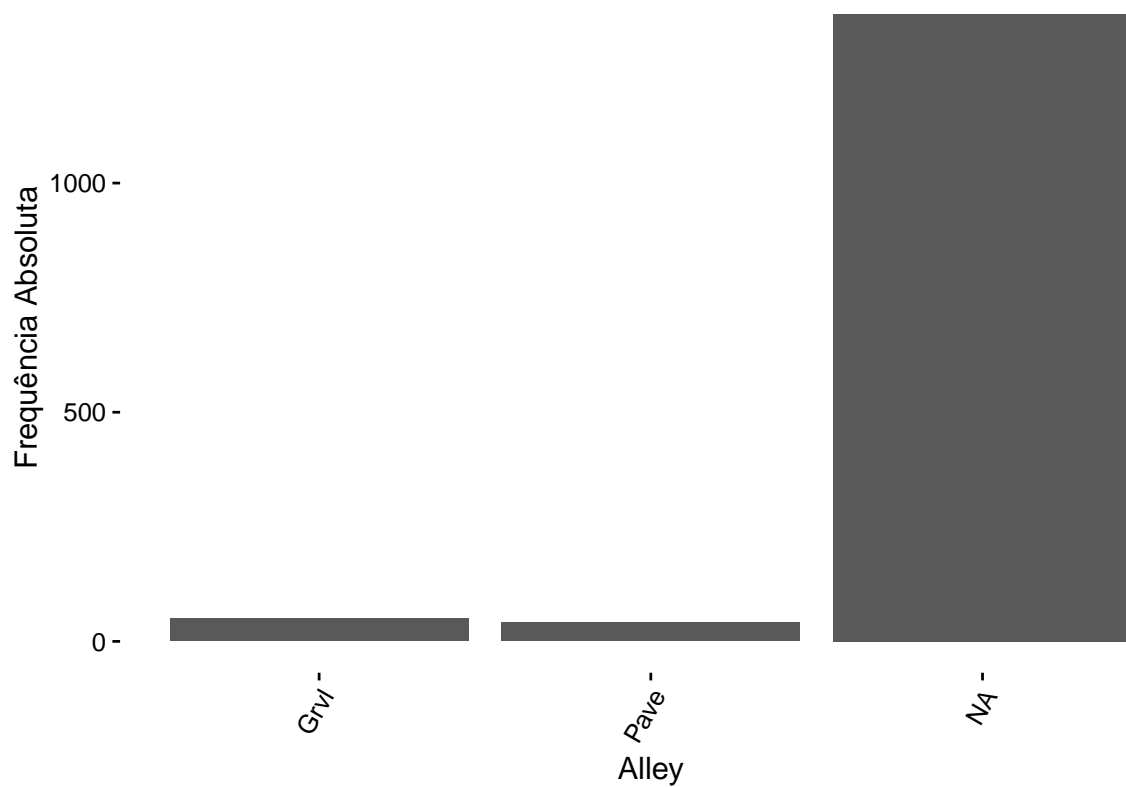
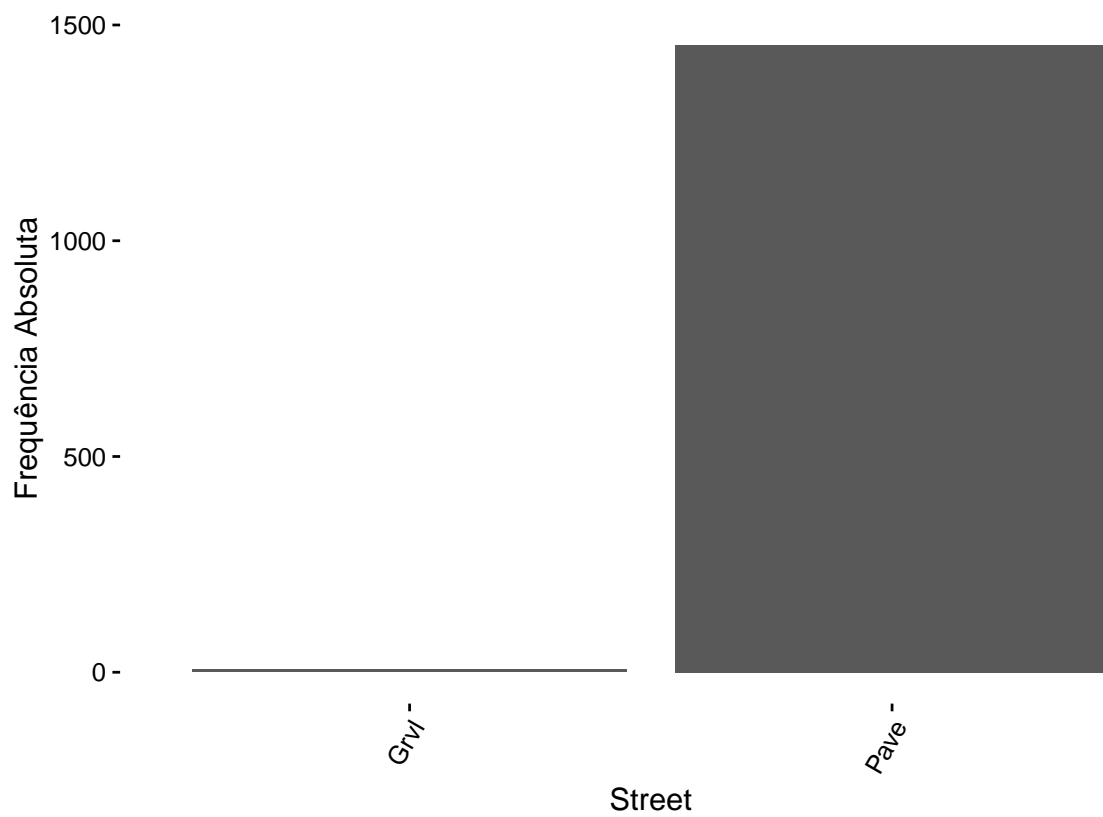
Ossada R. e Hamaguchi L.

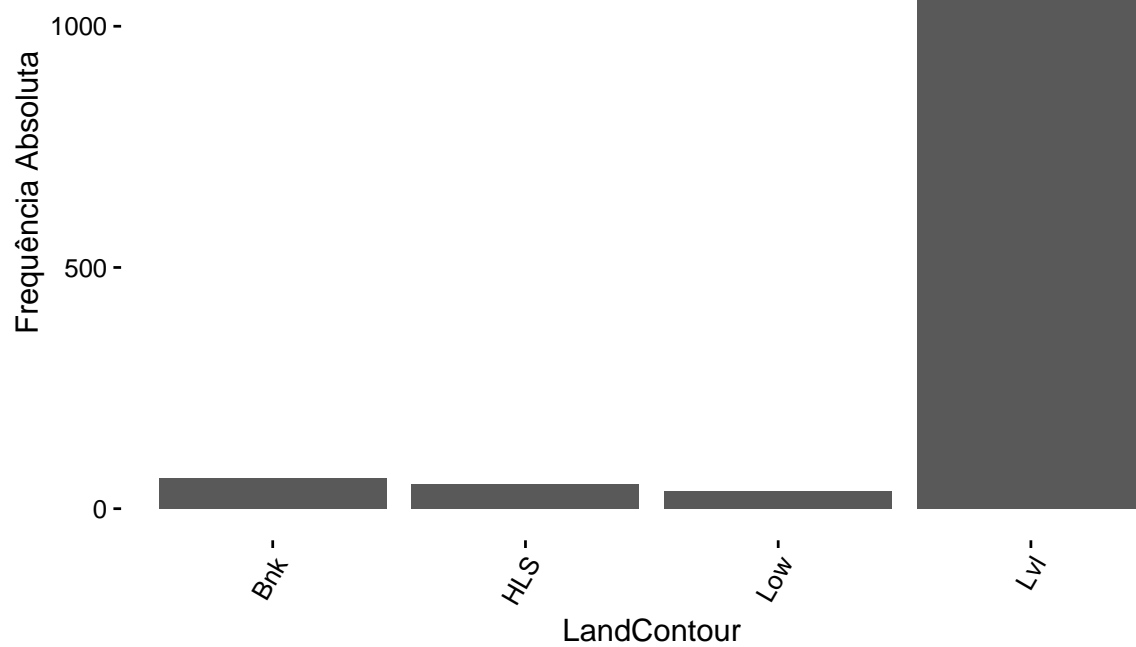
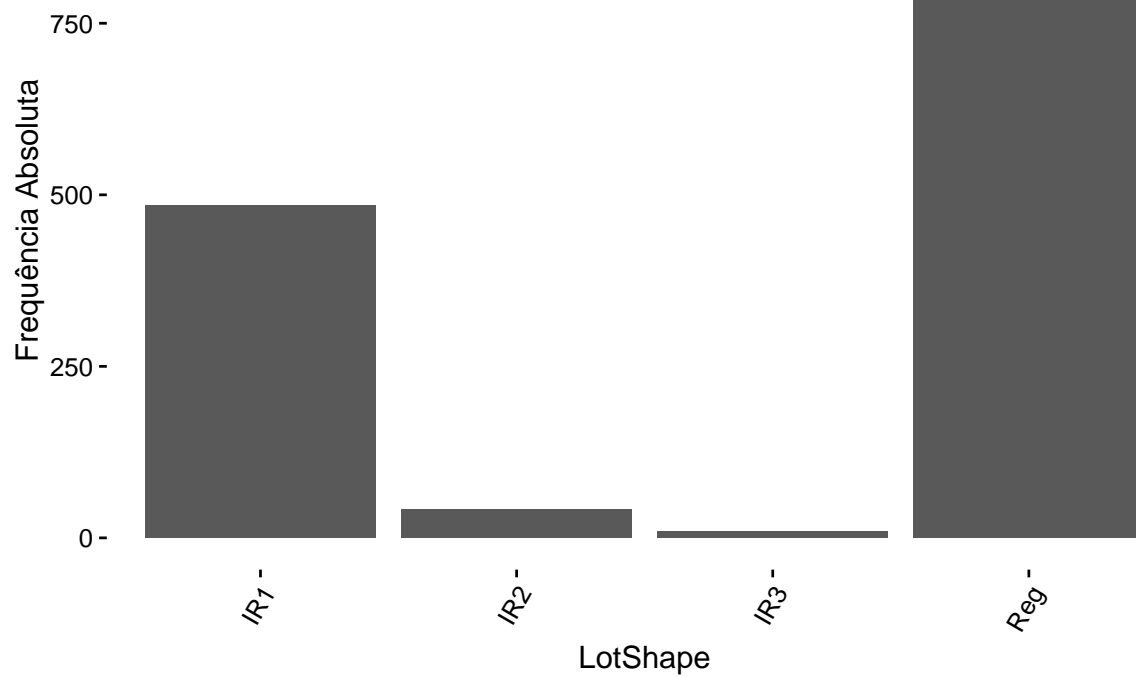
25 de julho de 2017

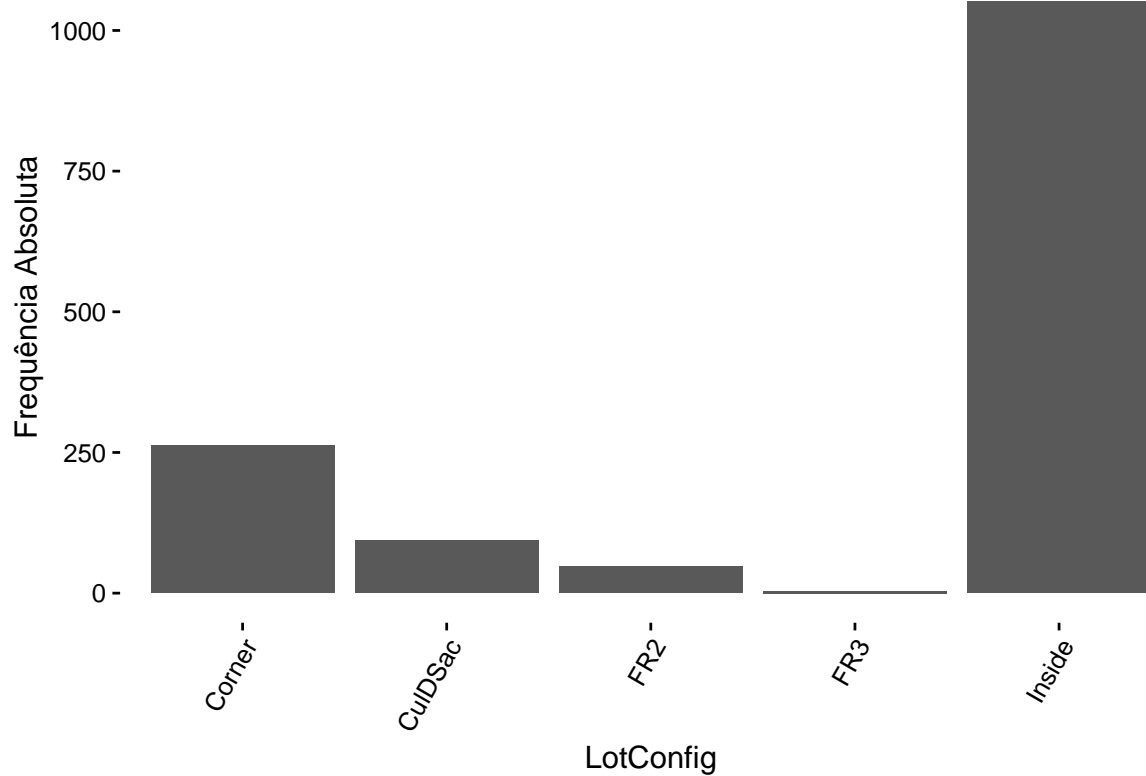
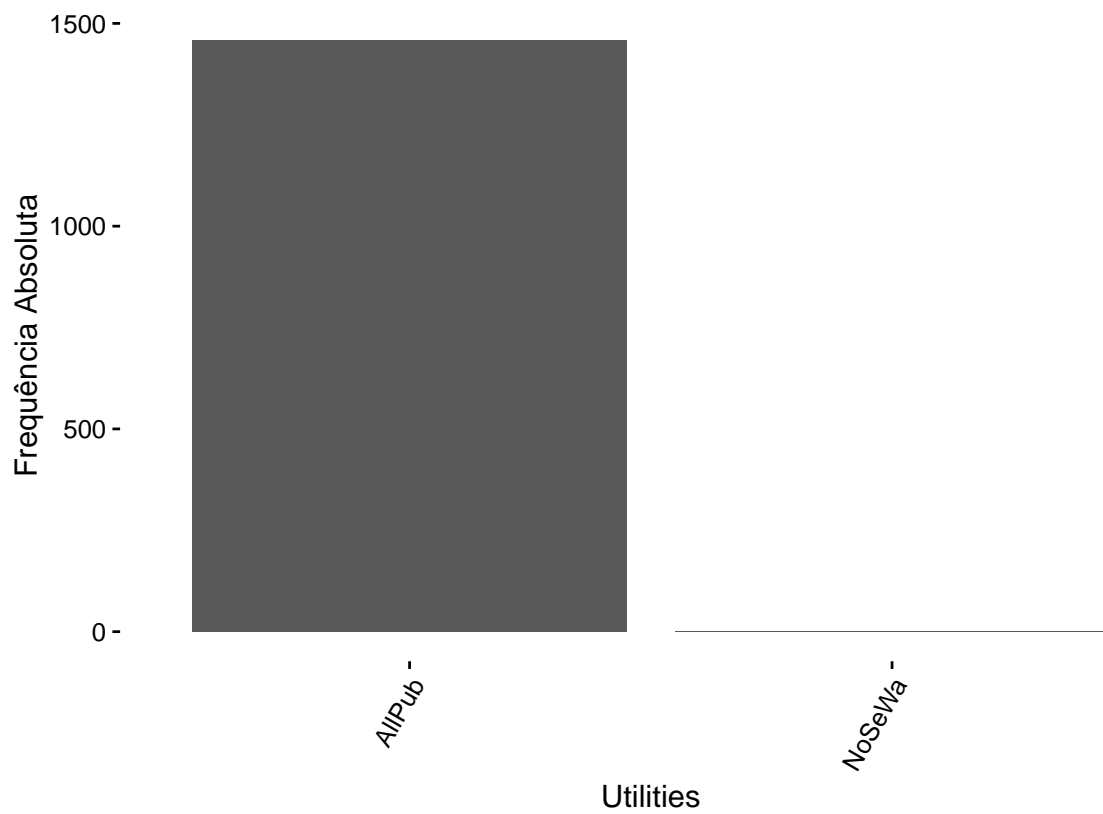
1 ANÁLISE EXPLORATÓRIA DOS DADOS

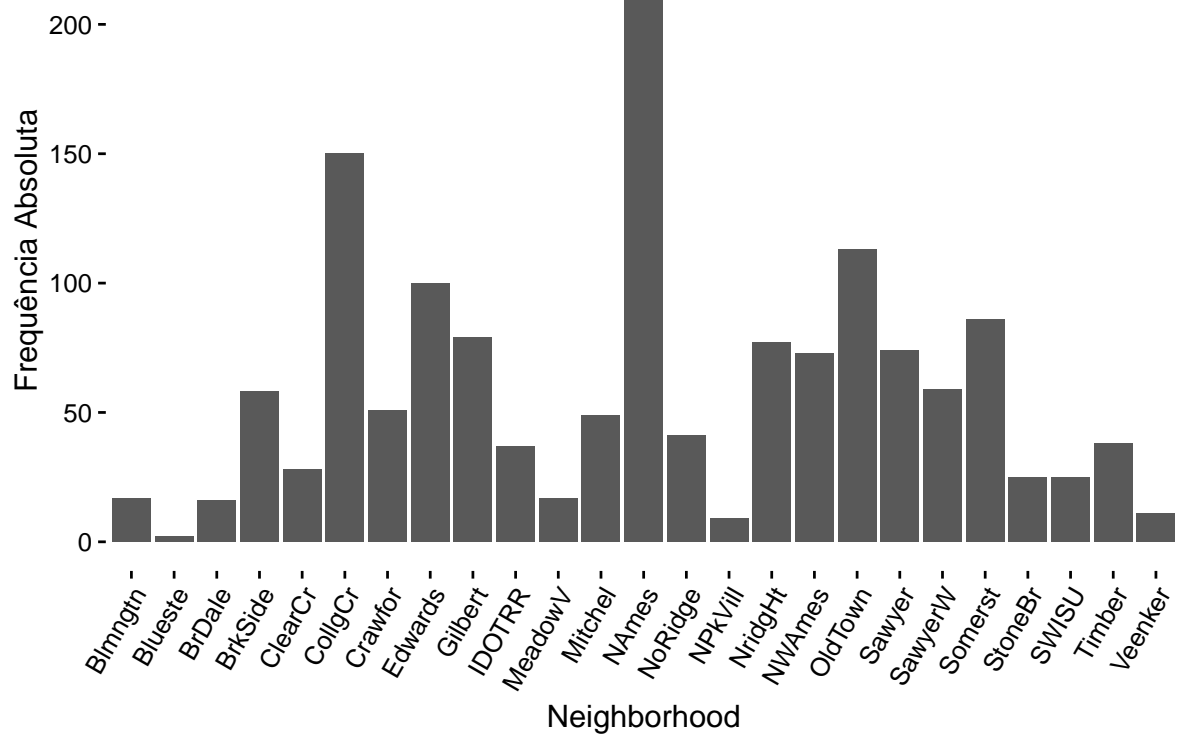
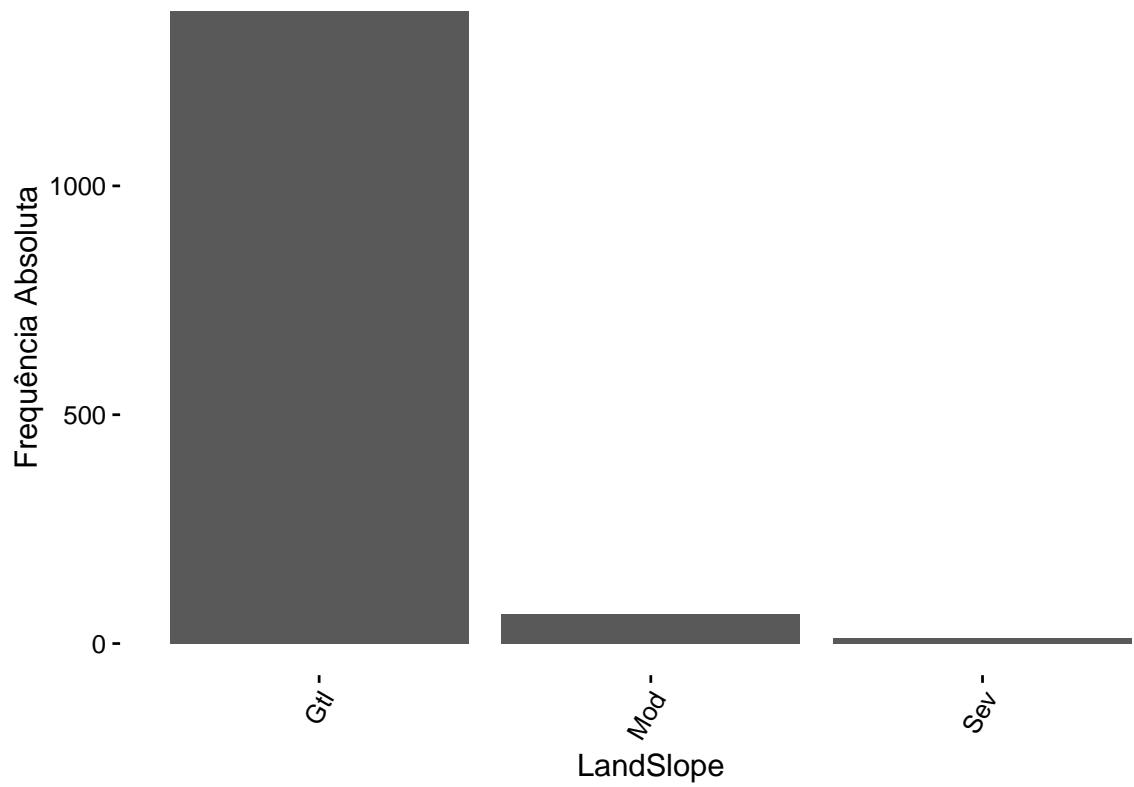
1.1 GRÁFICOS DAS VARIÁVEIS QUALITATIVAS

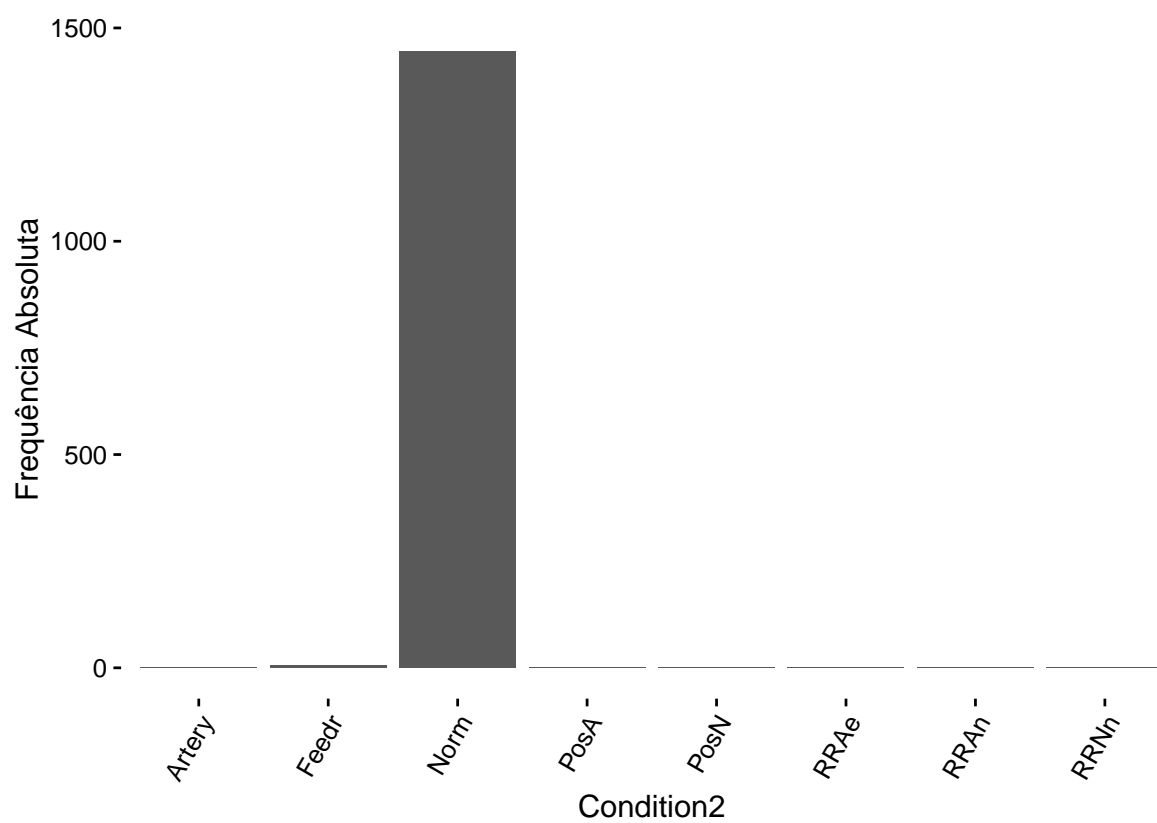
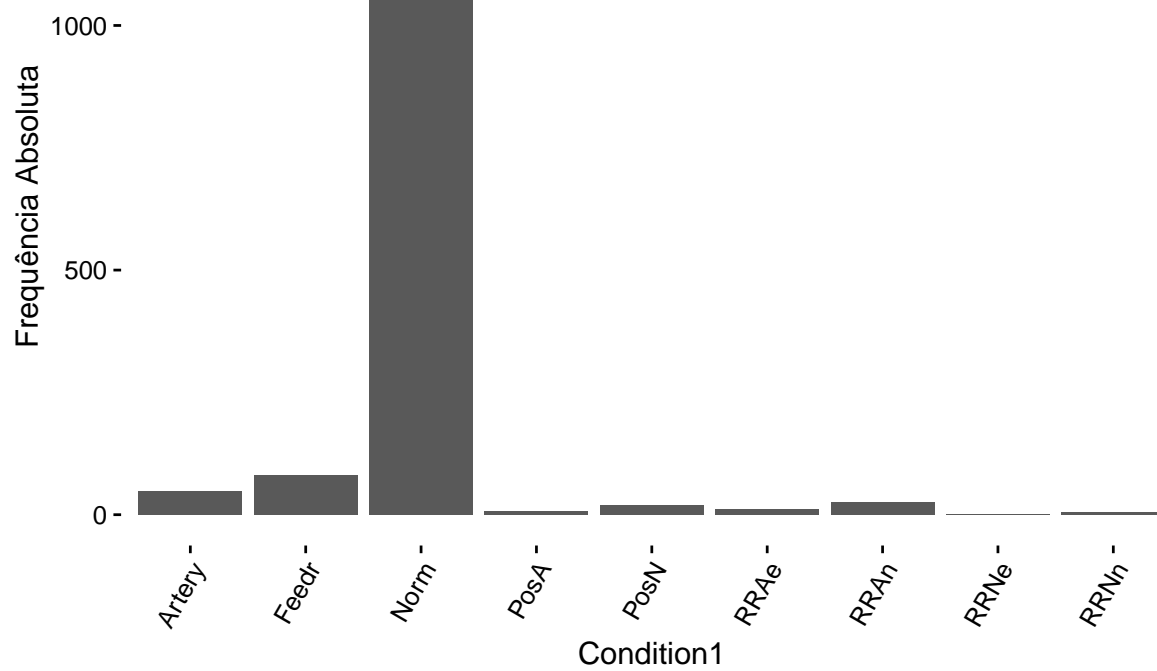


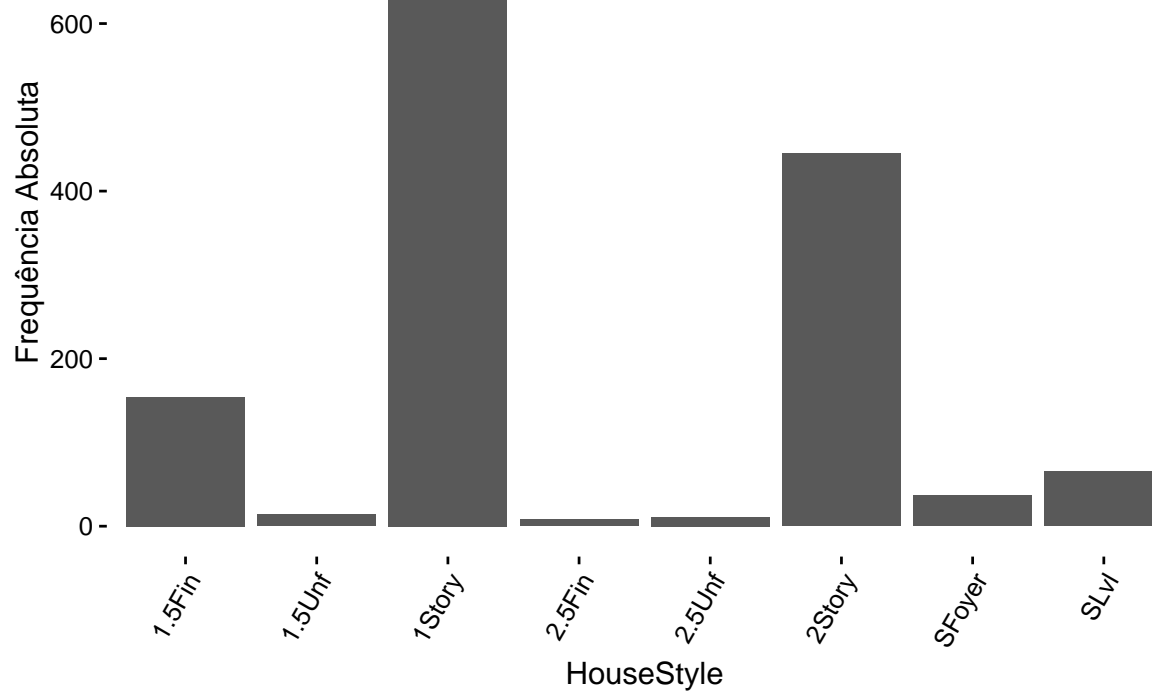
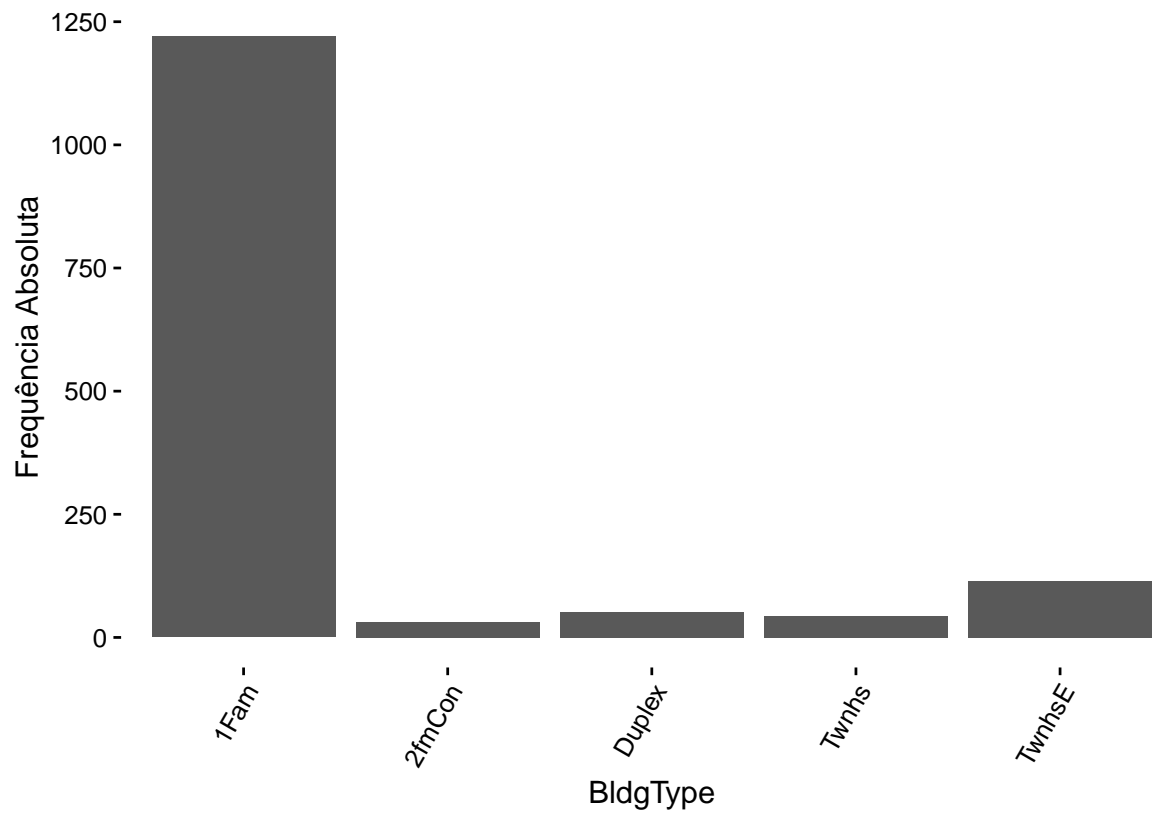


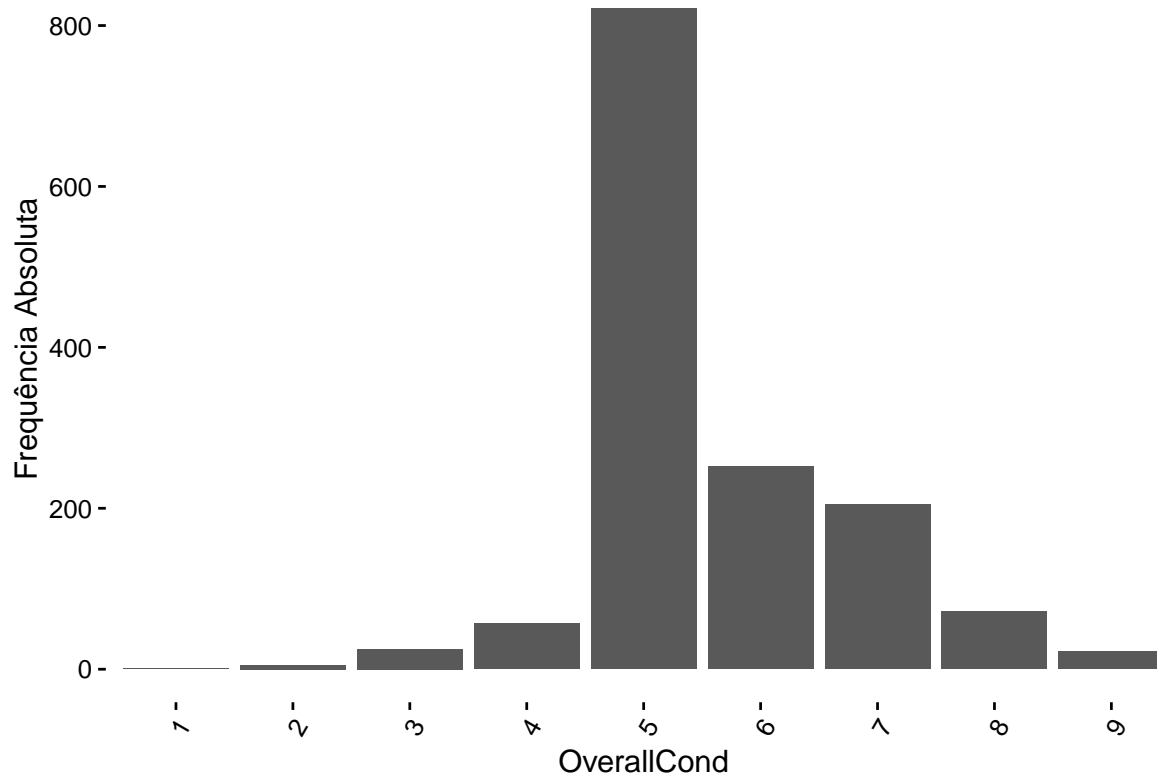
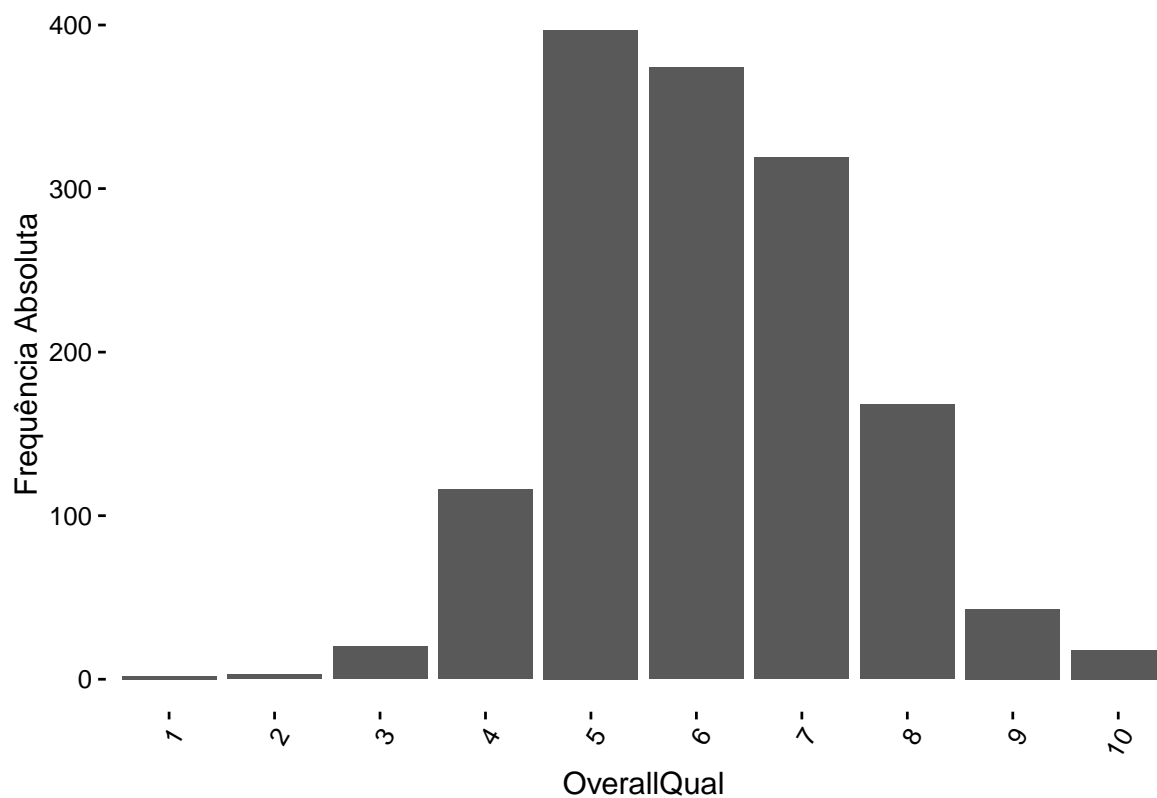


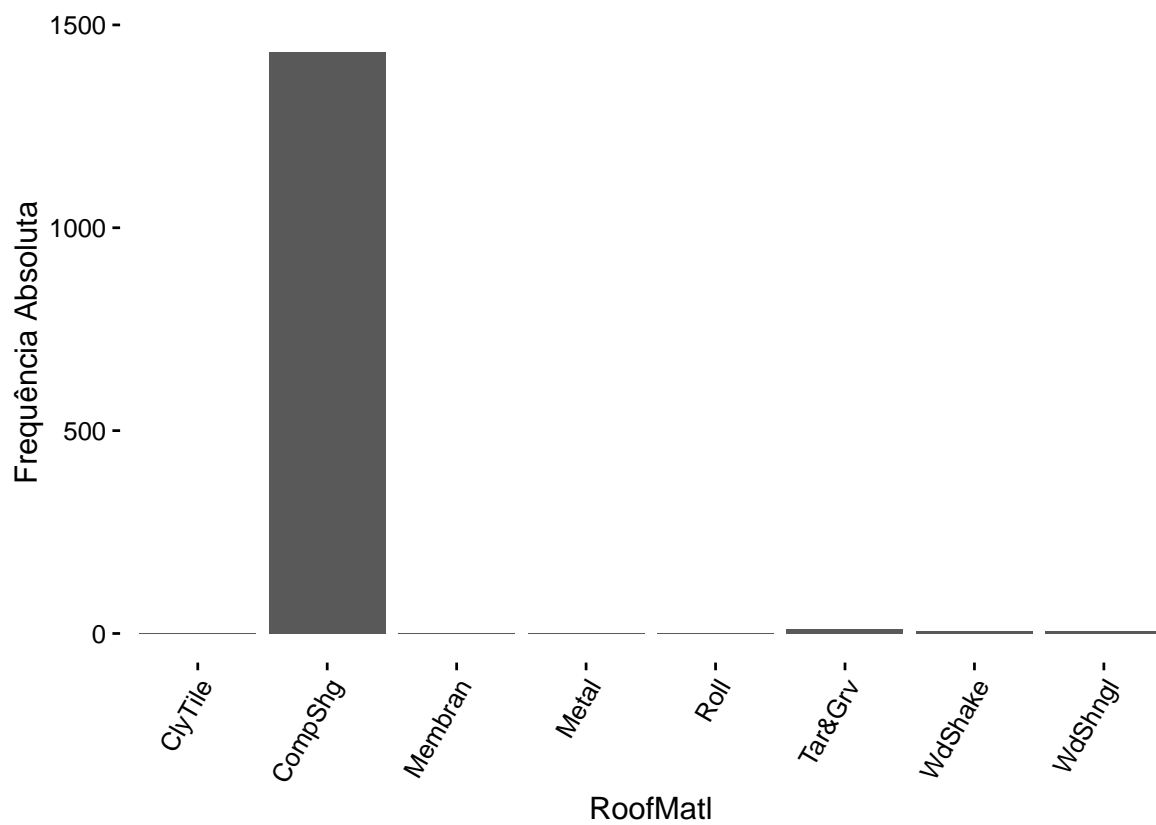
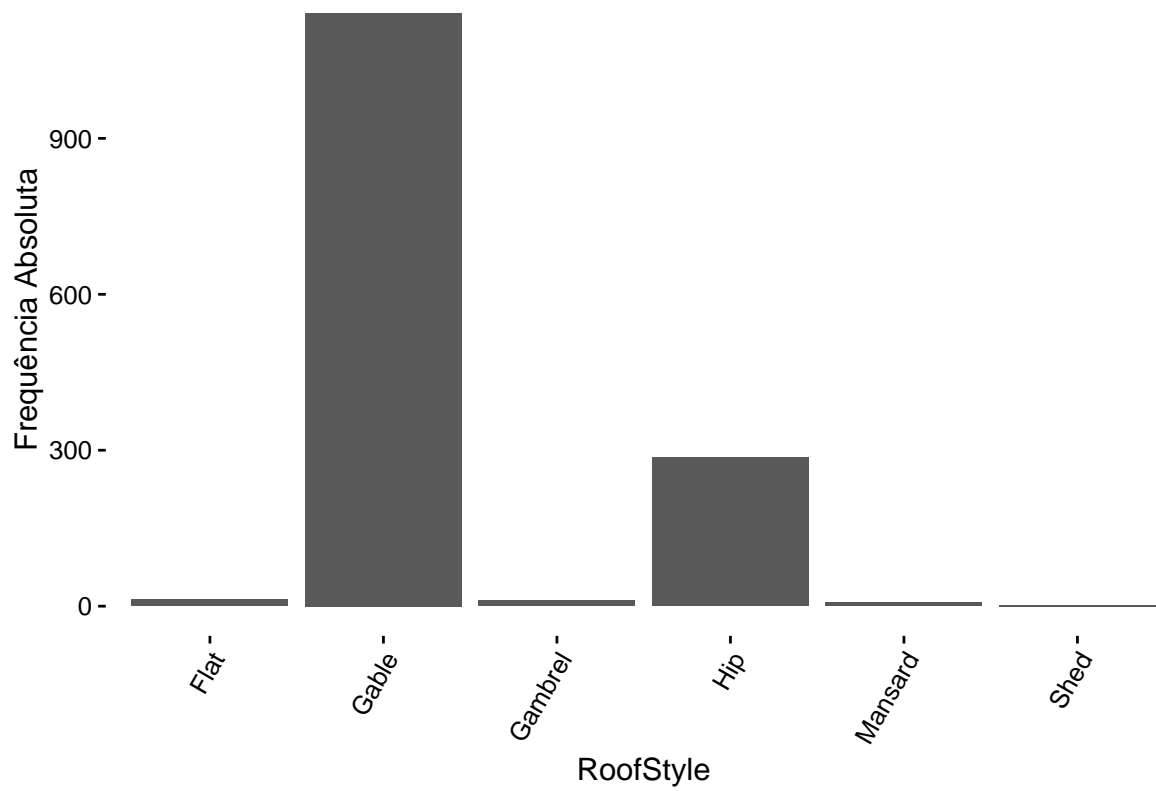


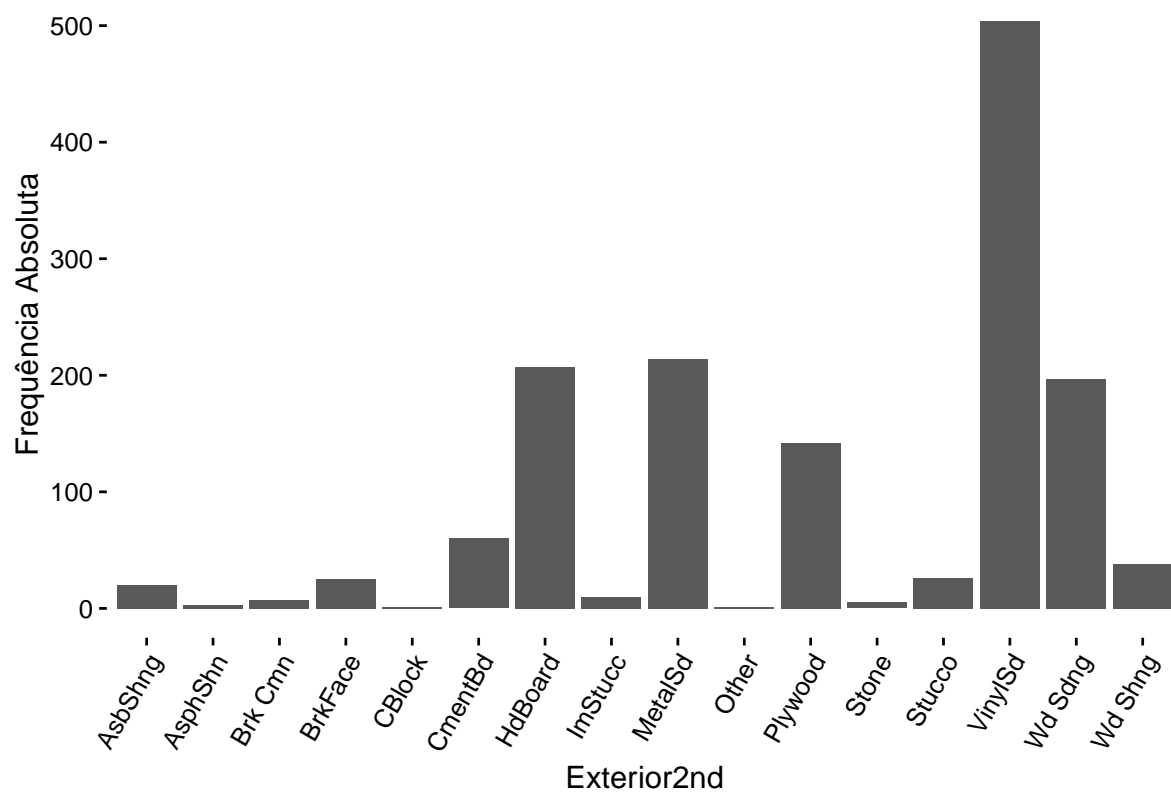
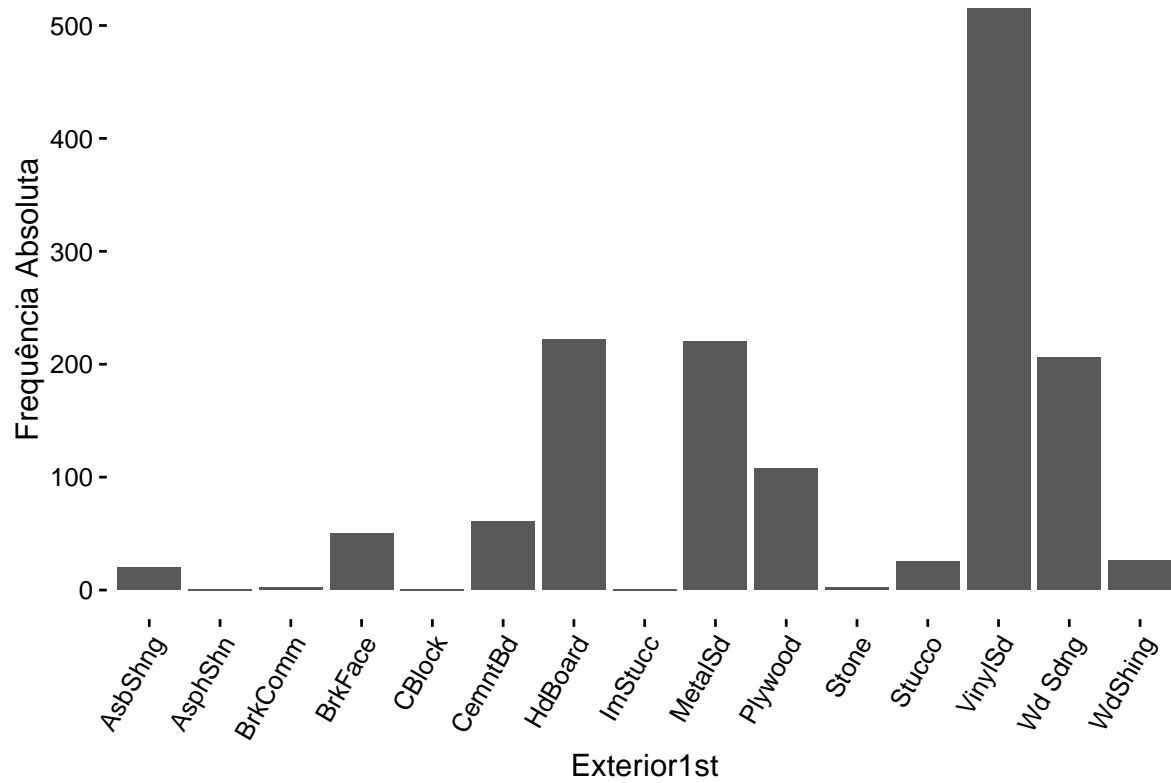


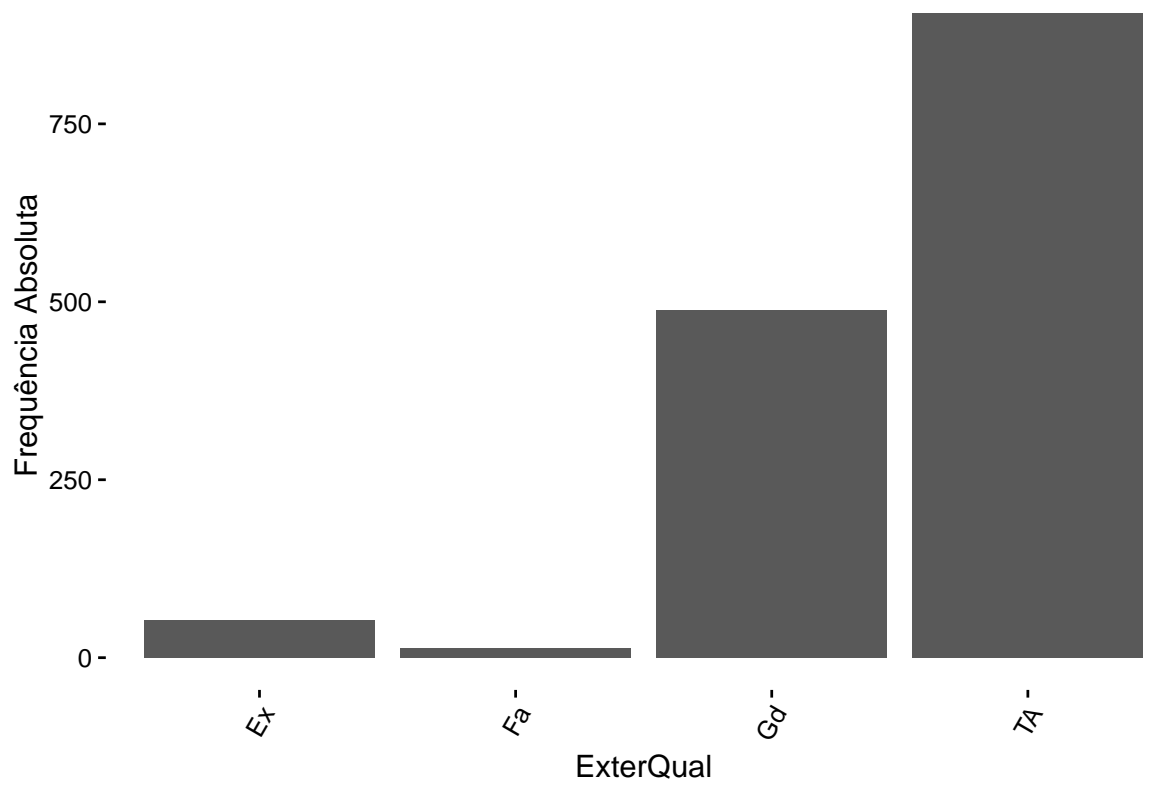
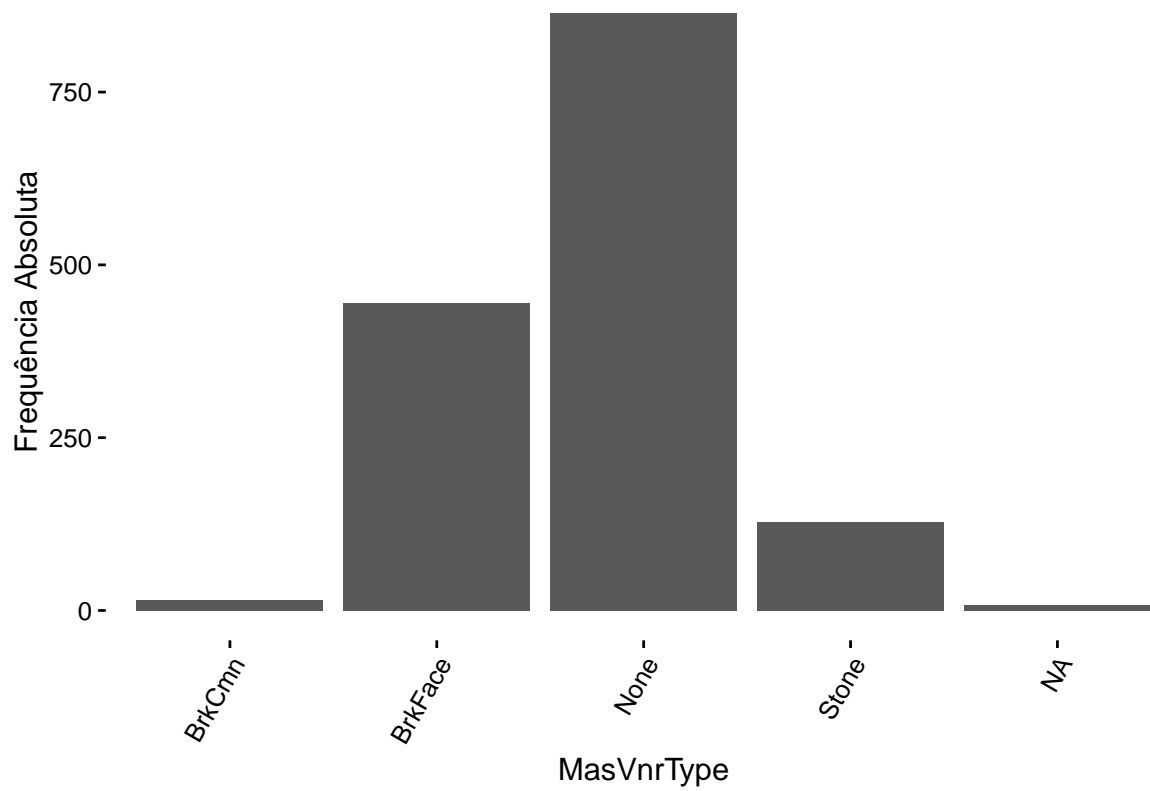


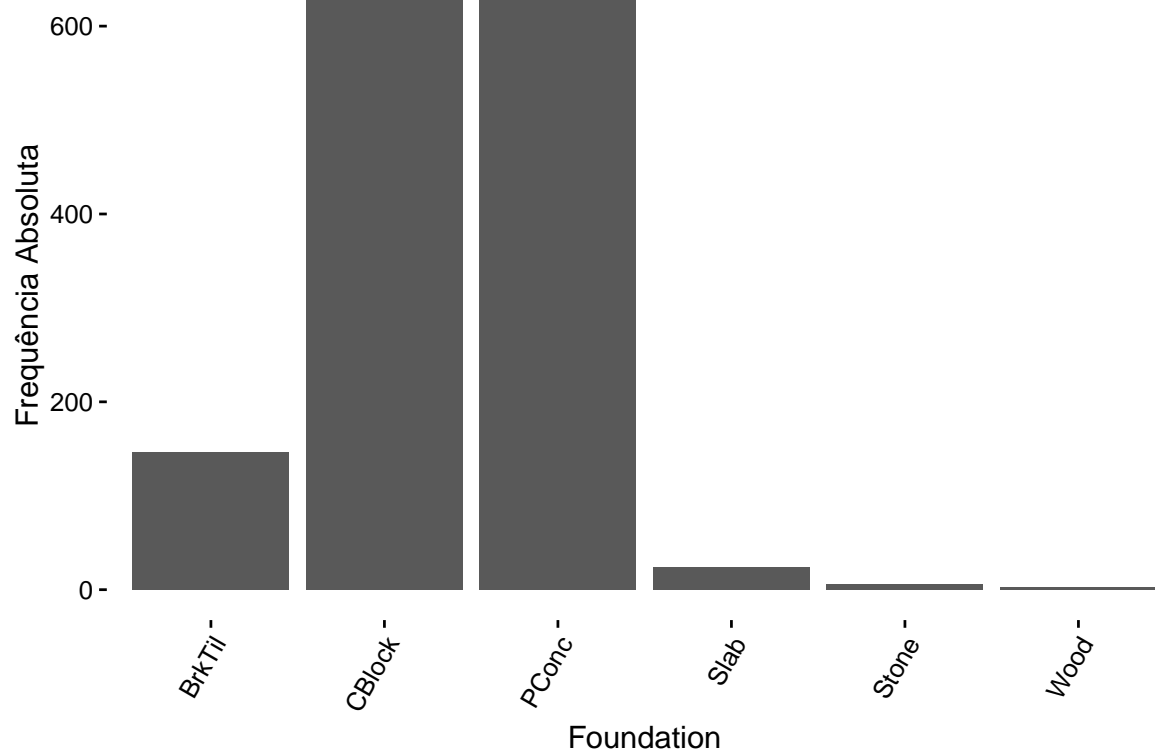
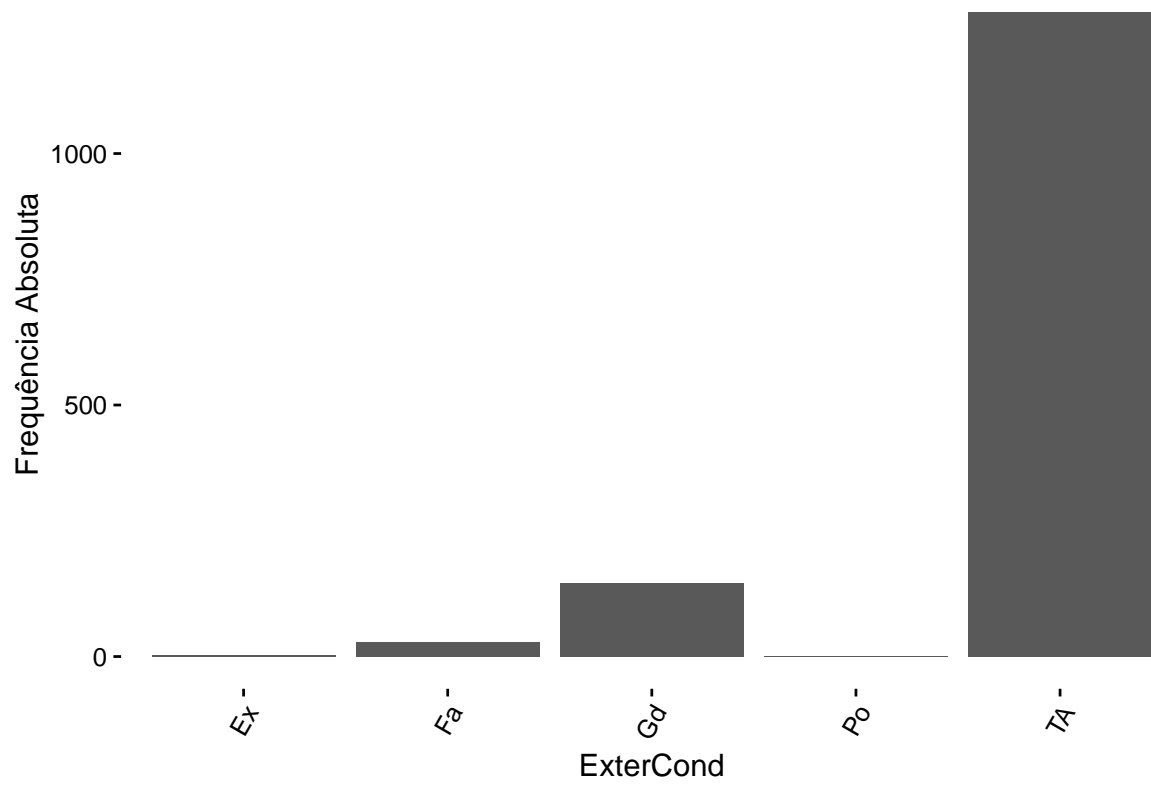


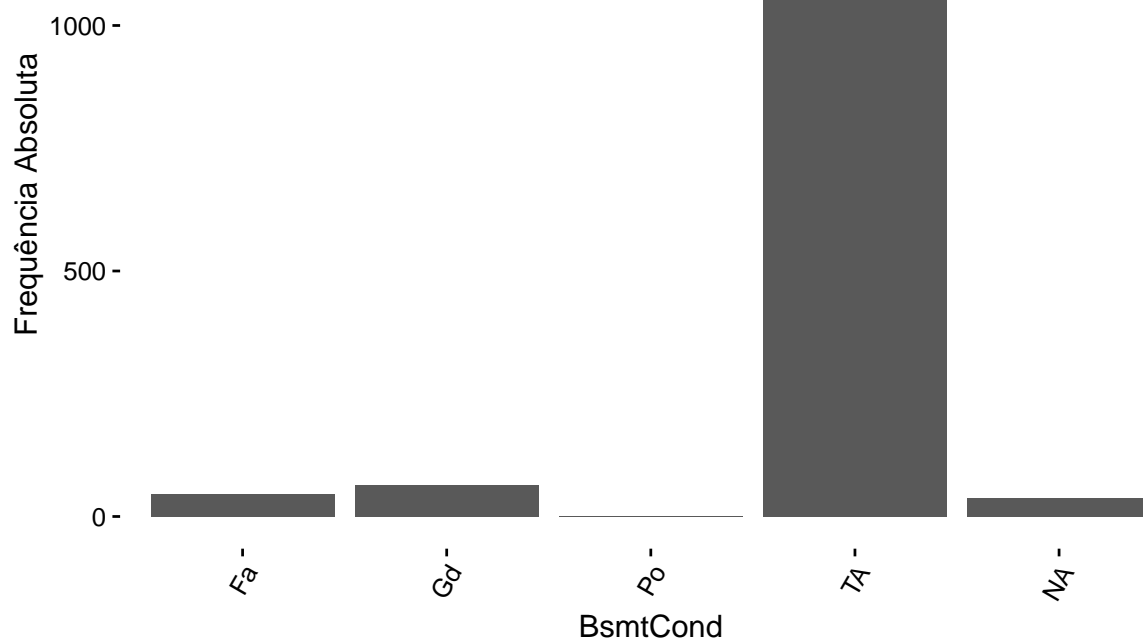
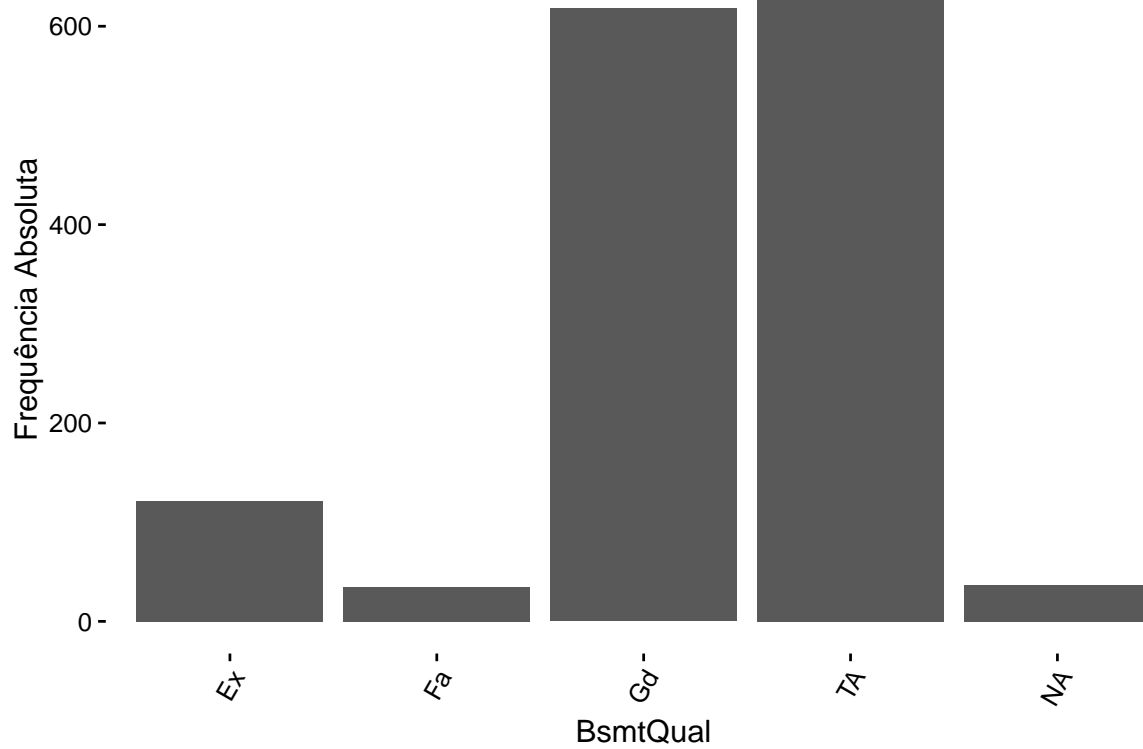


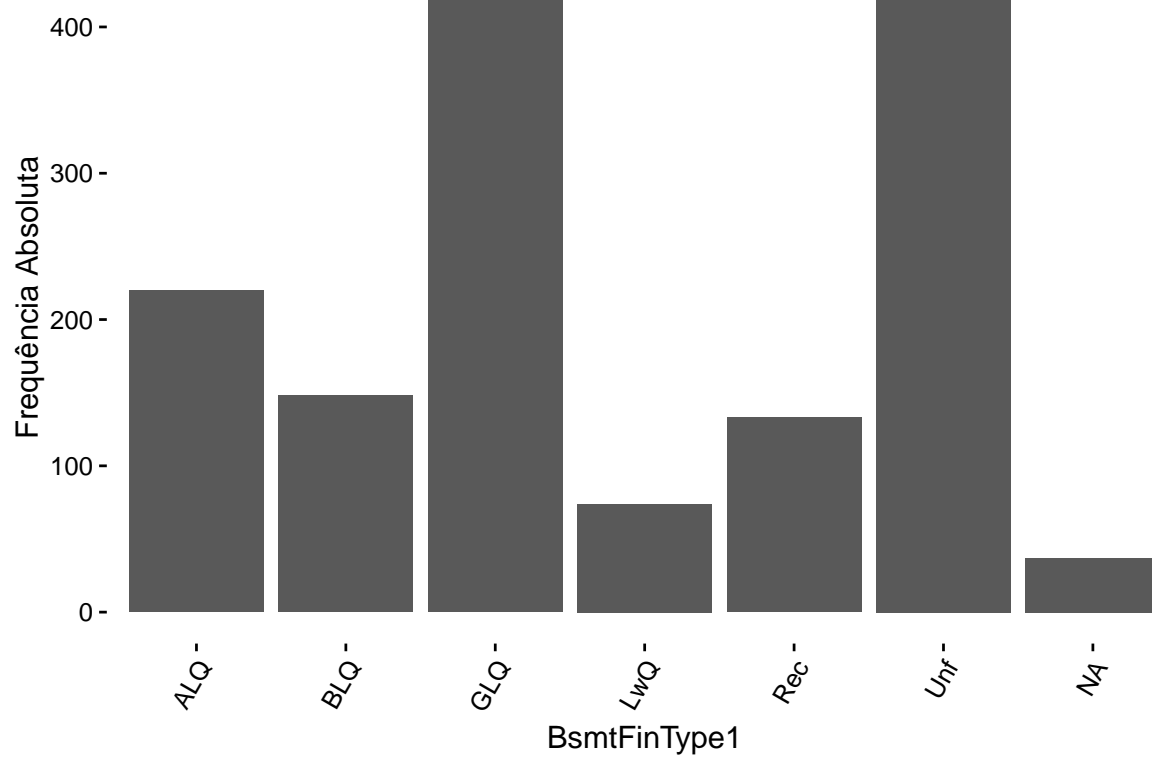
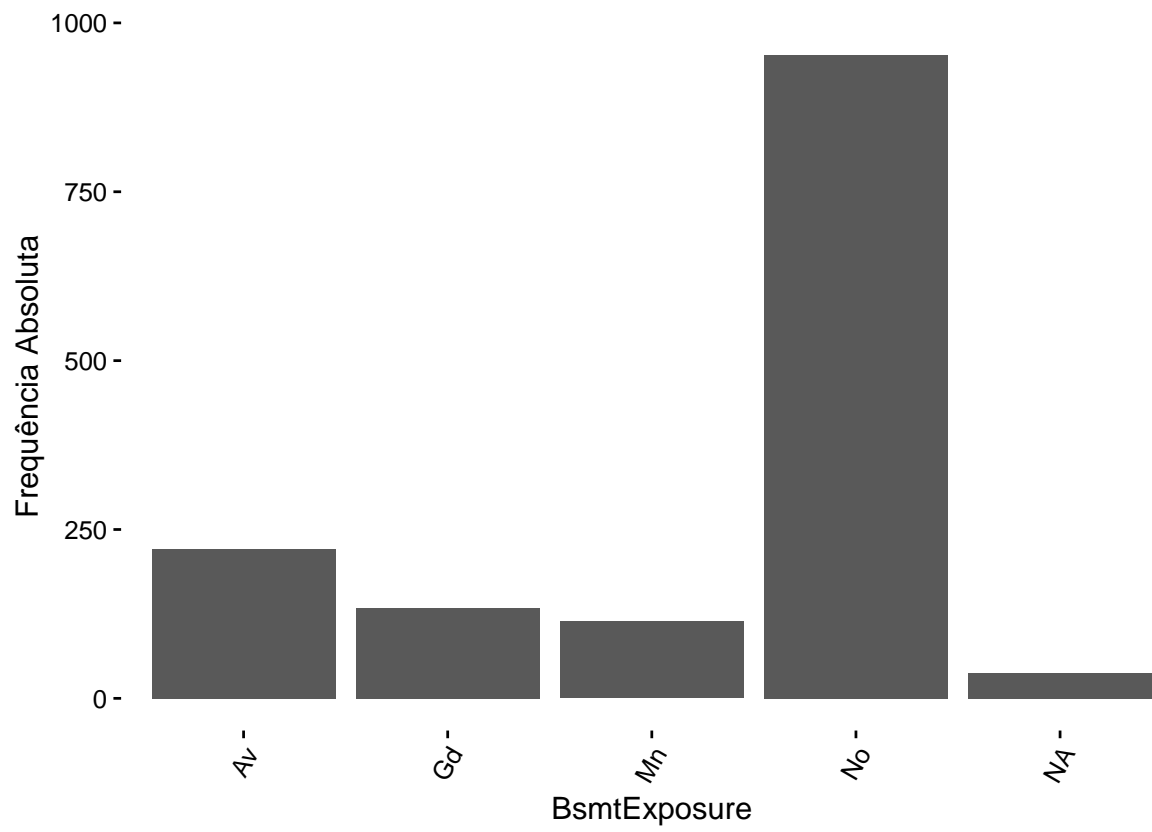


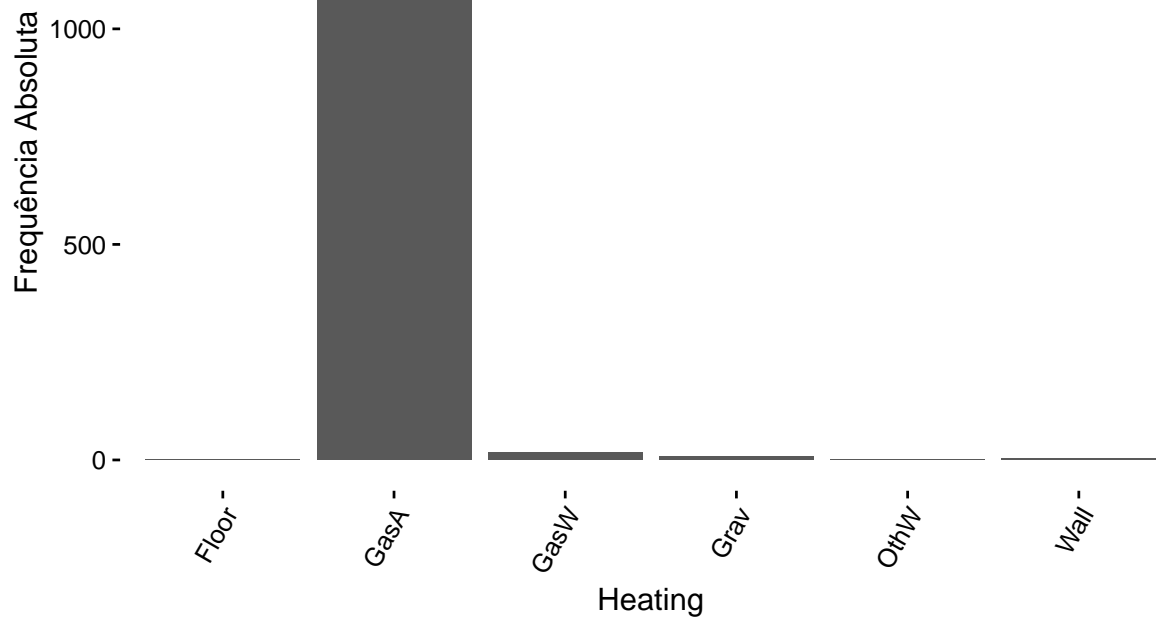
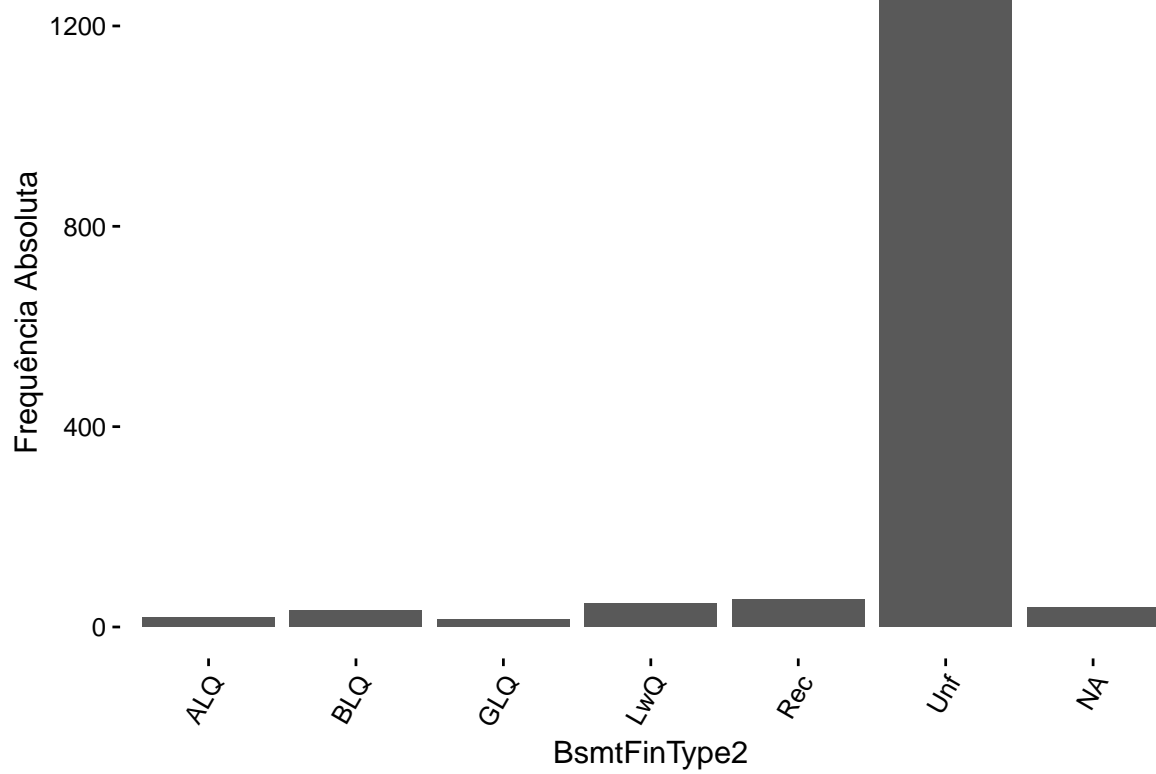


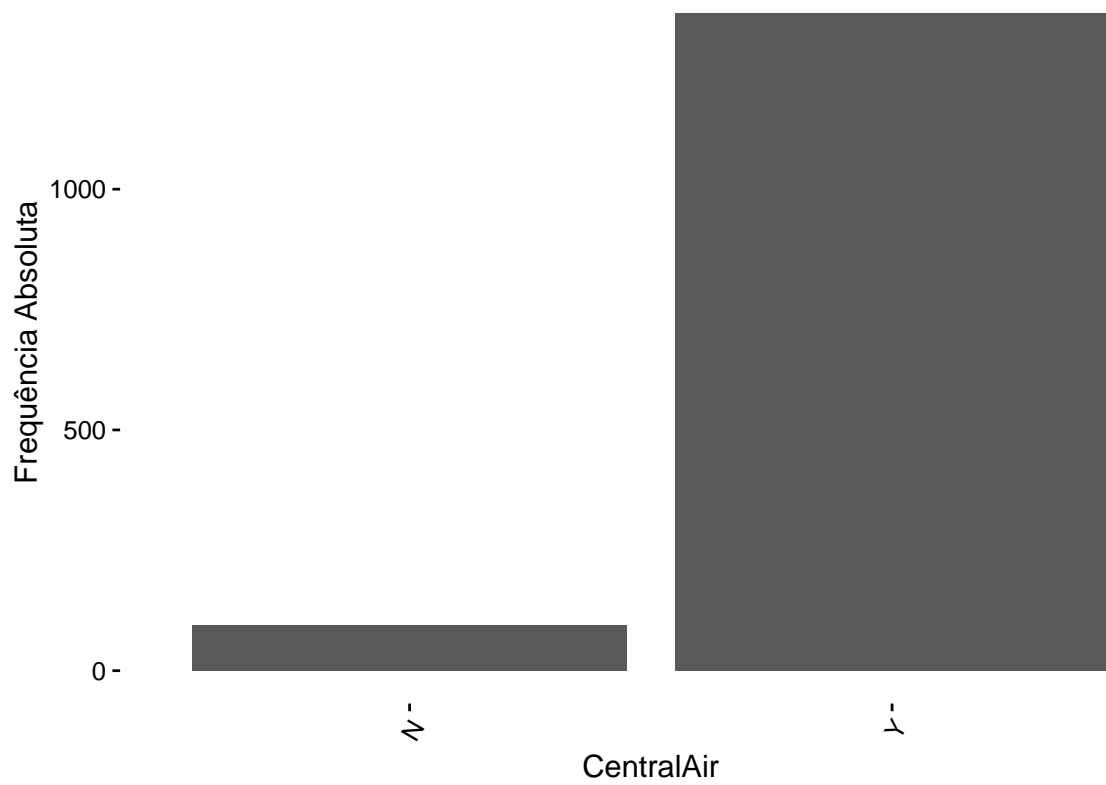
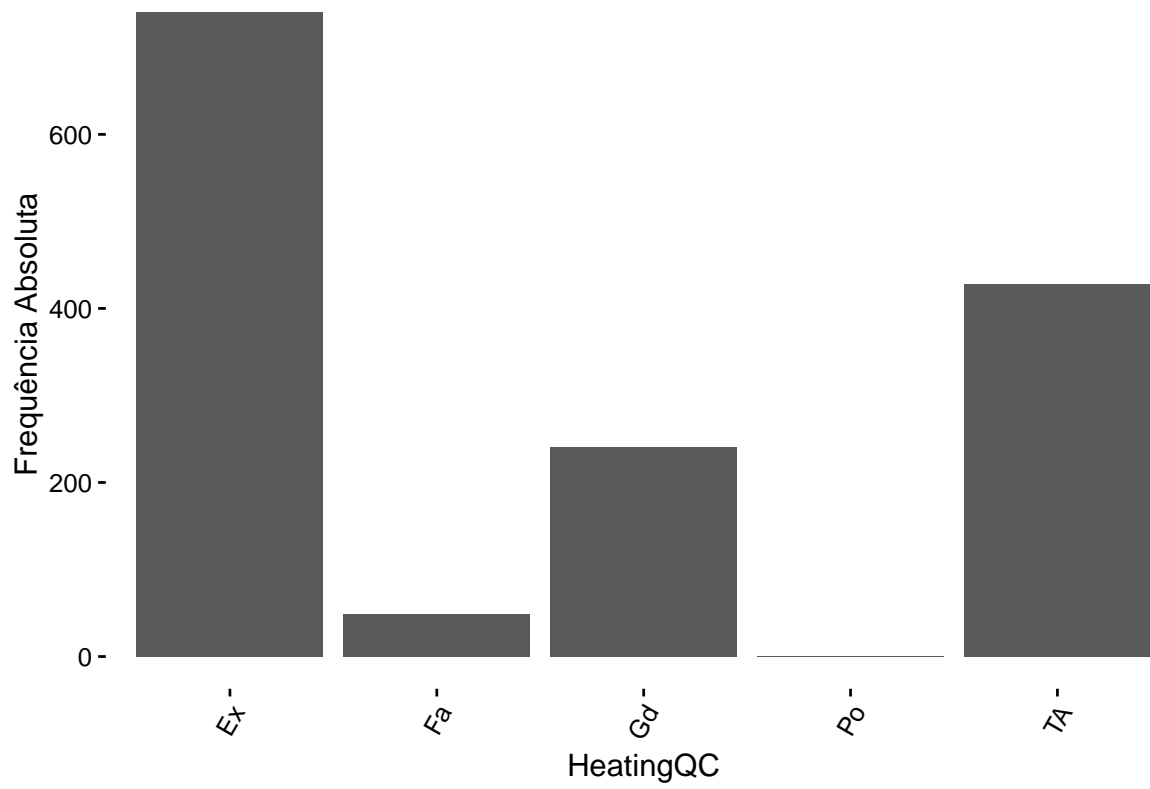


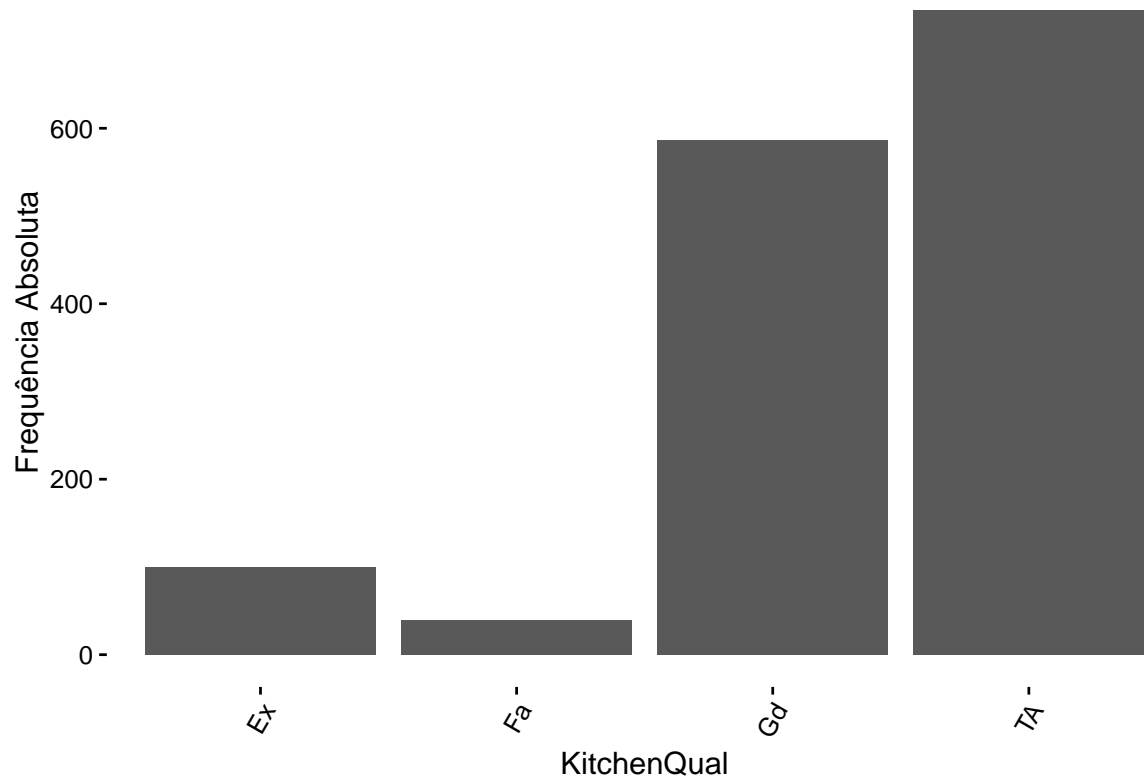
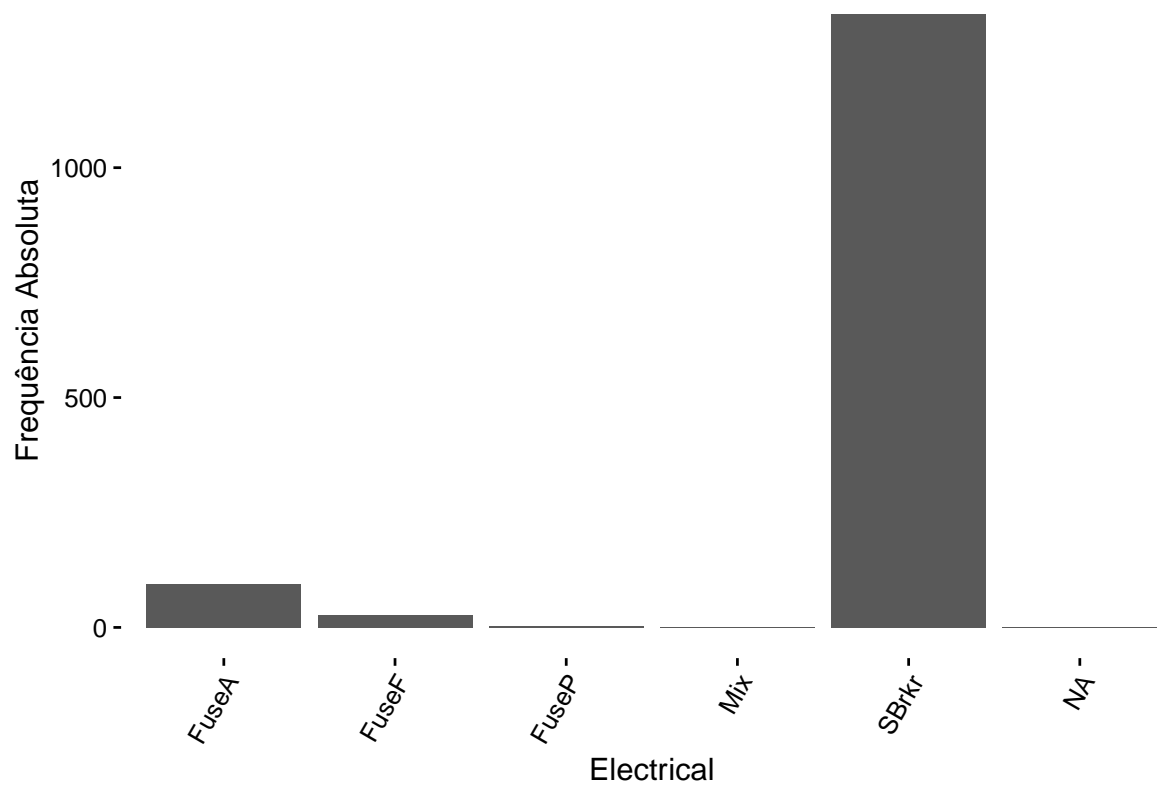


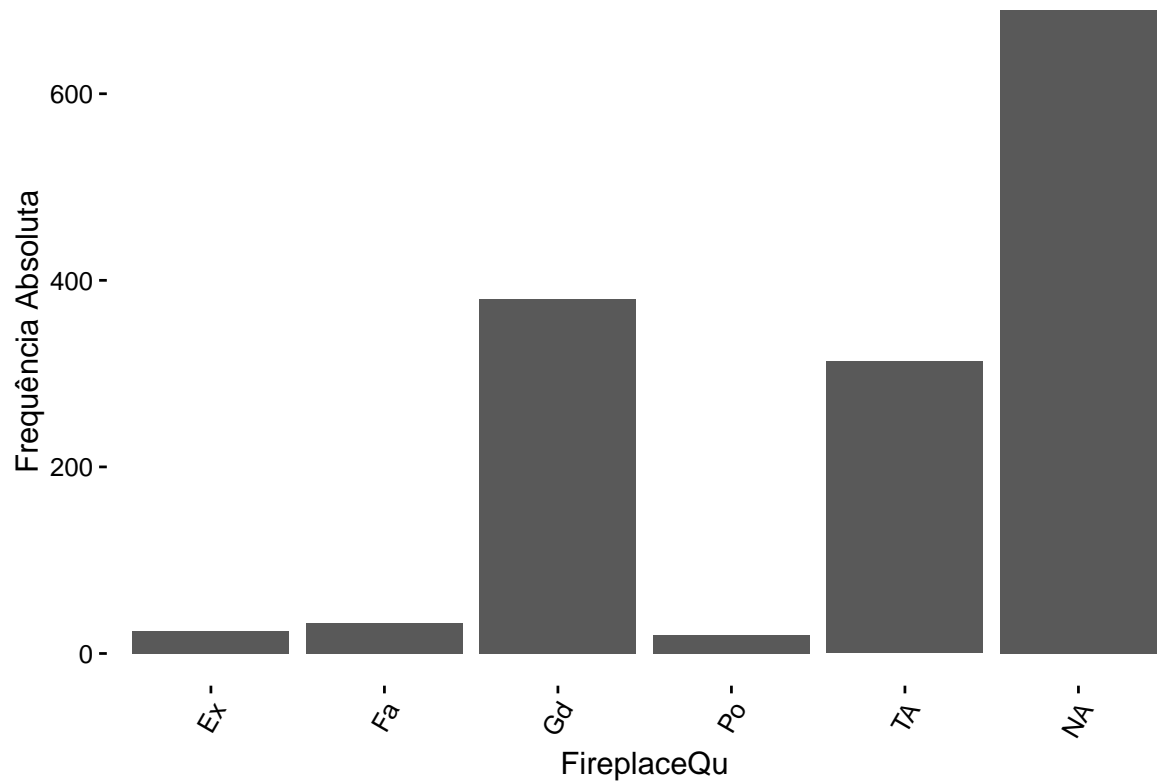
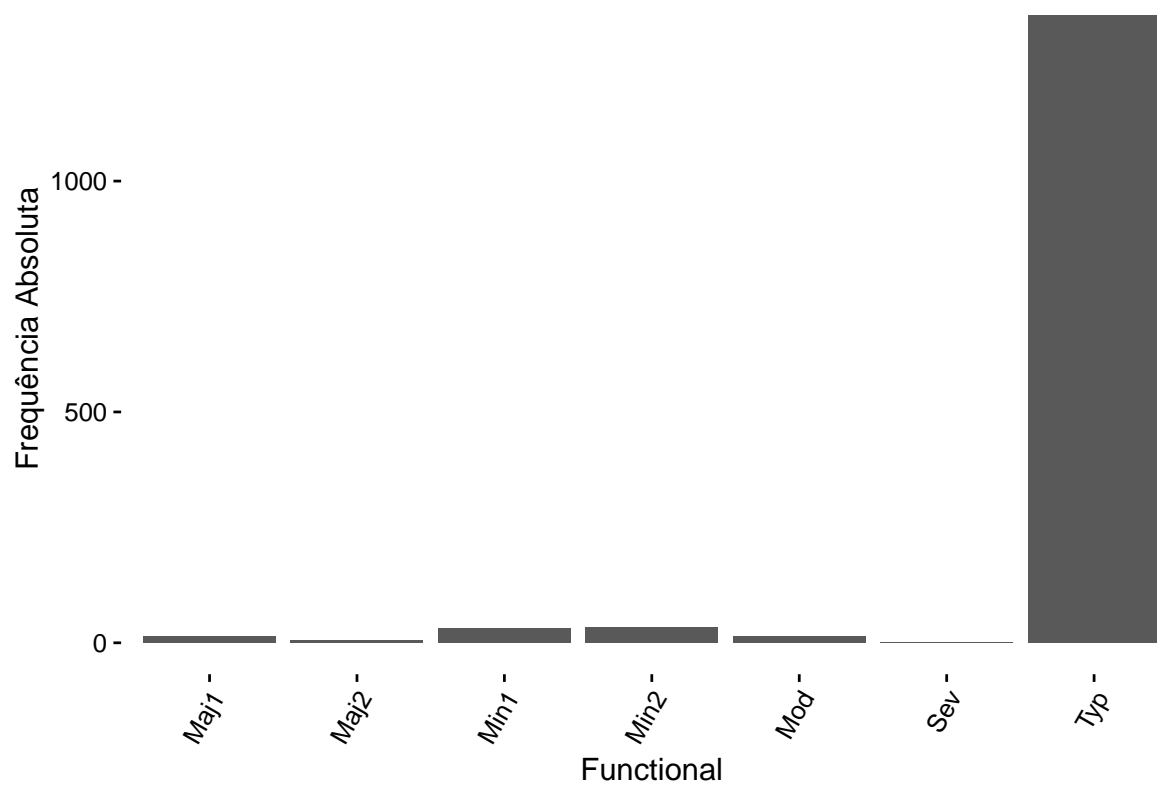


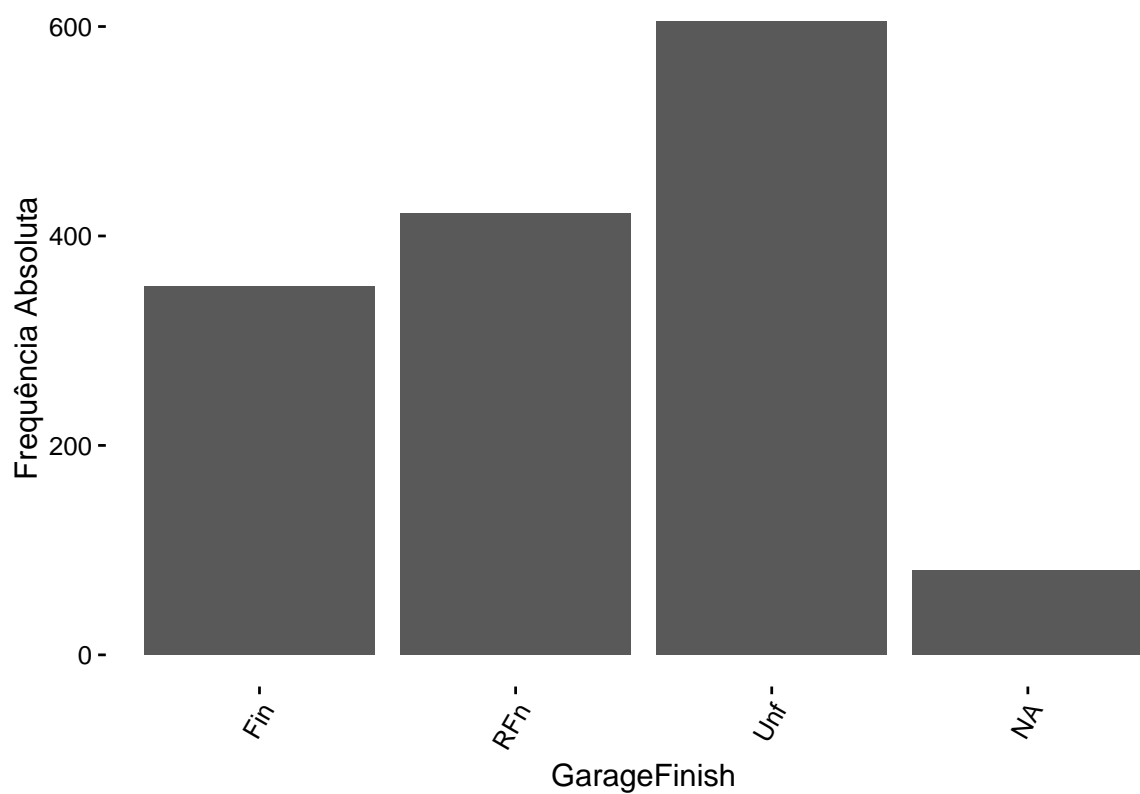
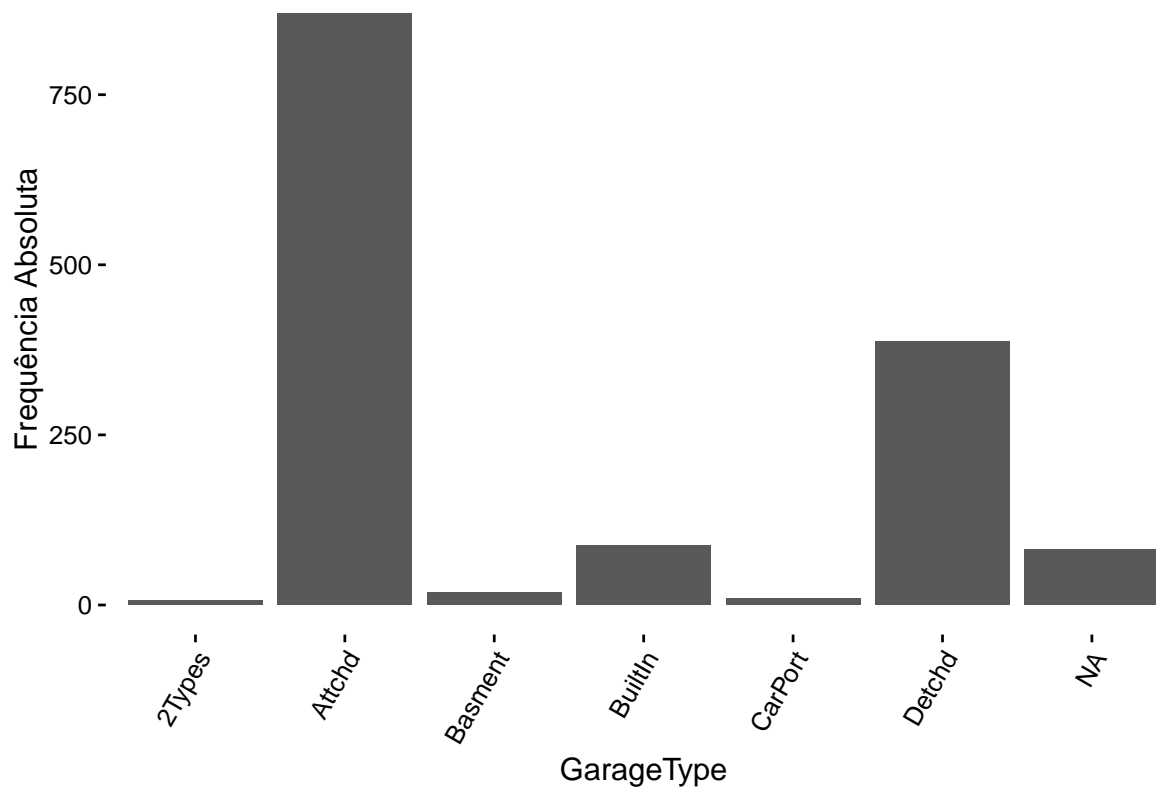


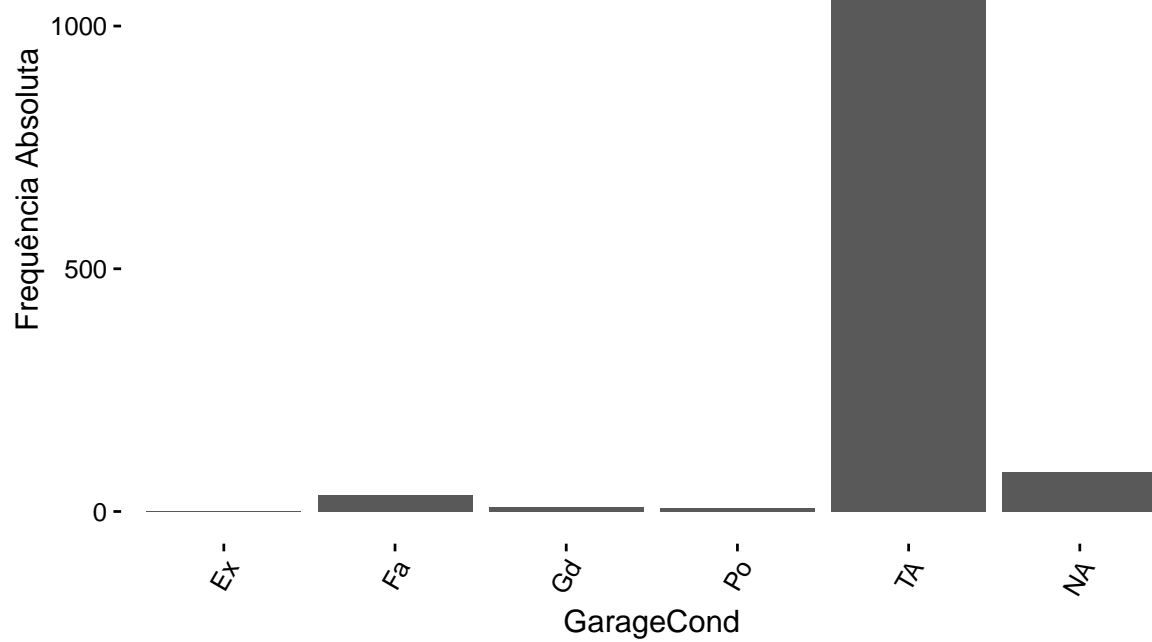
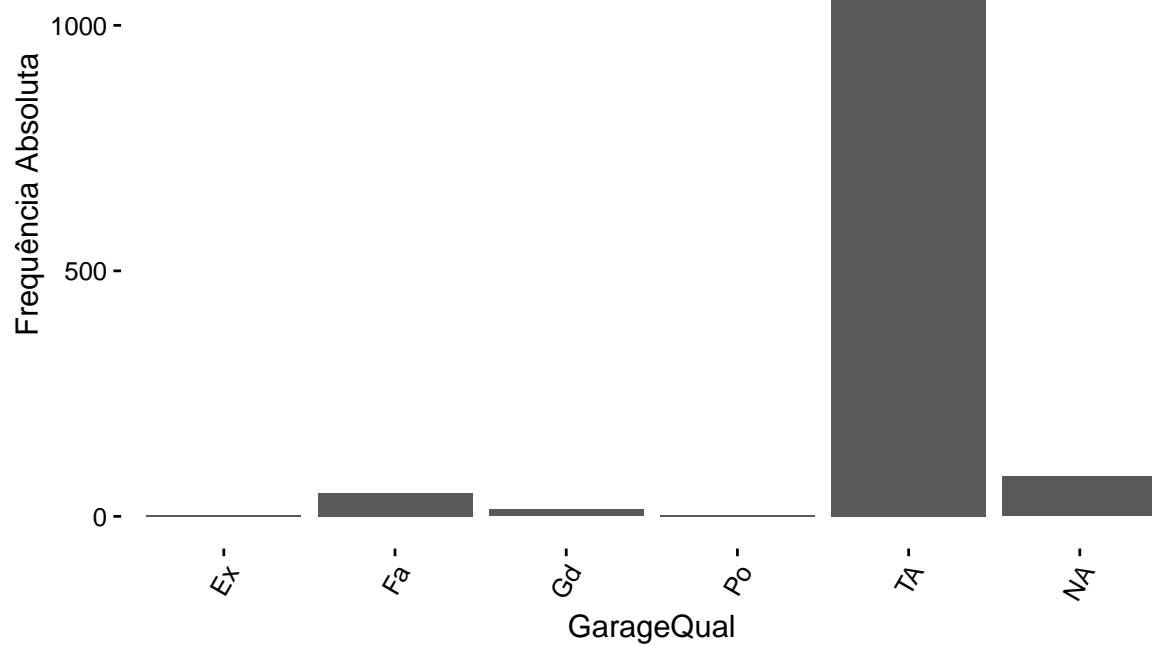


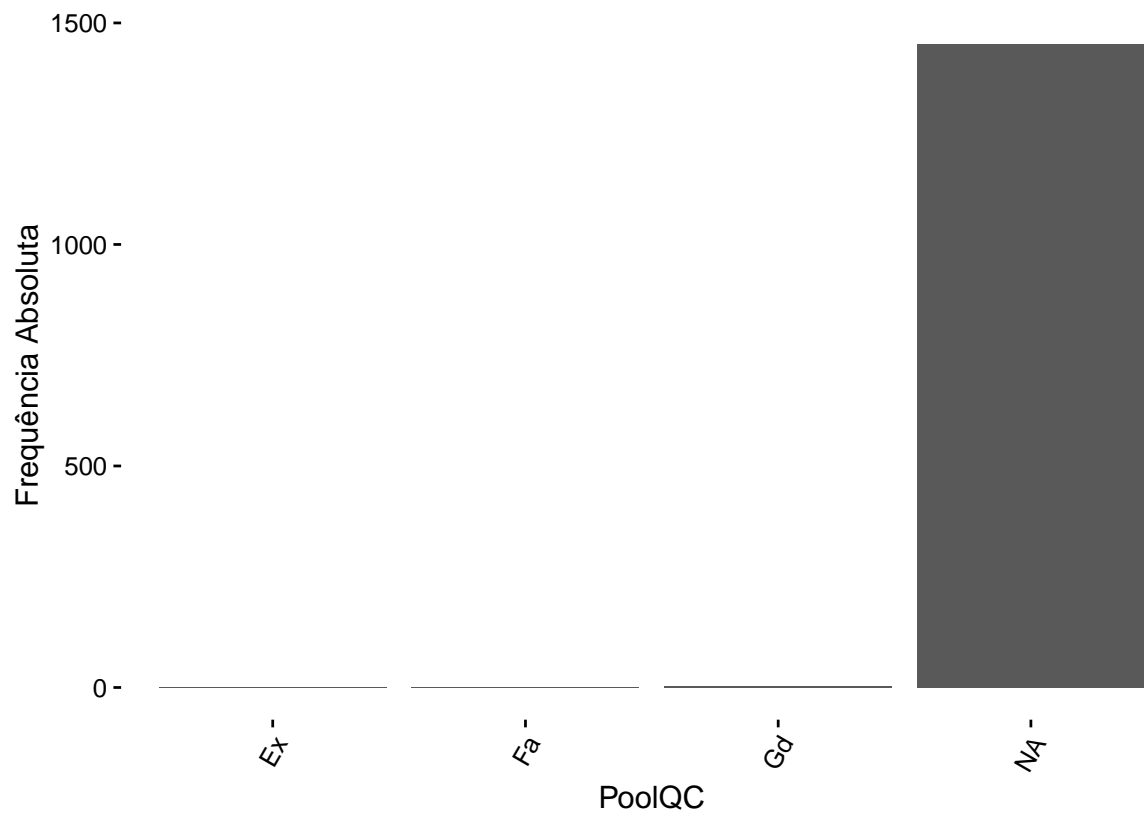
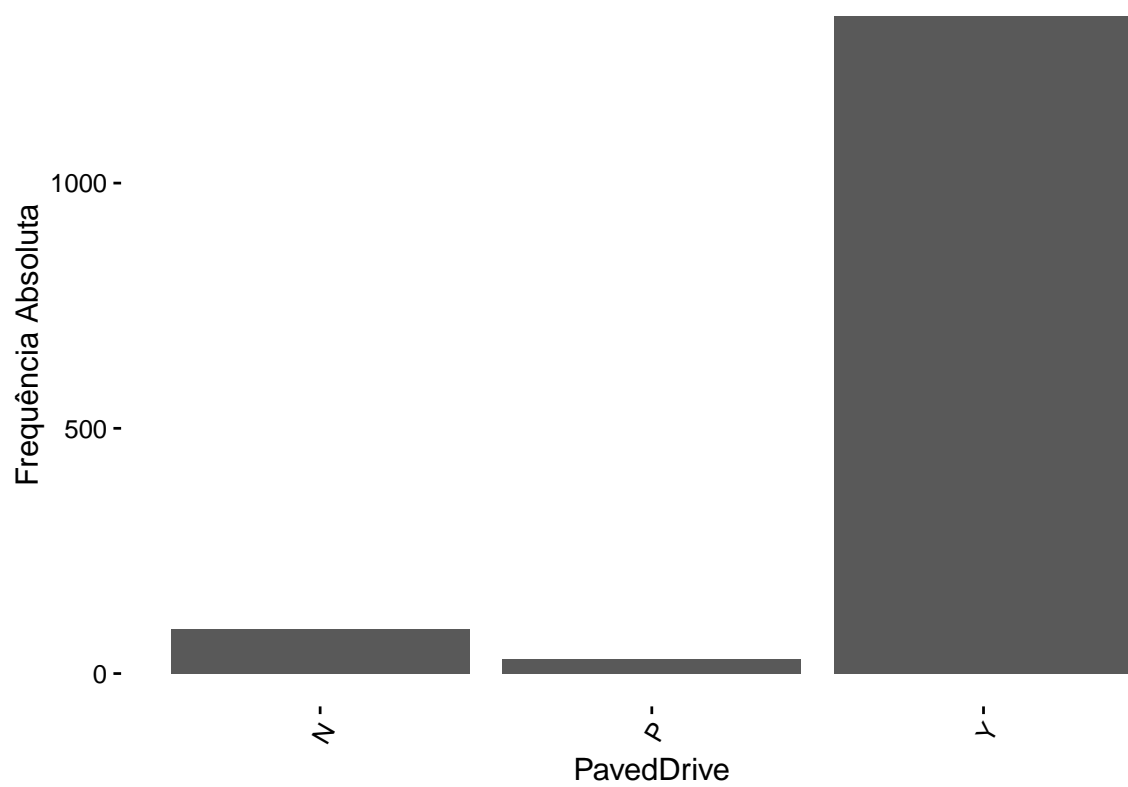


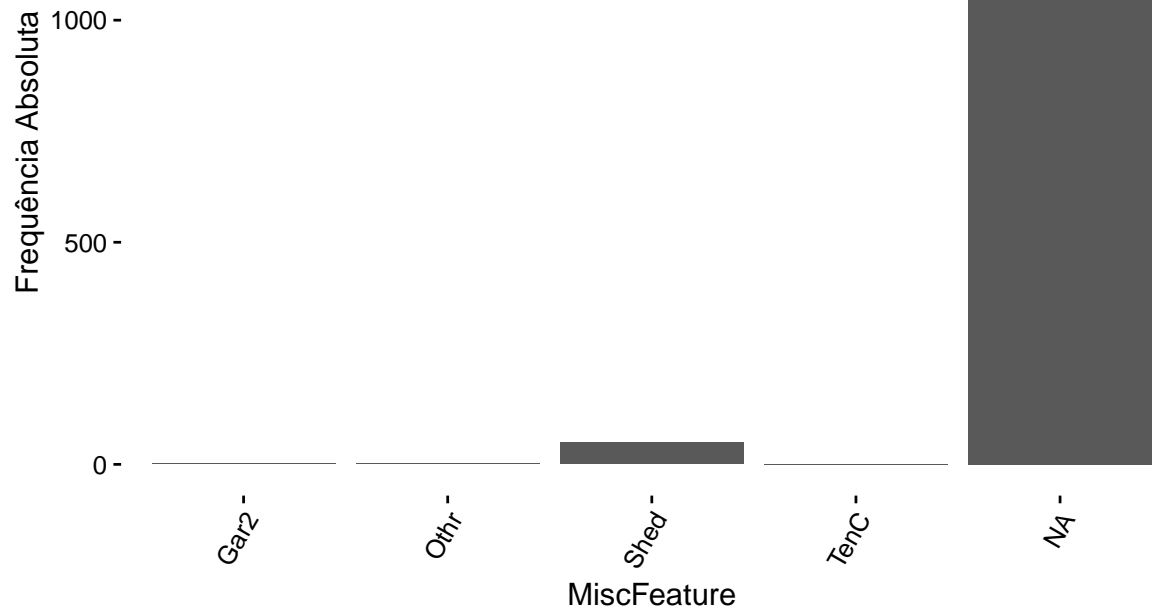
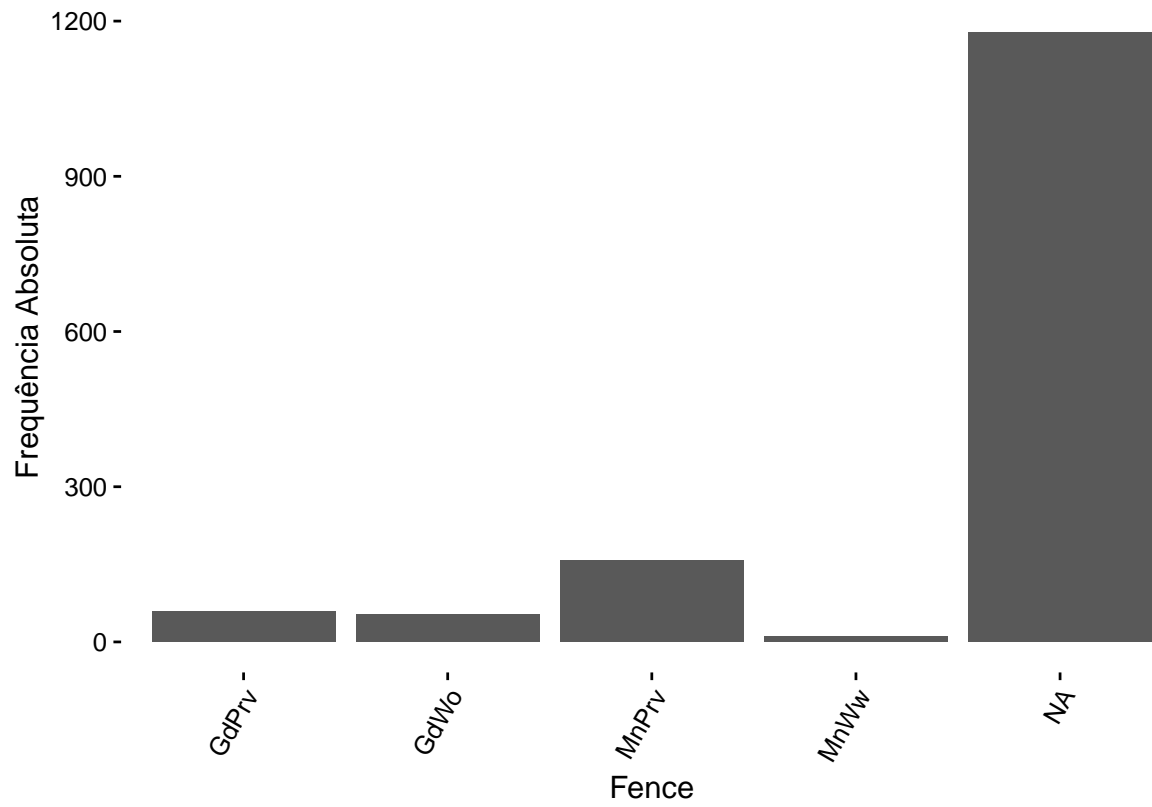


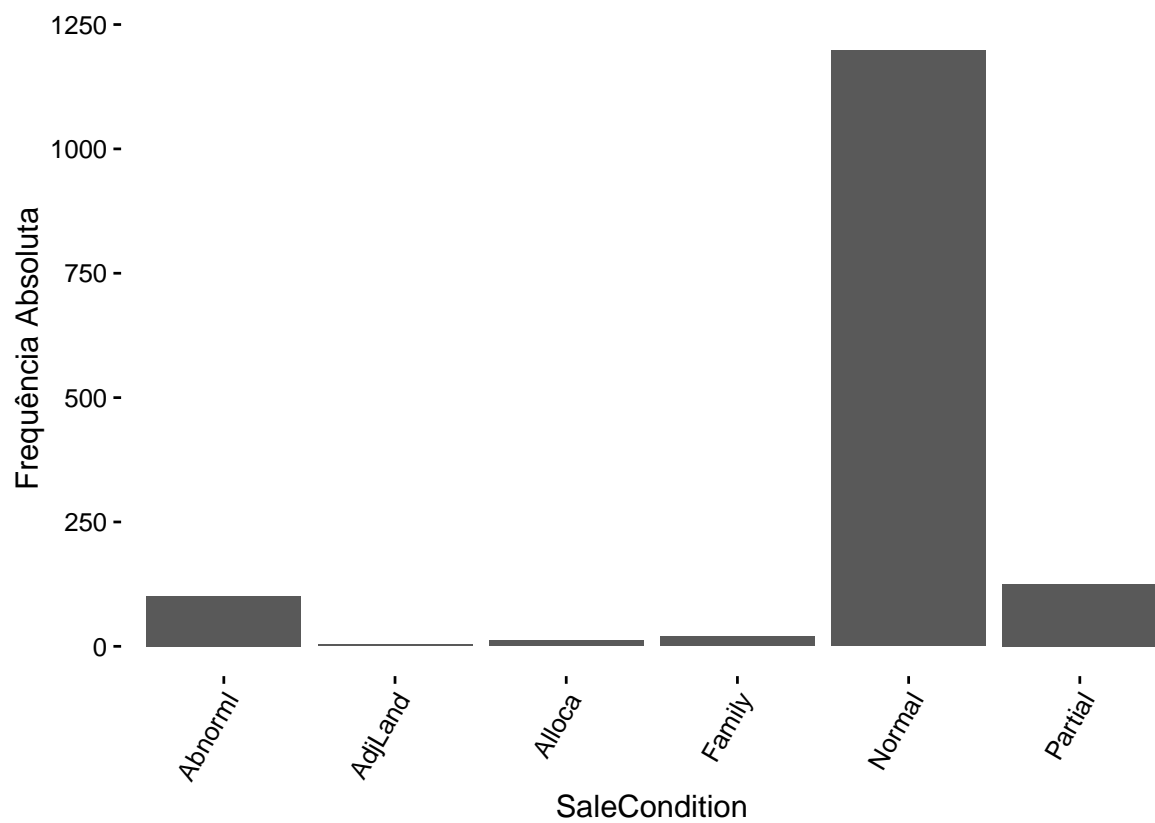
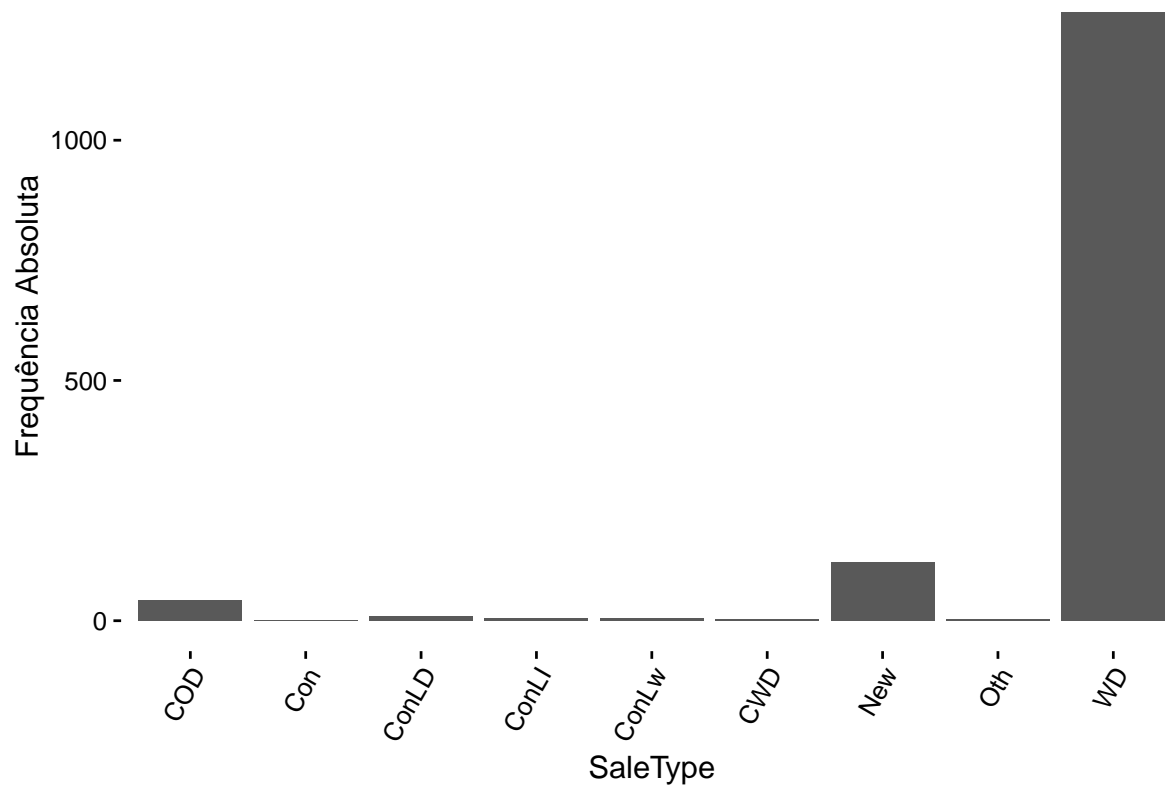






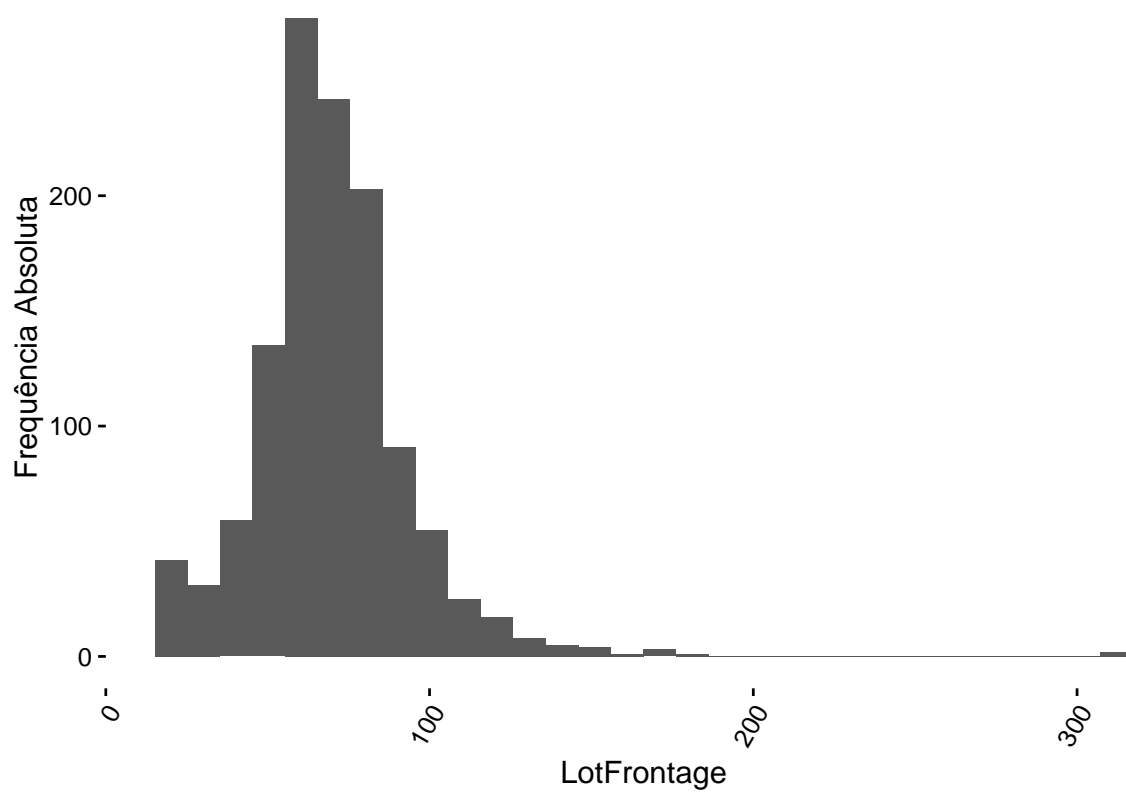




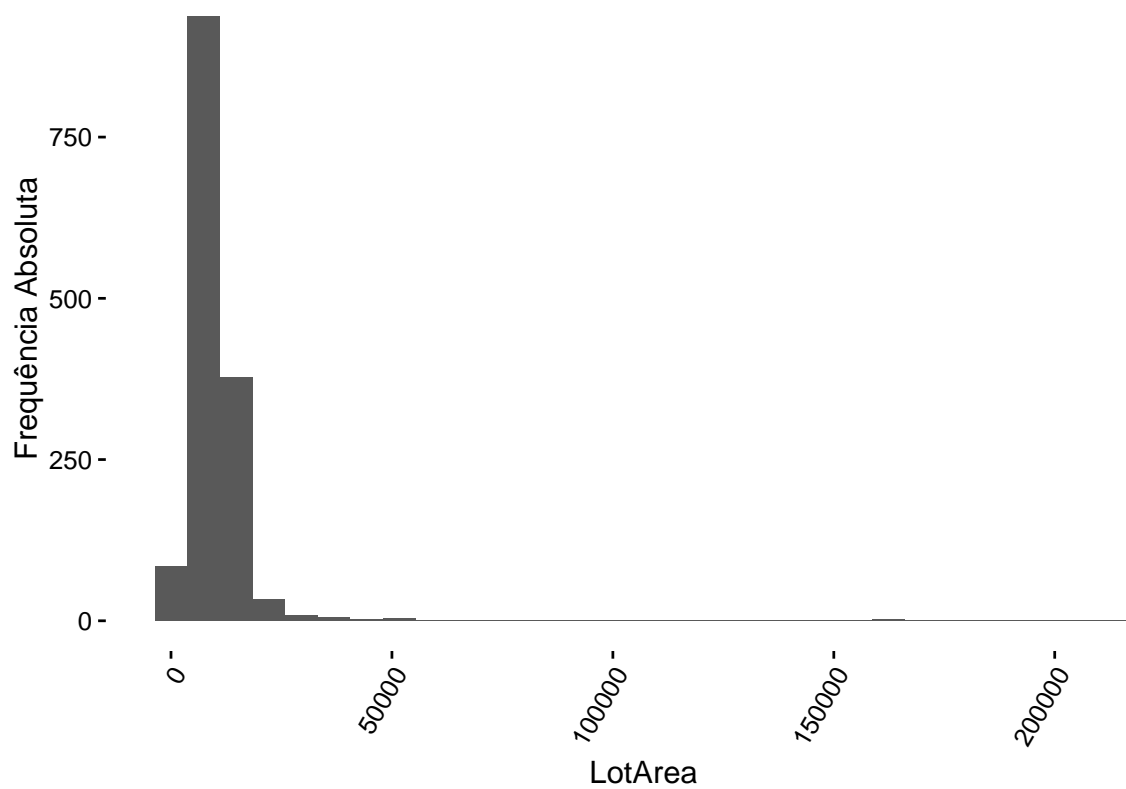


1.2 GRÁFICOS DAS VARIÁVEIS QUANTITATIVAS

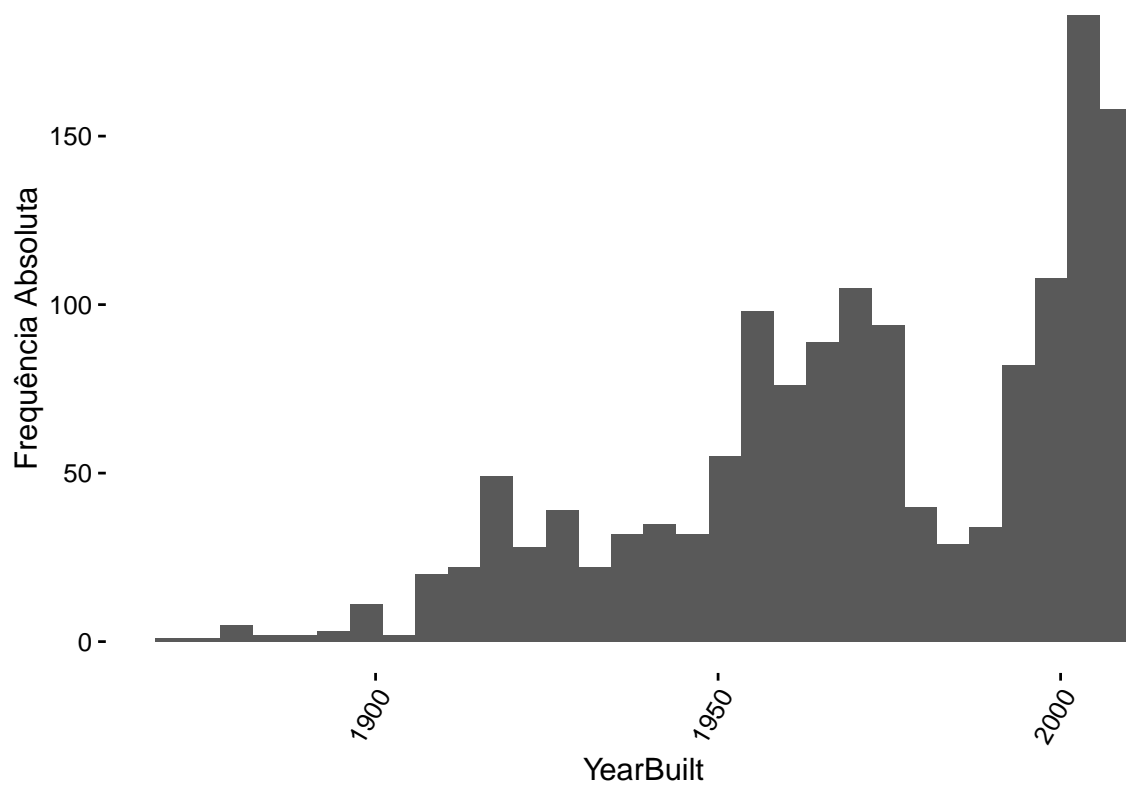
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



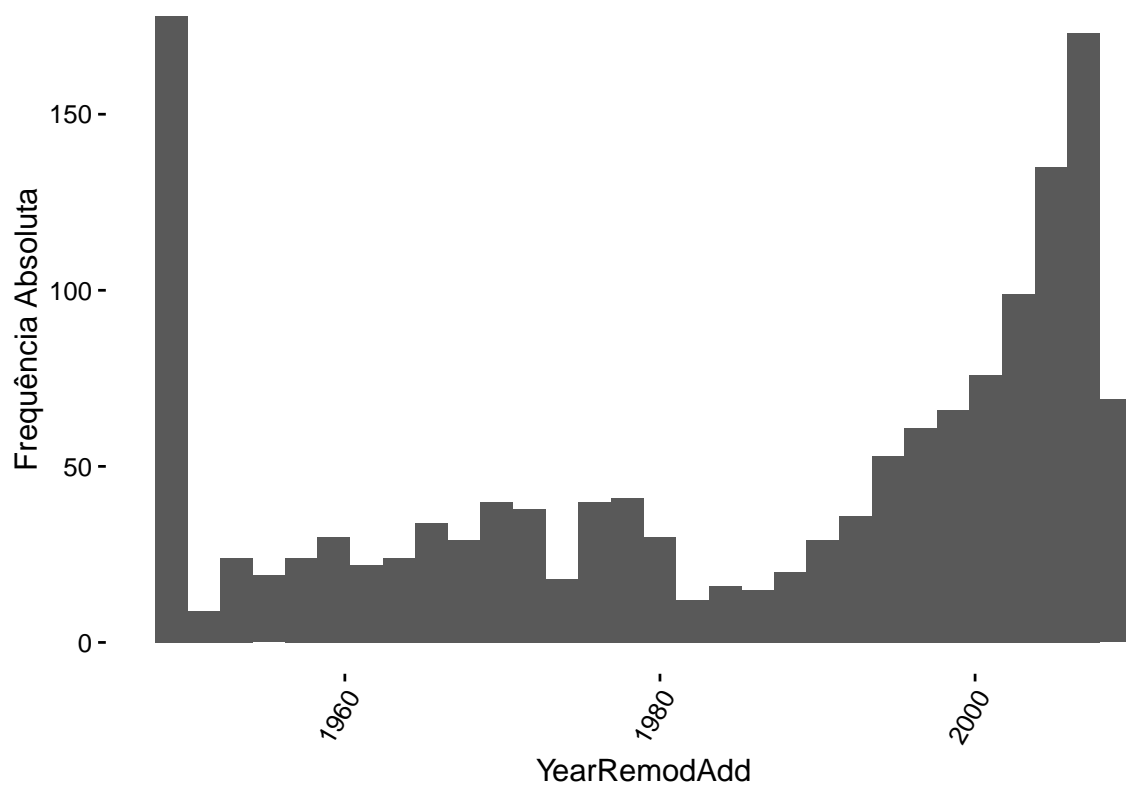
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



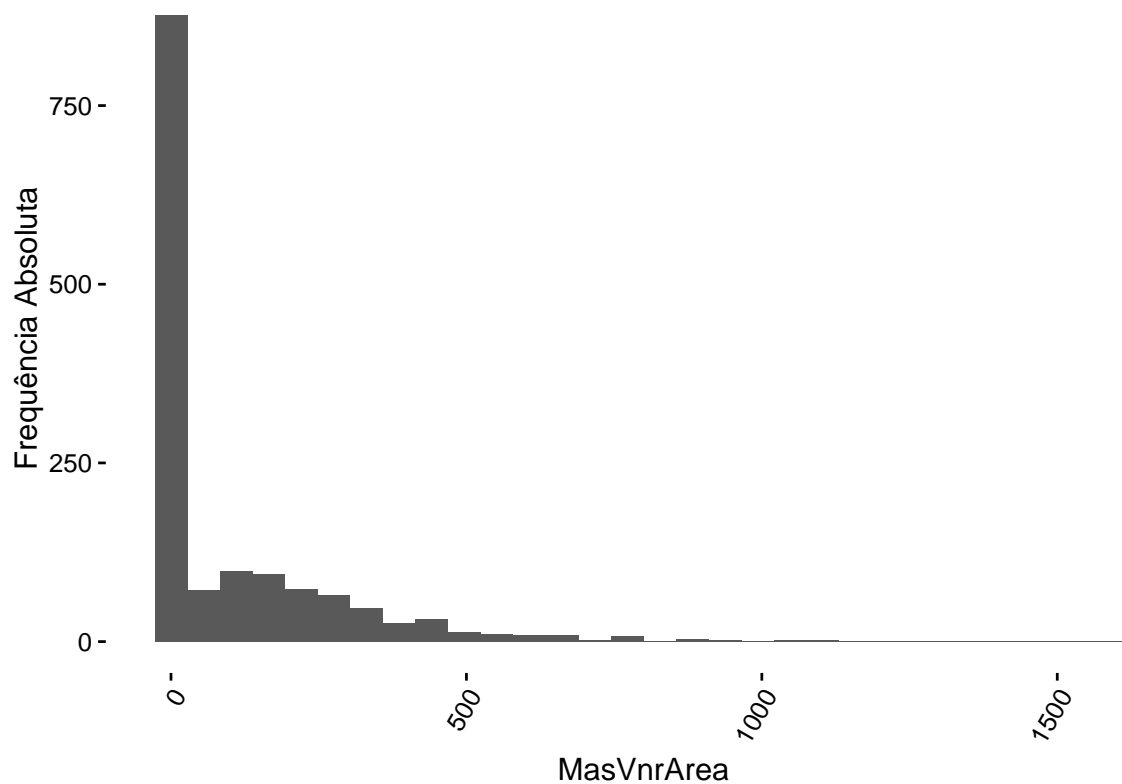
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



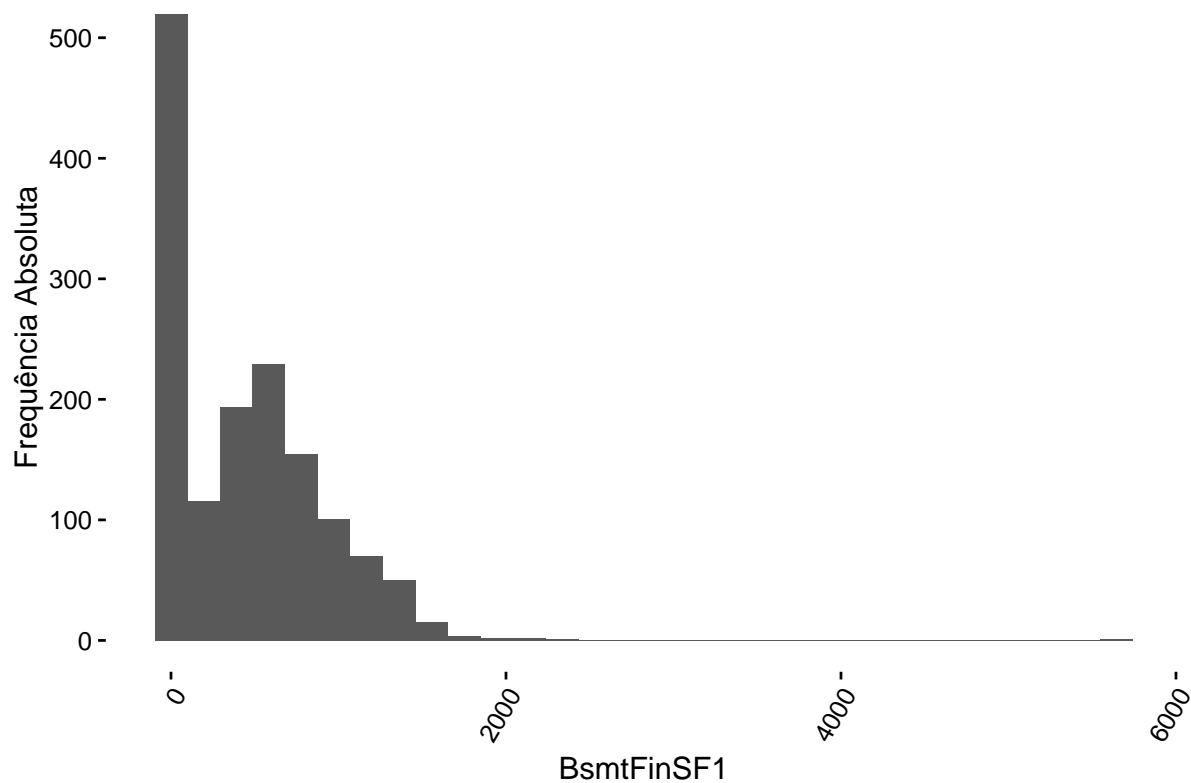
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



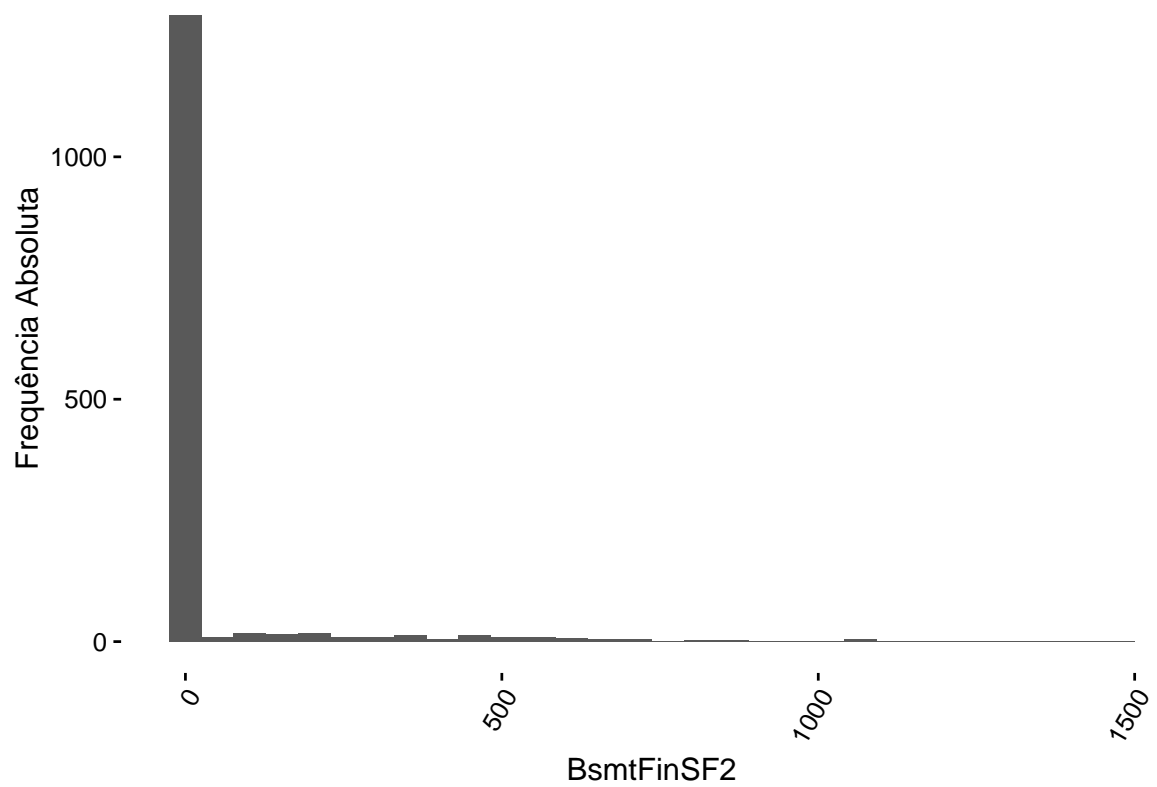
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



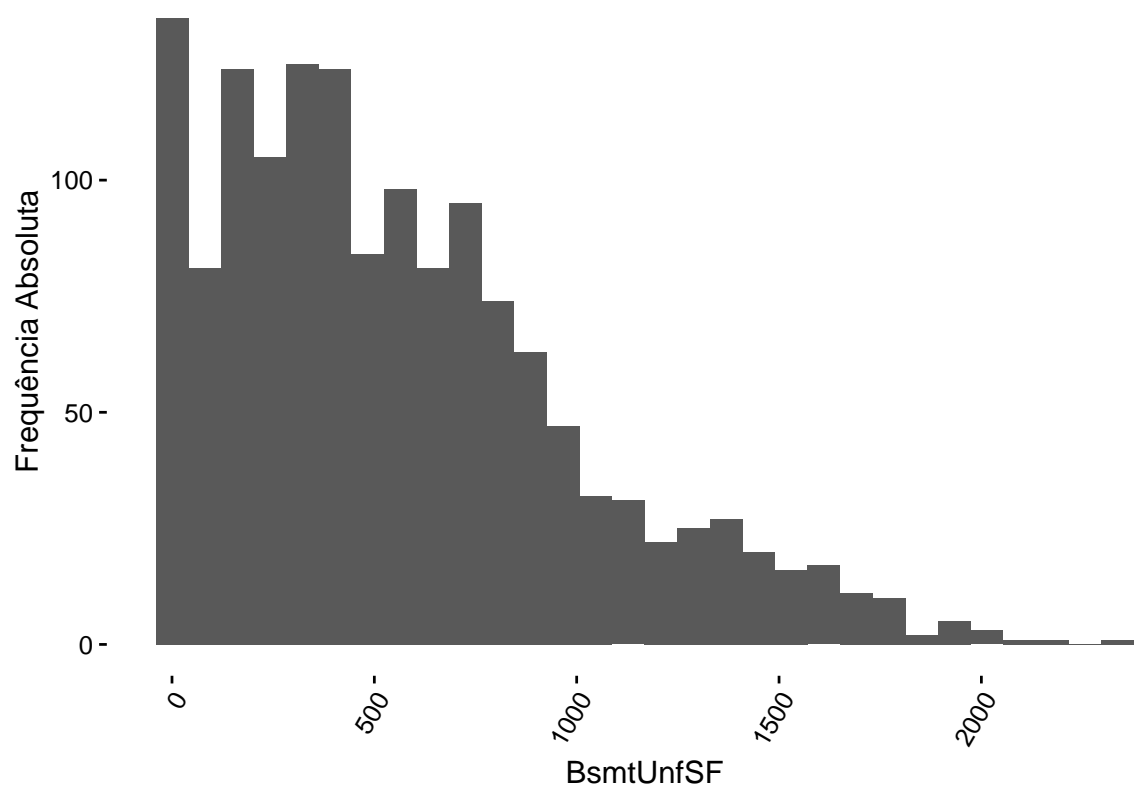
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



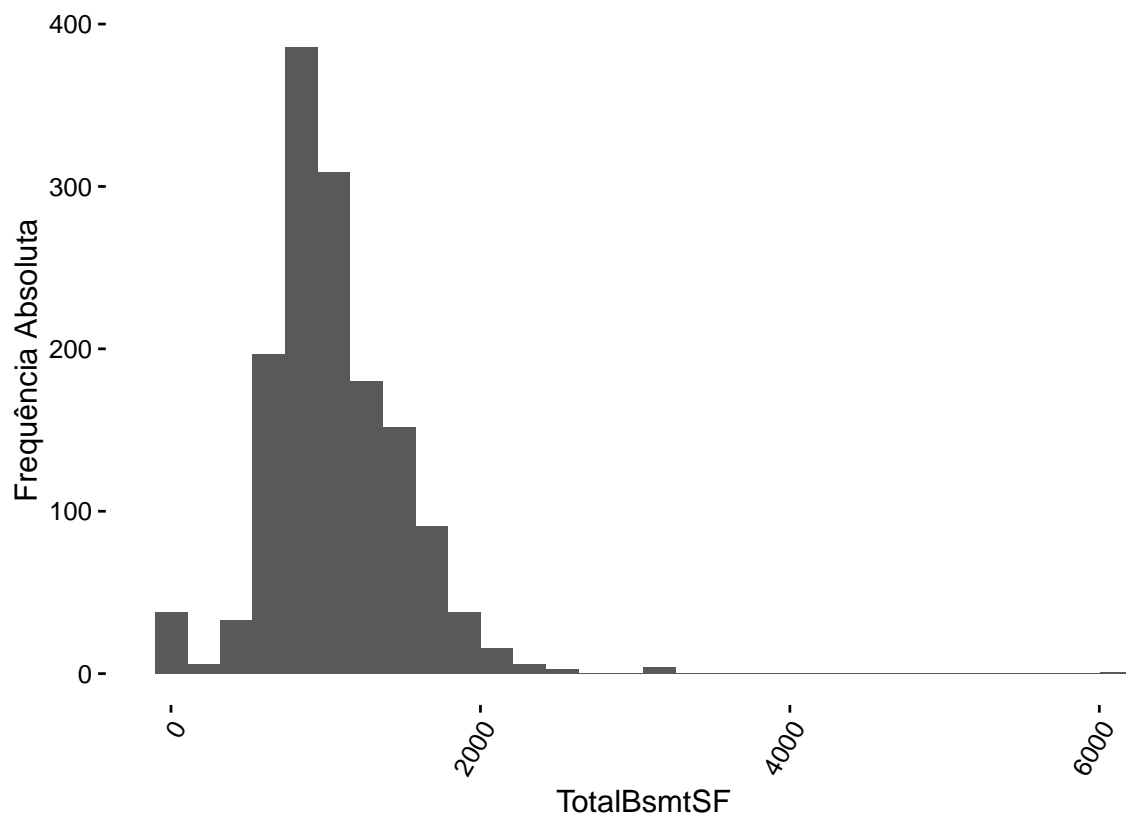
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



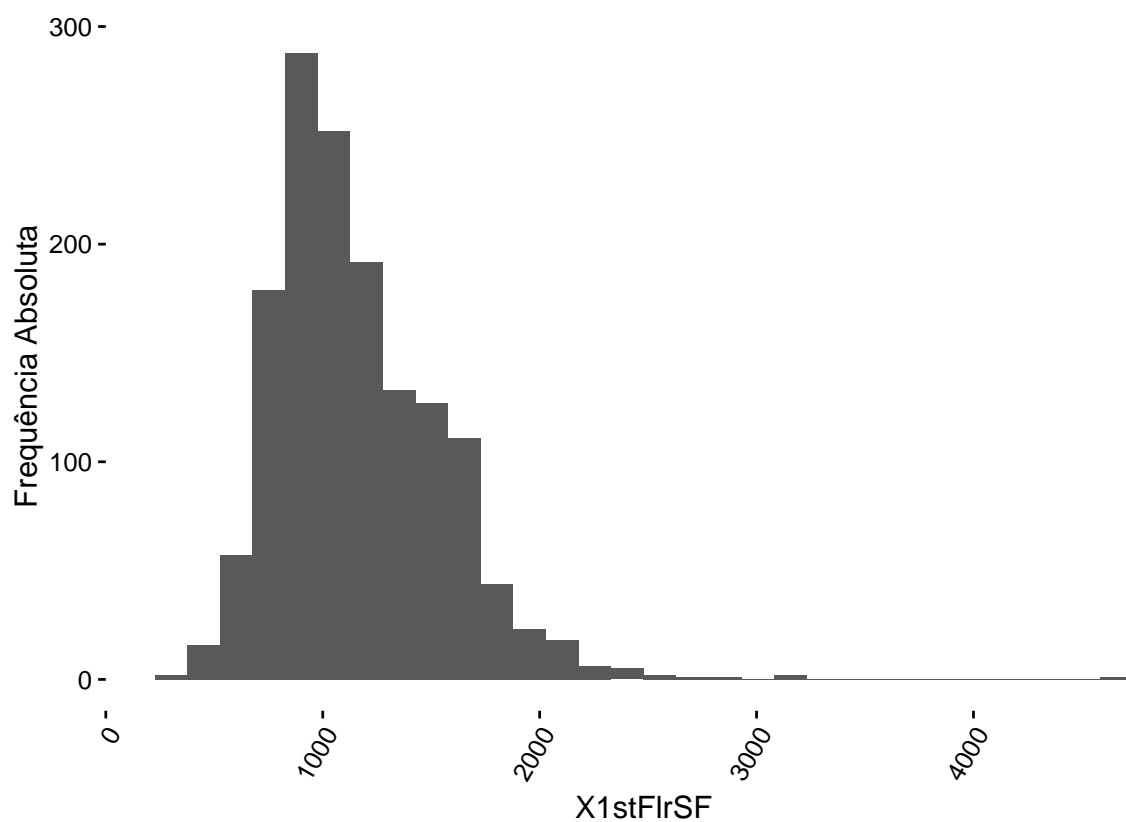
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



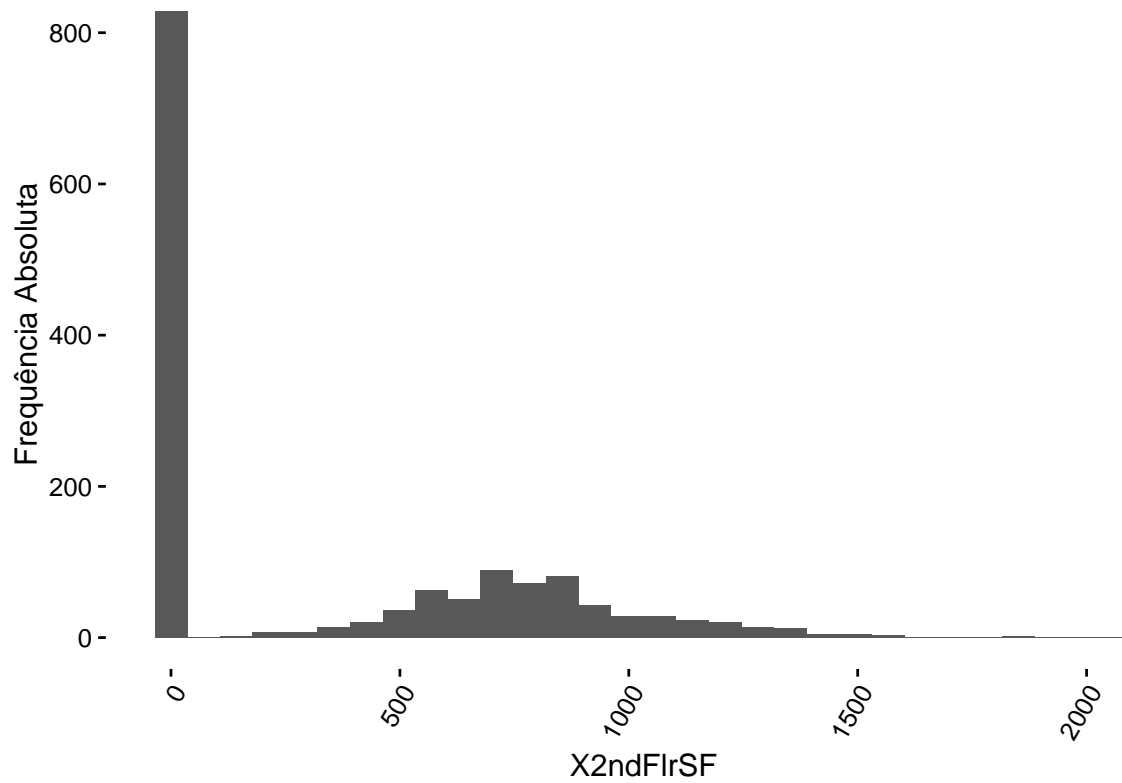
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



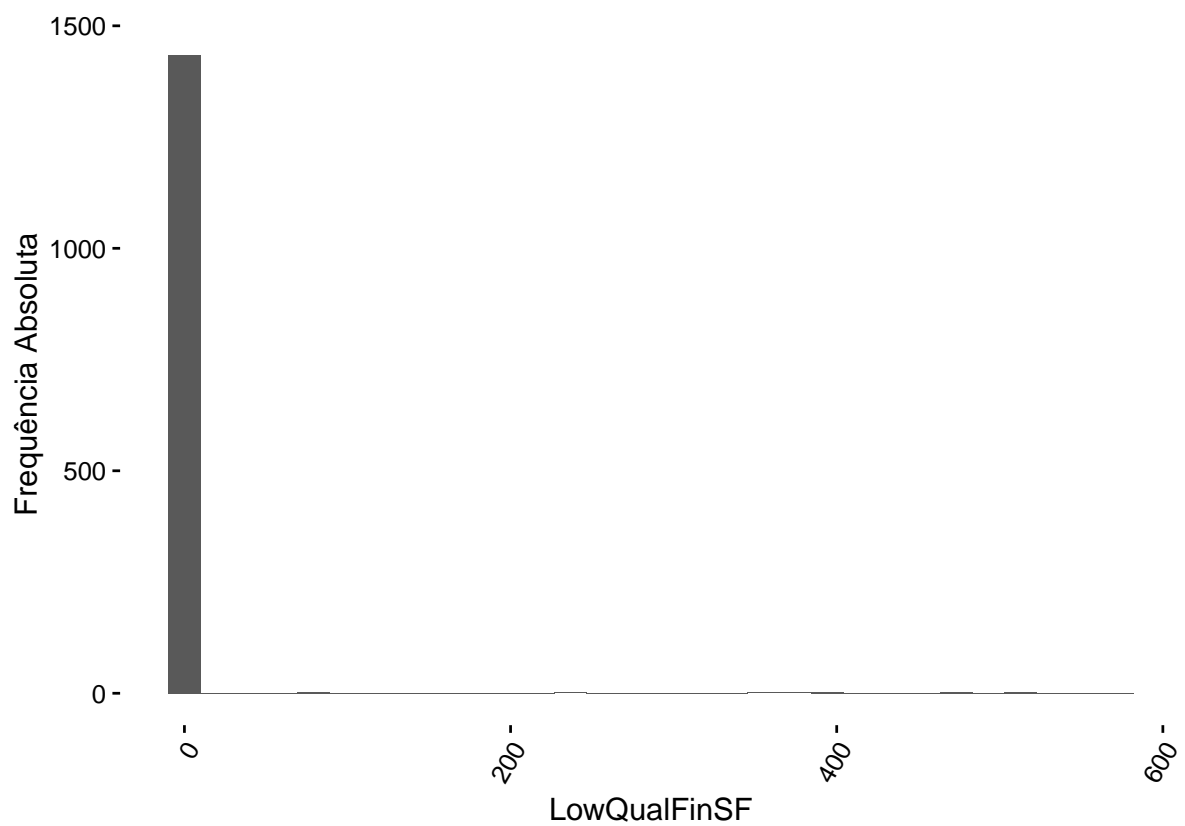
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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



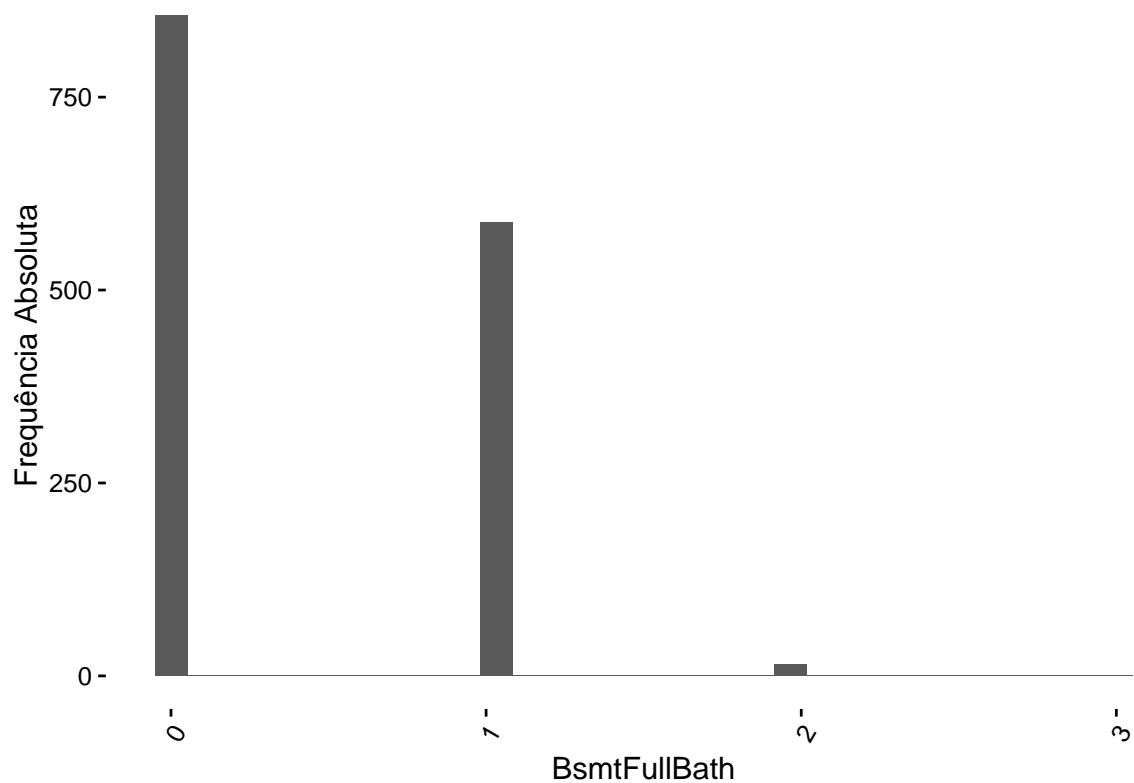
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



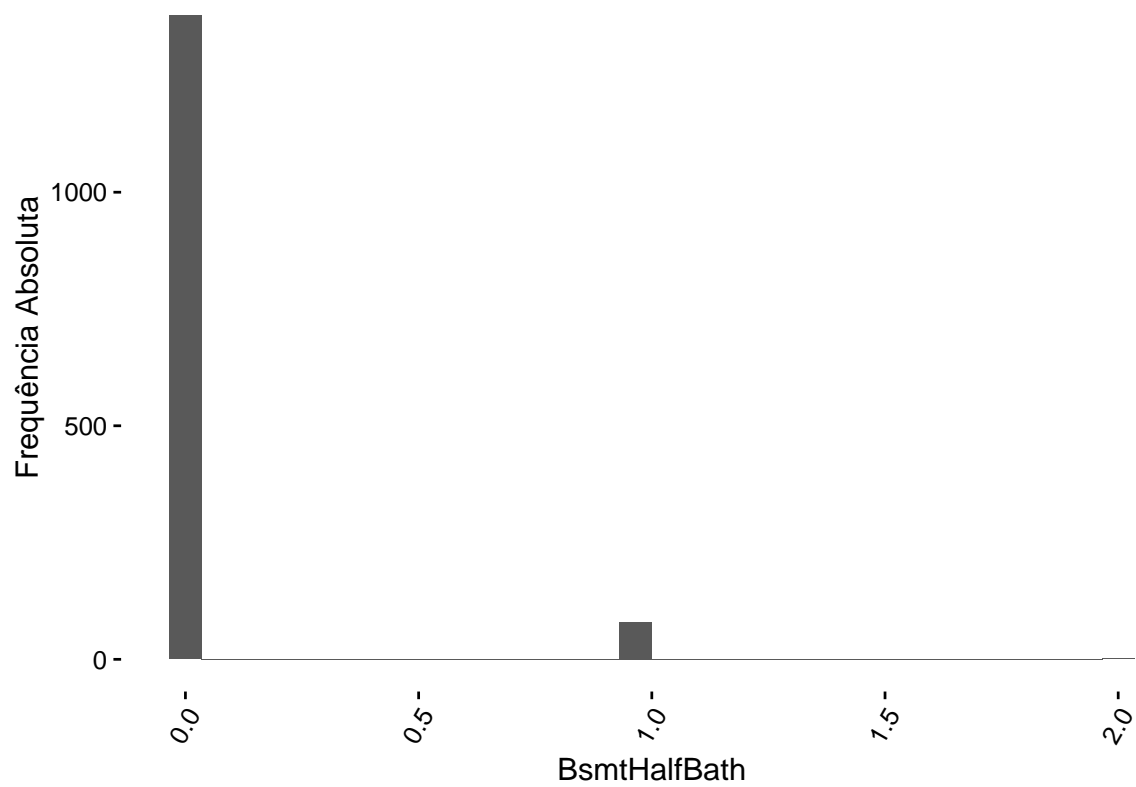
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



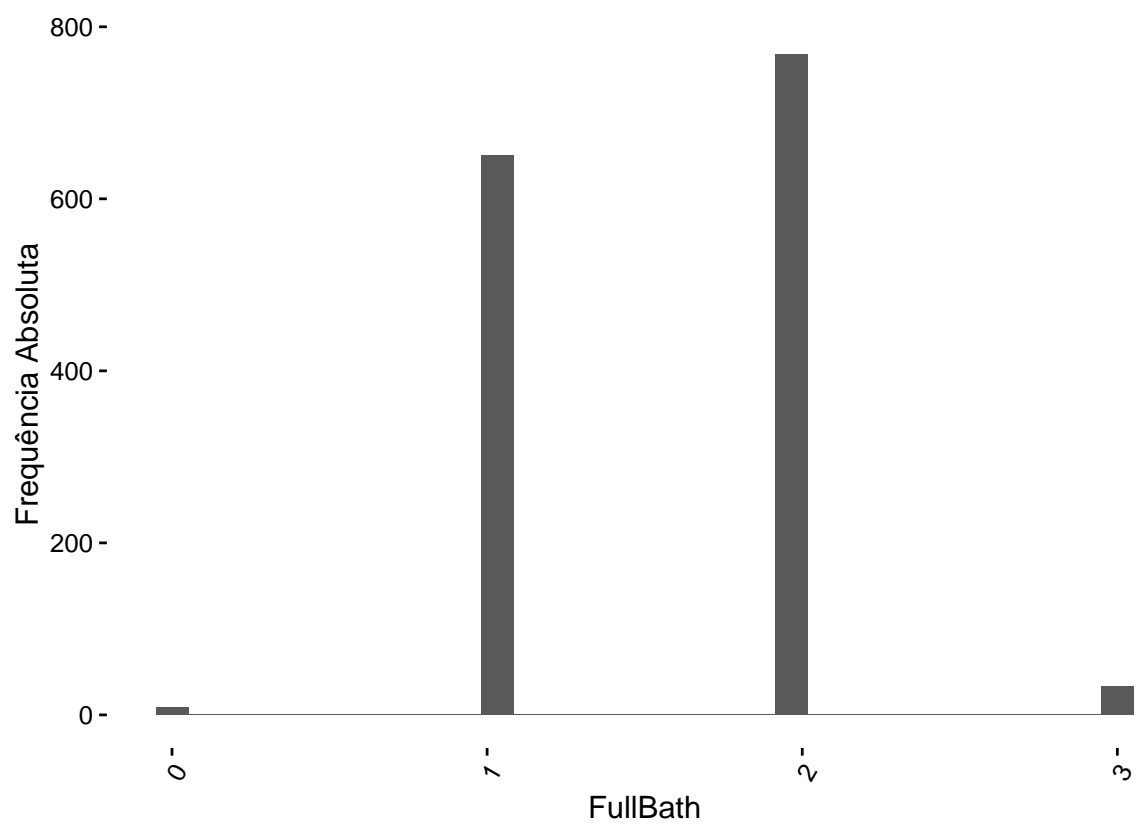
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



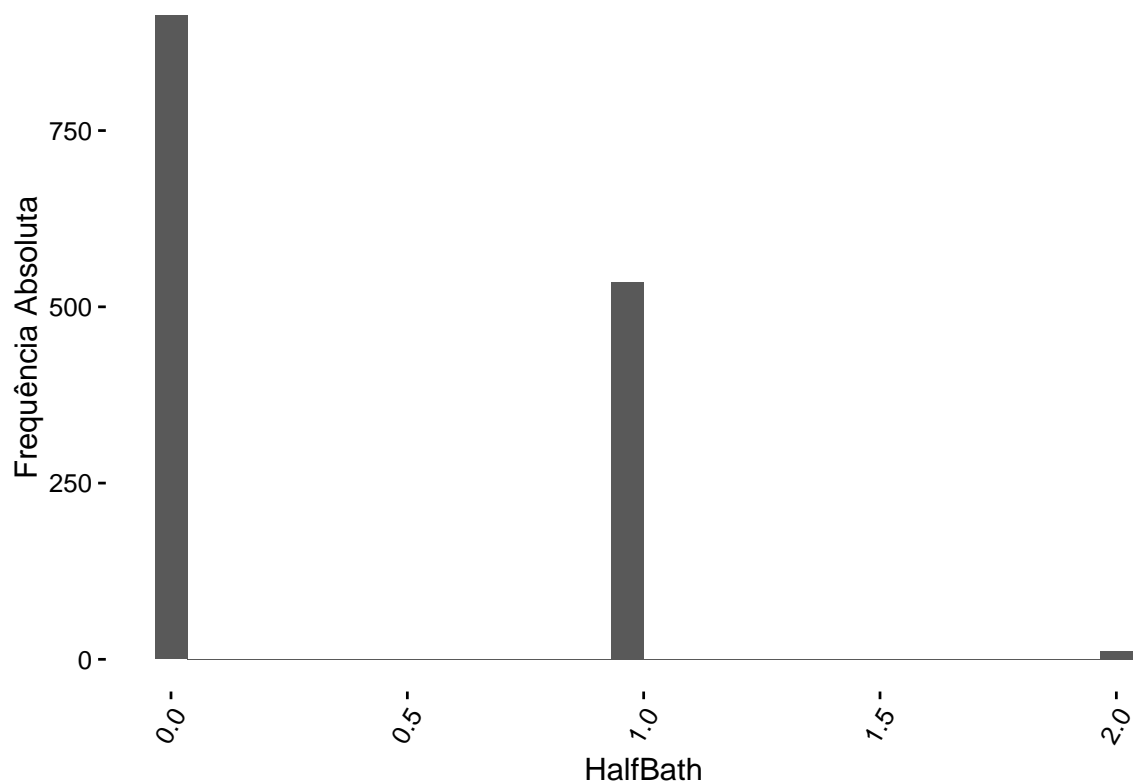
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



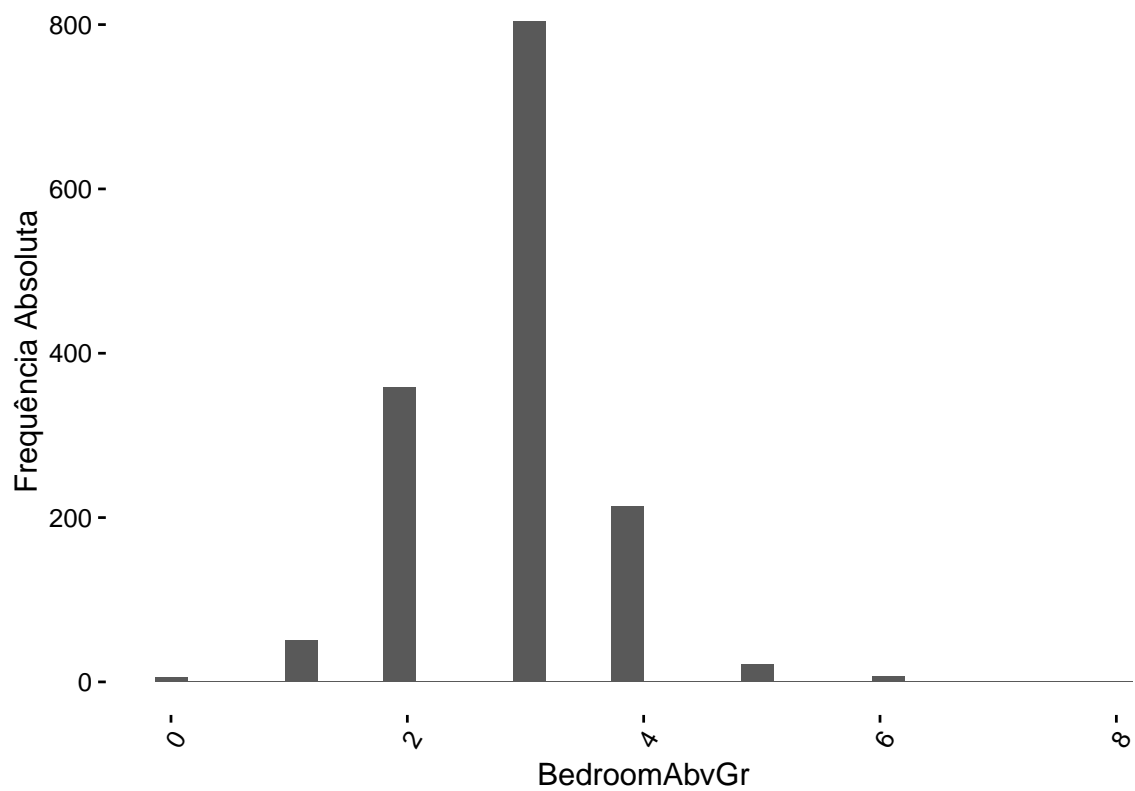
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



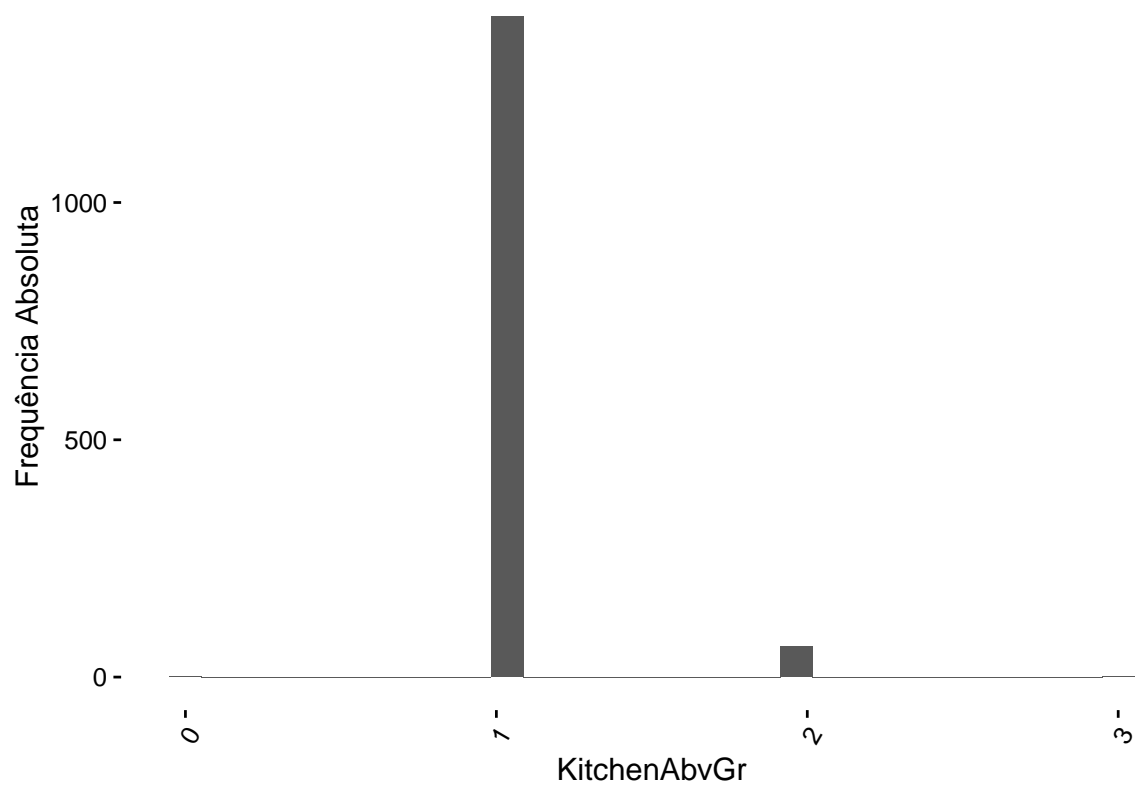
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

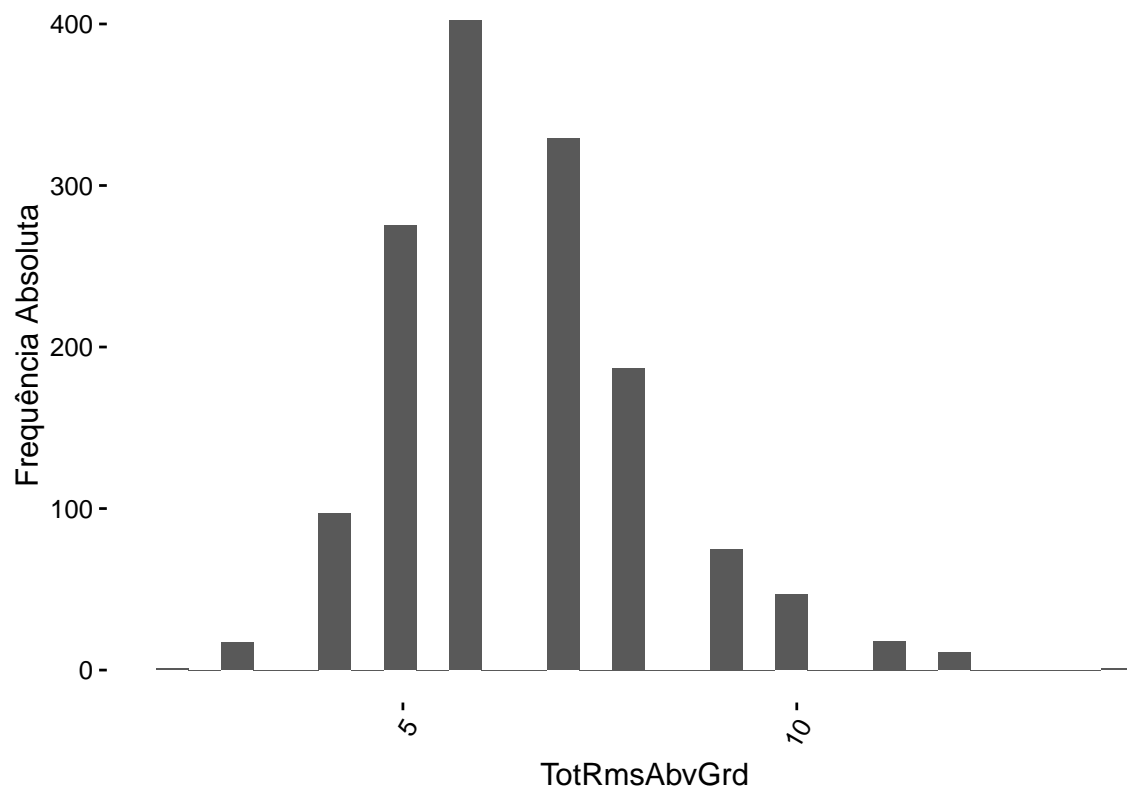
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



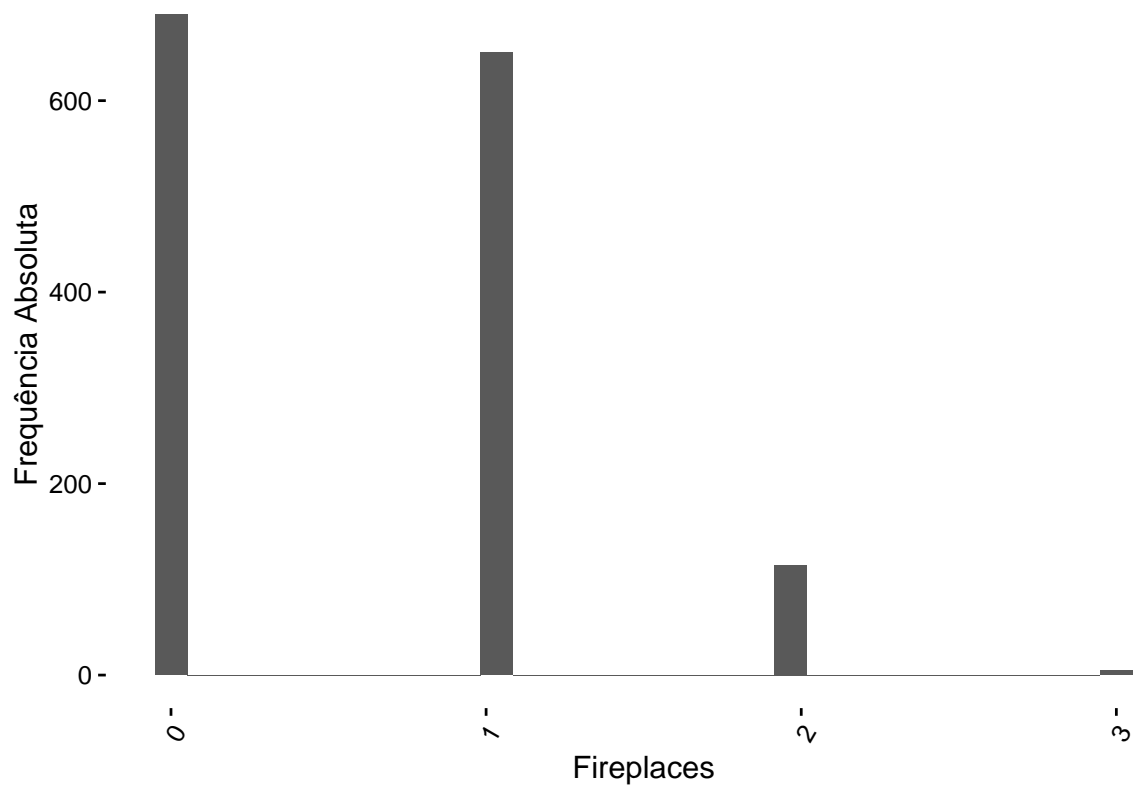
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



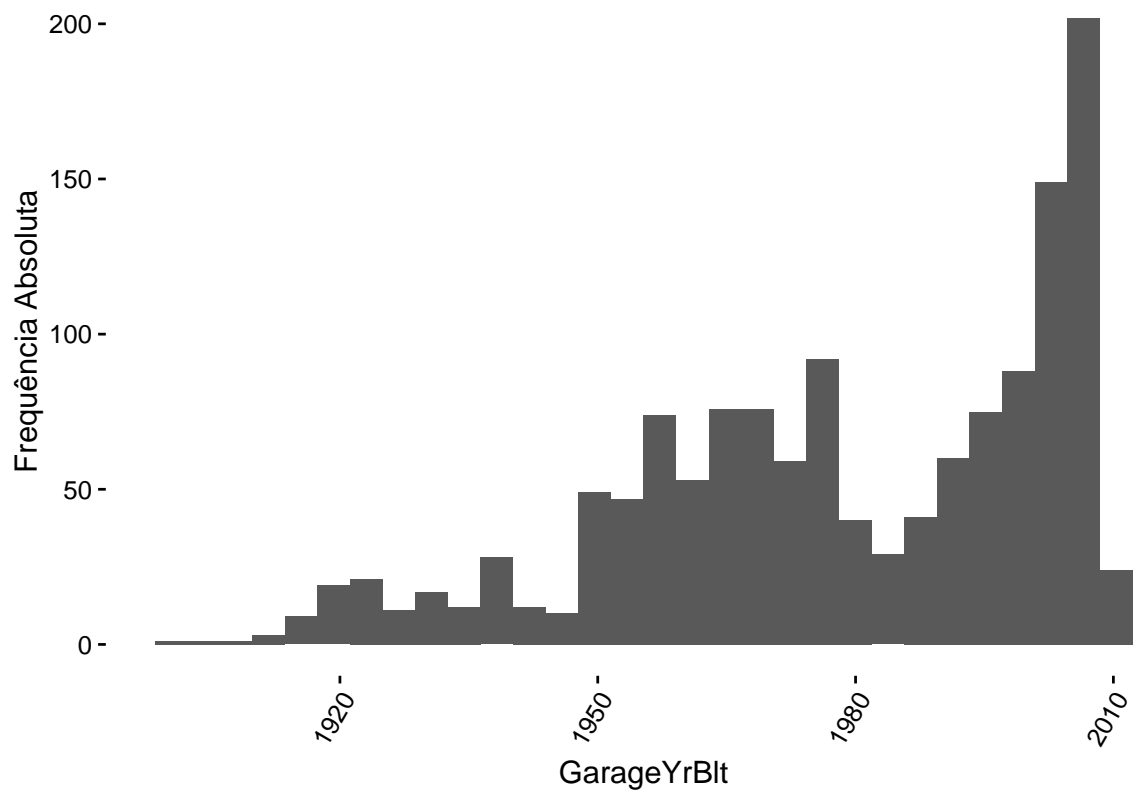
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



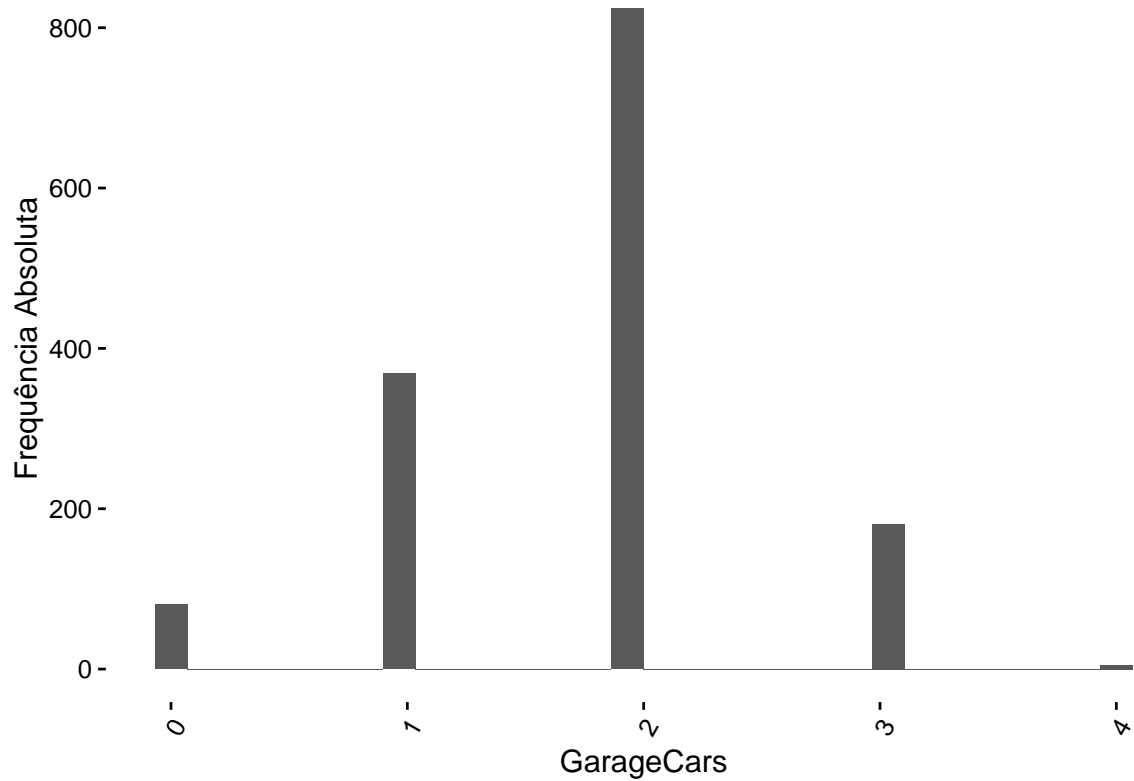
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



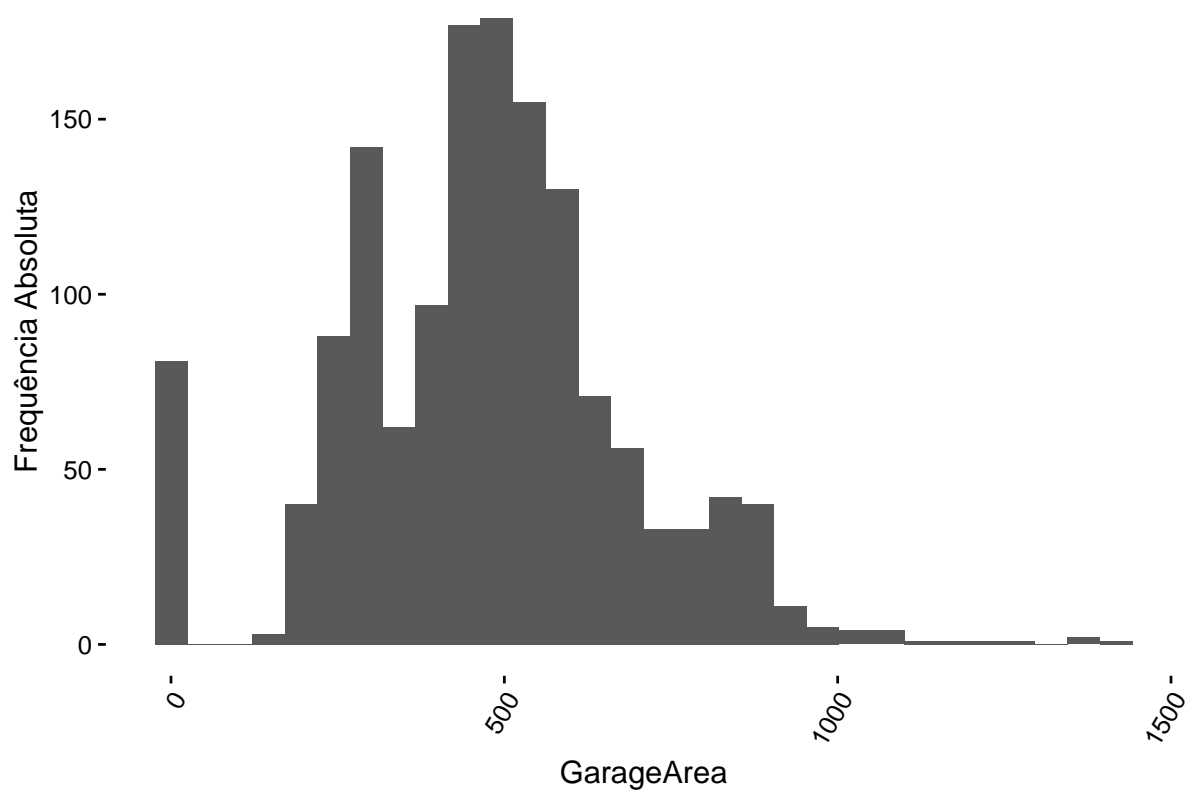
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



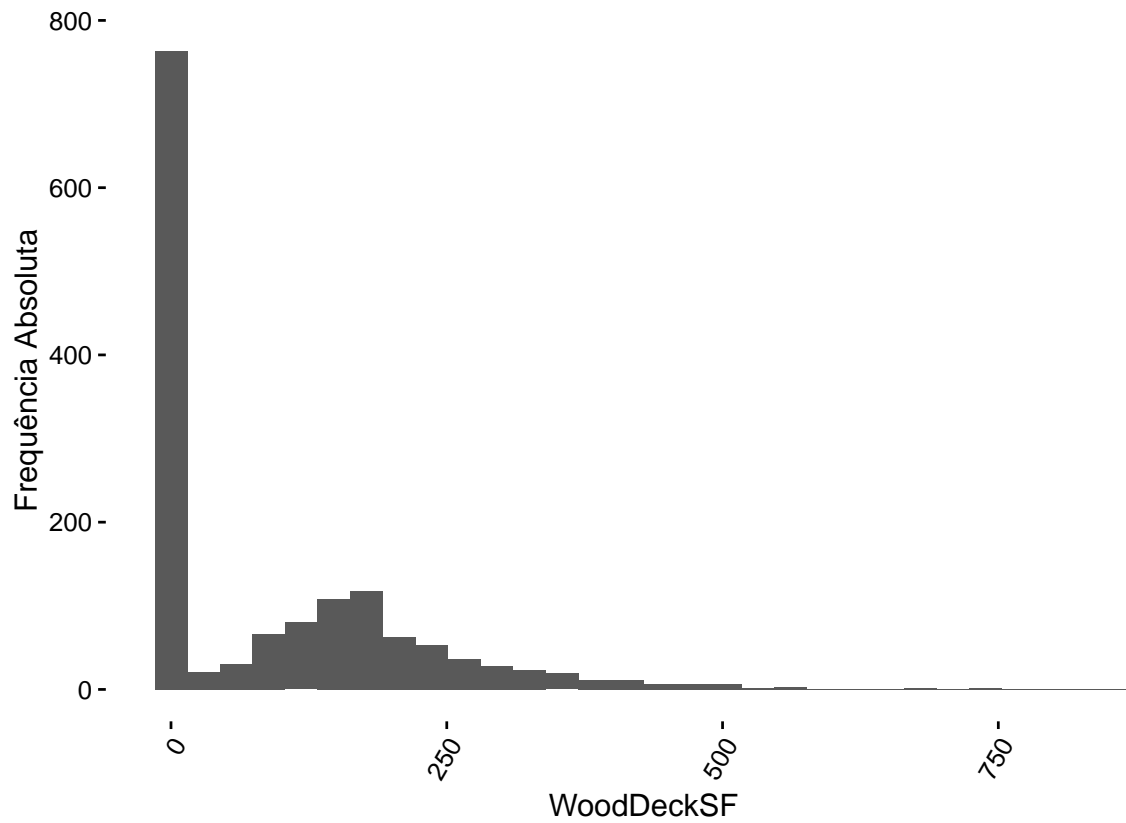
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



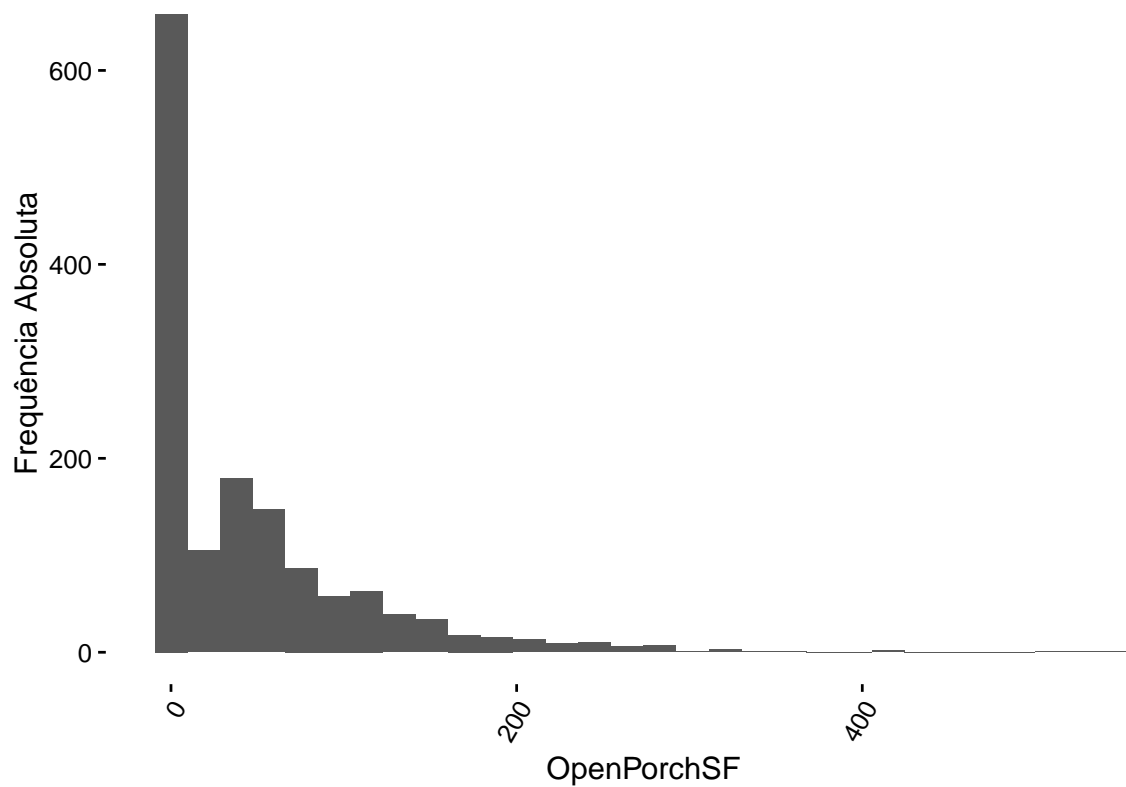
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



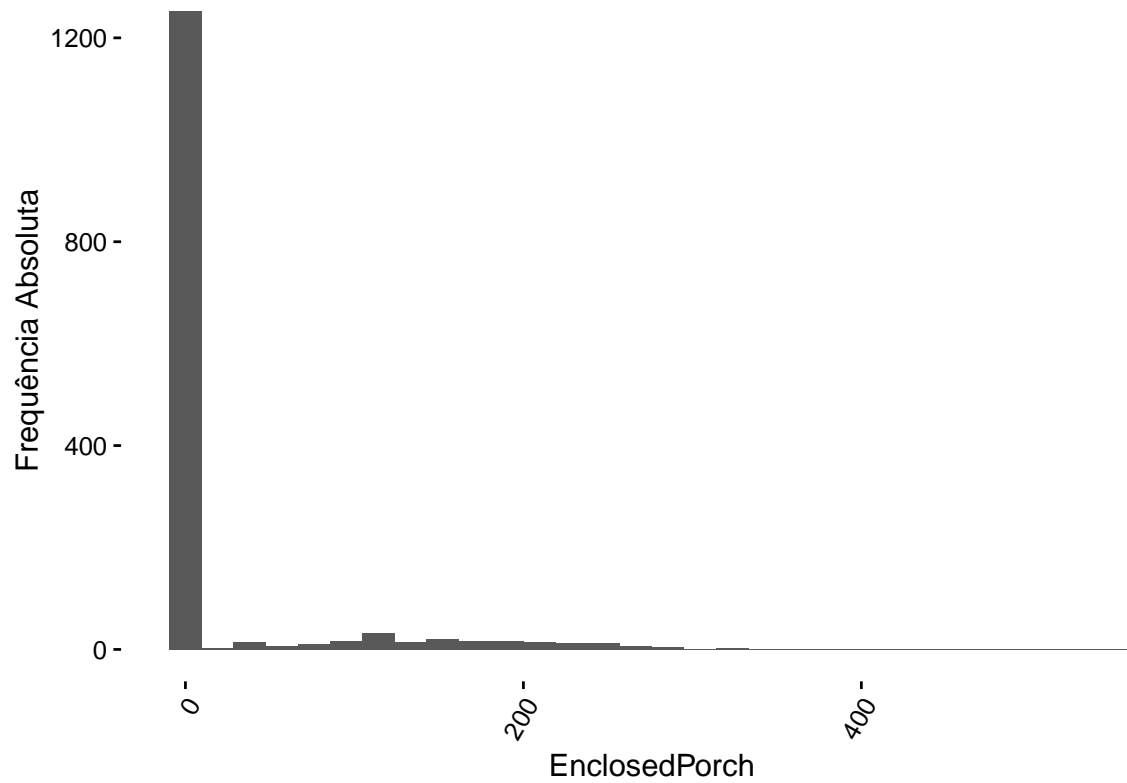
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



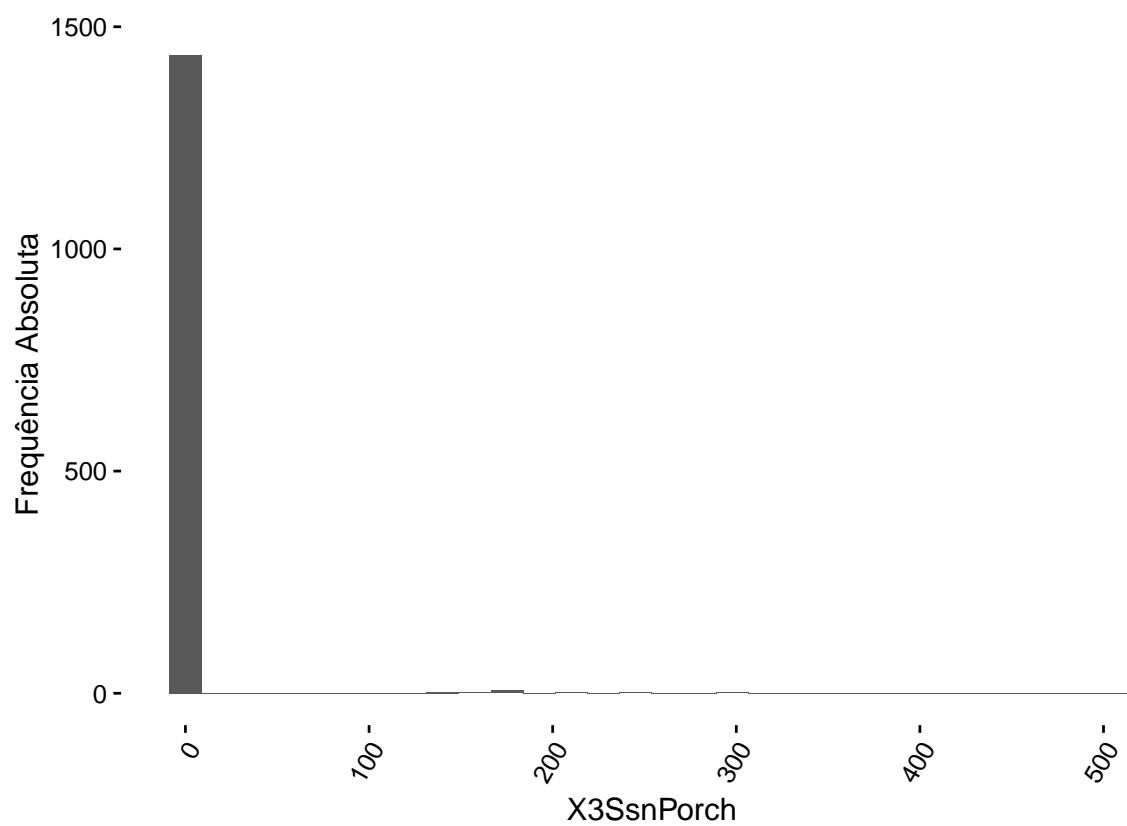
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



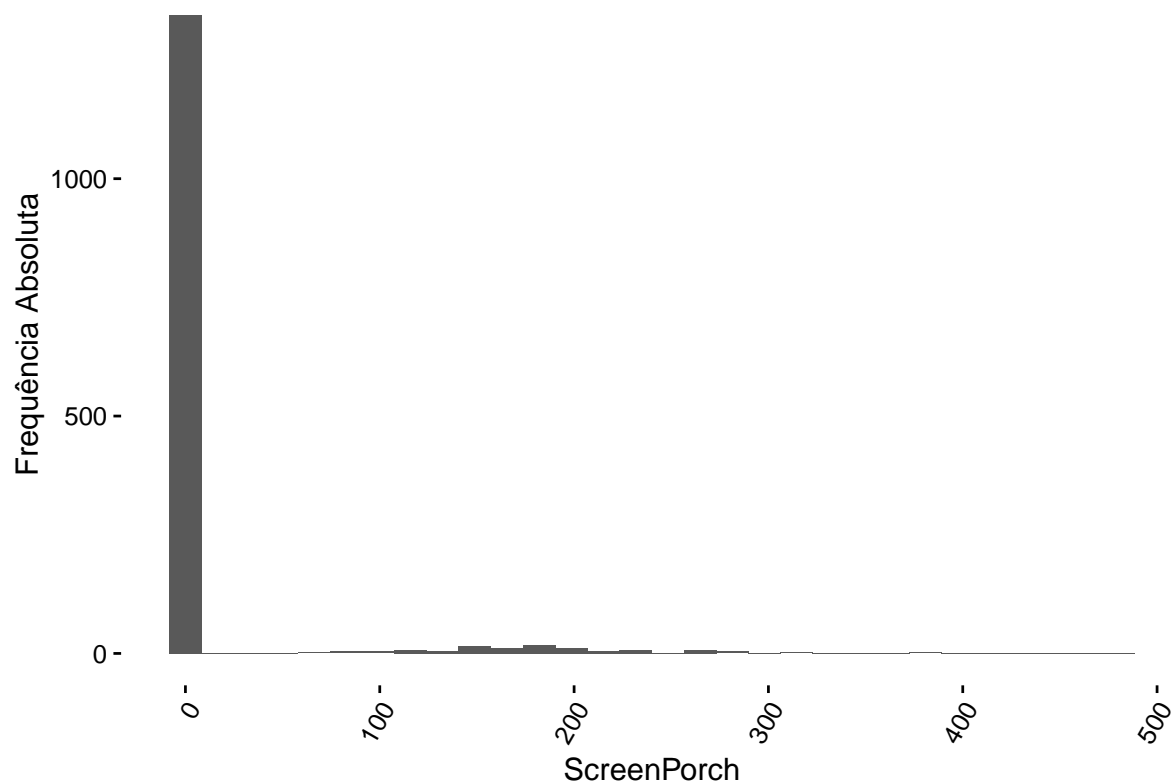
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



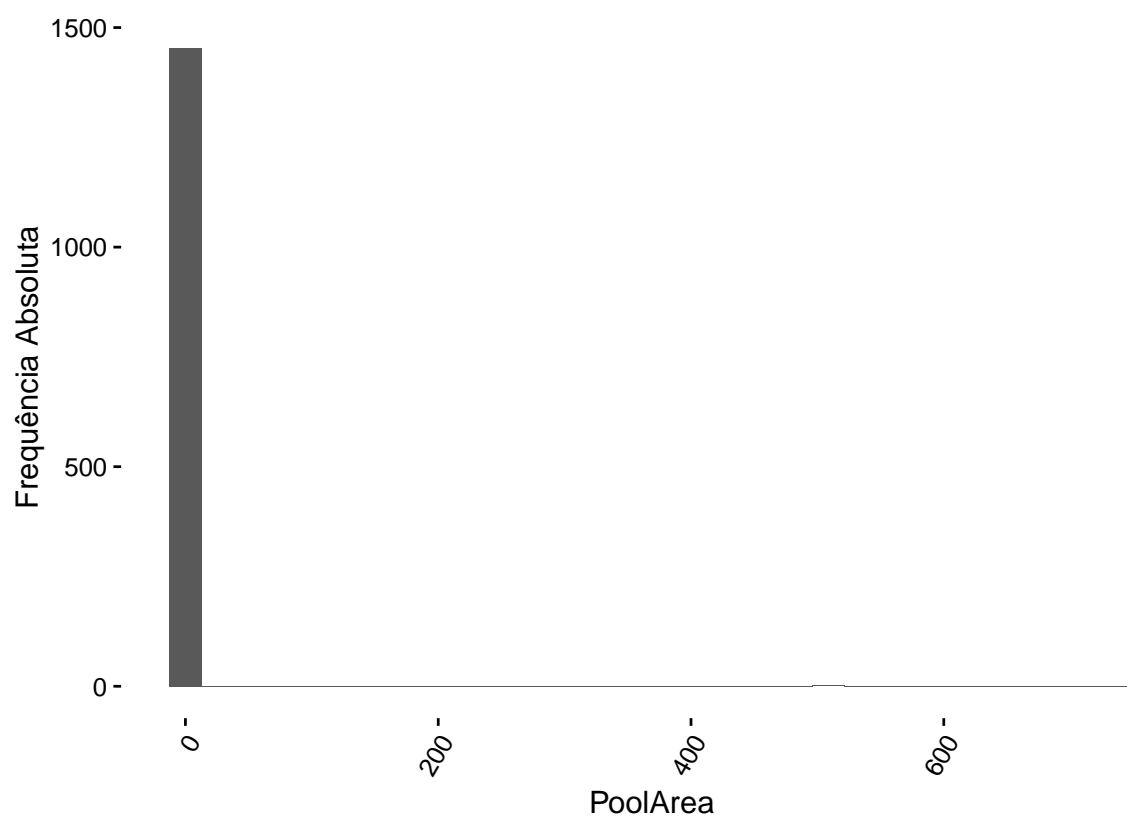
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



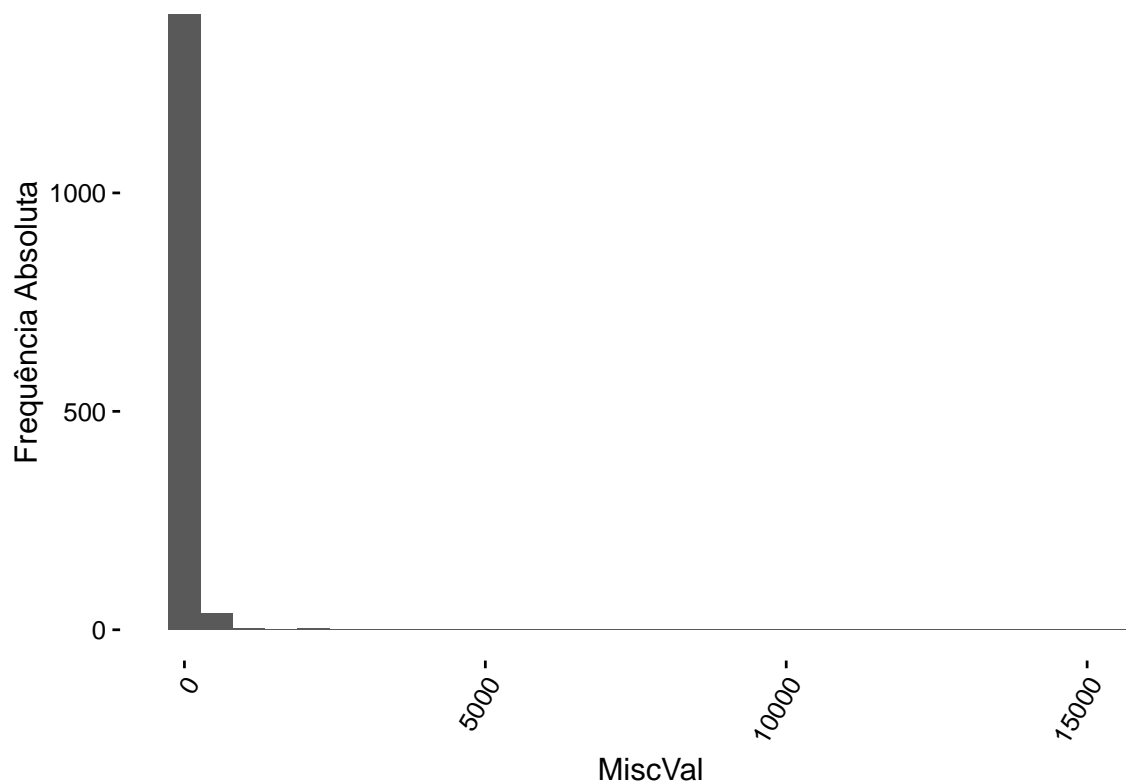
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



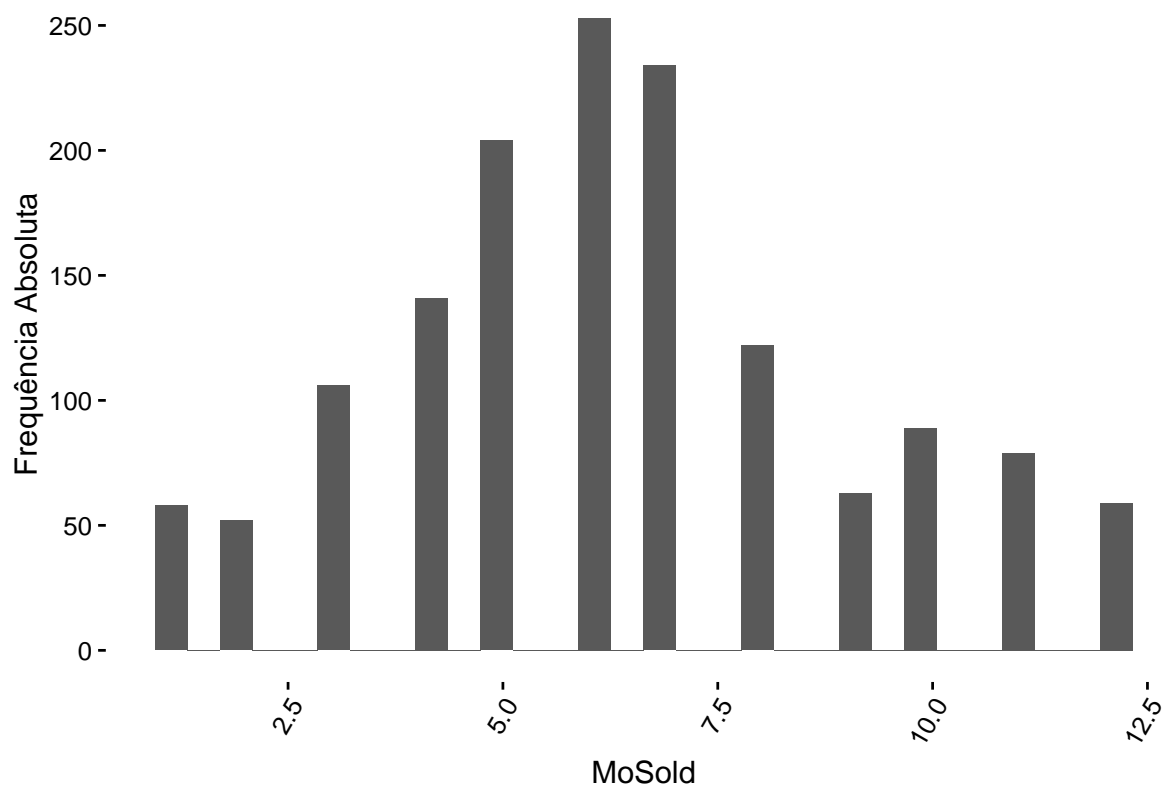
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



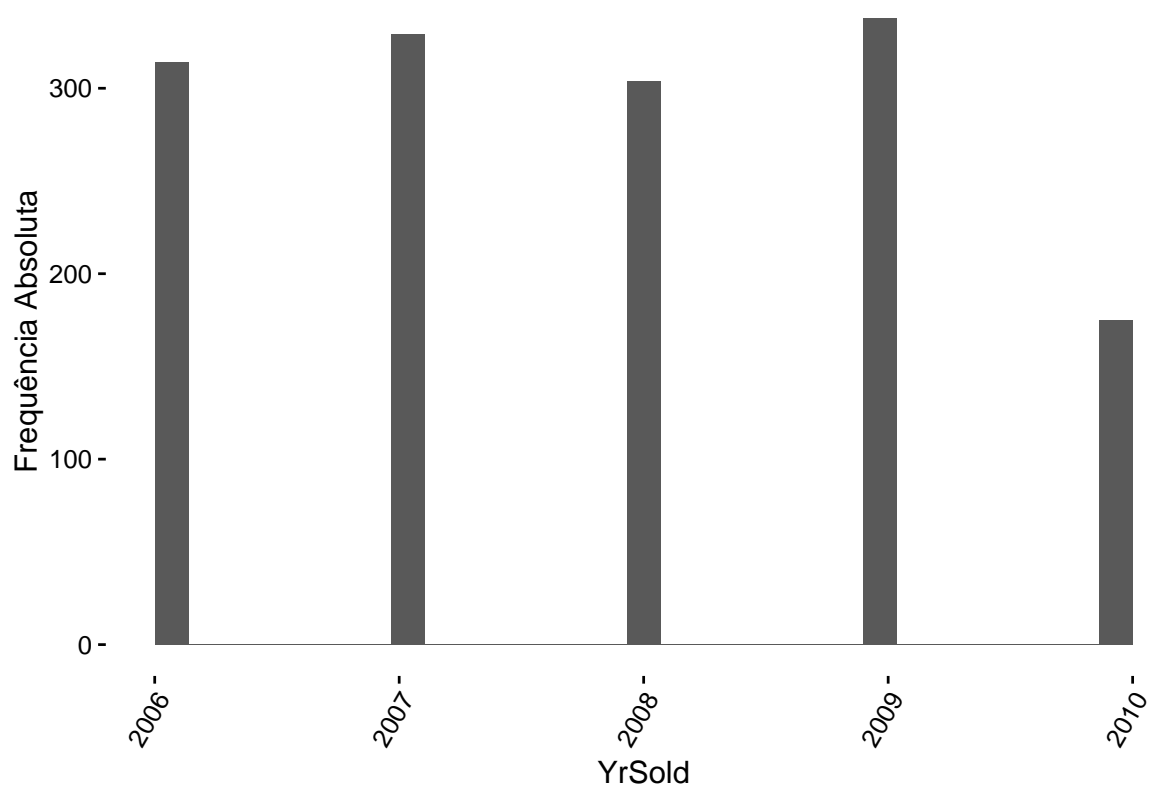
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



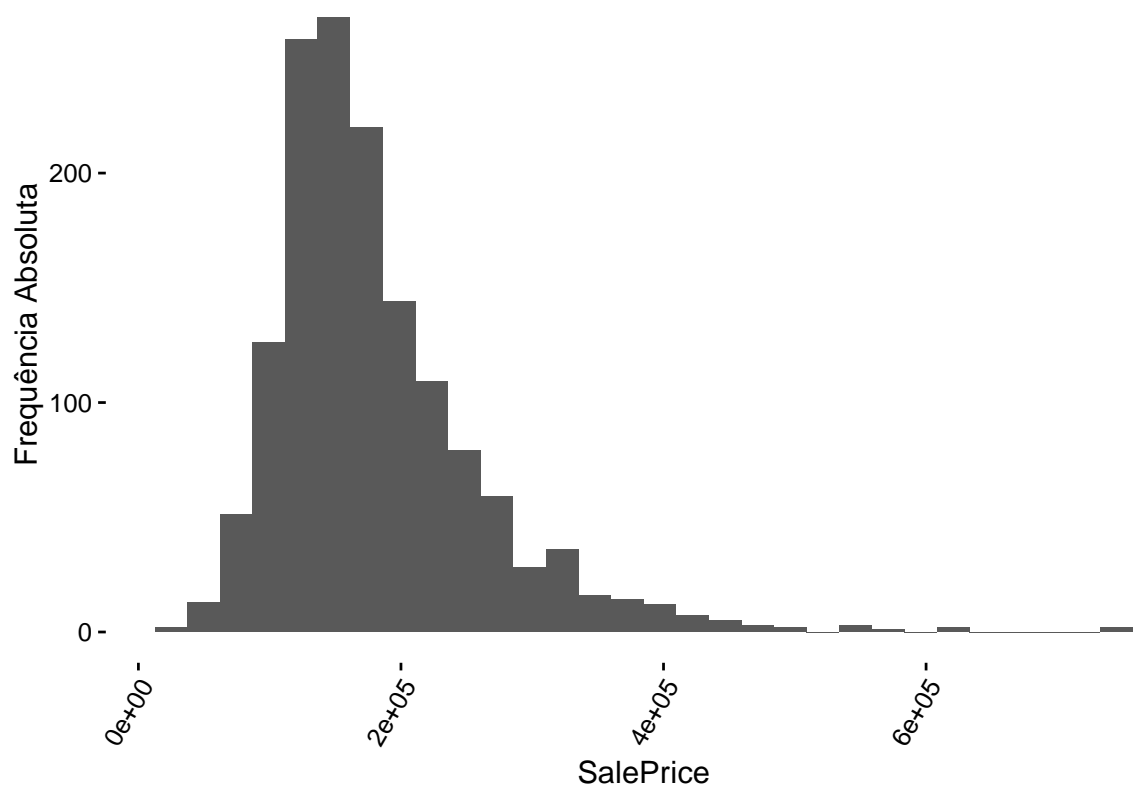
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

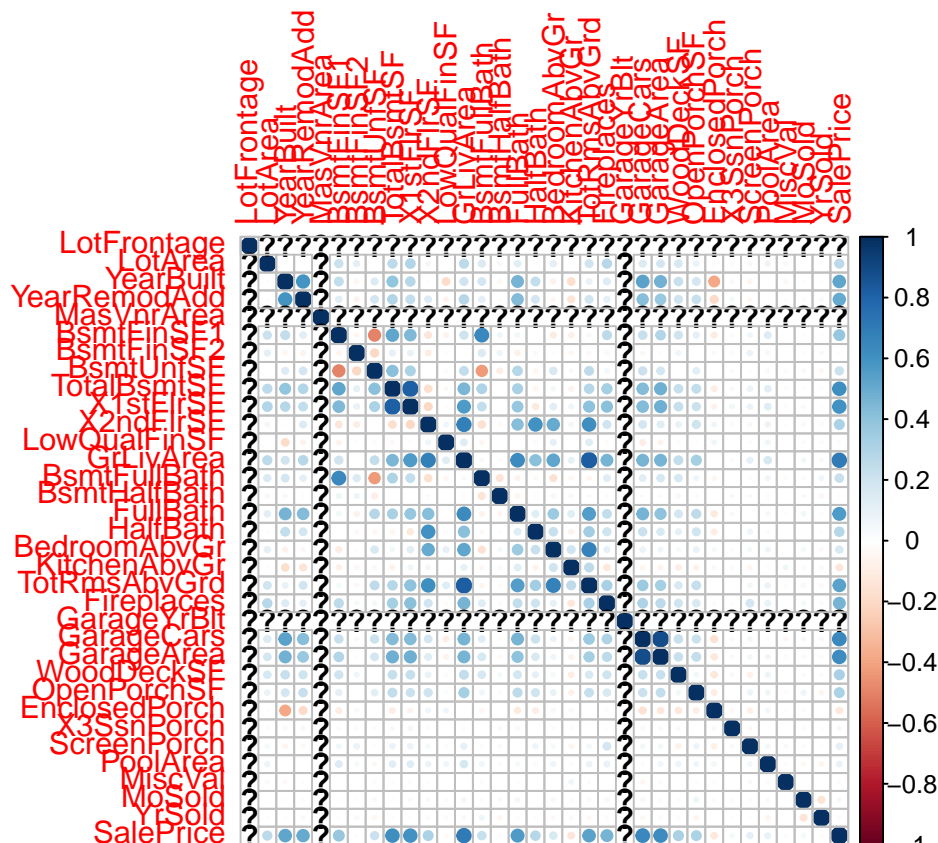


```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

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





APÊNDICES

```
#Pacotes que serão utilizados
#ir incluindo em libs novos pacotes
libs = c('ggplot2','corrplot');
for(ii in libs){
  if(!( ii %in% installed.packages())){
    install.packages(ii);
  }
  library(ii, character.only=TRUE );
}
```

```
# Antes de começar dê um setwd na pasta HousePrices
# (onde quer que ela esteja no seu pc)
```

```
# setwd("Exploratory/");
```

```
# Importando dados
training_data <- read.csv(file="../Data/train.csv", as.is=FALSE);
remover_colunas <- c("Id"); #Boa Raul!
training_data <- training_data[ , !(names(training_data) %in% remover_colunas ) ];

# Transformando as variáveis qualitativas que aparecem como um código numérico em fatores
training_data$MSSubClass <- as.factor(training_data$MSSubClass);
training_data$OverallQual <- factor(training_data$OverallQual,levels= c(1,2,3,4,5,6,7,8,9,10));
training_data$OverallCond <- factor(training_data$OverallCond,levels= c(1,2,3,4,5,6,7,8,9,10));
```

```
# Tipos das variáveis
vars_tipo <- sapply(X=training_data, FUN=class);
```

```

qualitativas <- which(vars_tipo=="factor");

for(ii in qualitativas){
  gg <- ggplot(data=training_data, mapping=aes_string(x=colnames(training_data)[ii]) ) +
    geom_bar() +
    scale_y_continuous(name="Frequência Absoluta") +
    theme_classic()+
    theme(axis.text.x = element_text(angle = 60, hjust = 1))
  print(gg);
}

quantitativas <- which(vars_tipo=="integer");

for(ii in quantitativas){
  gg <- ggplot(data=training_data,
    mapping=aes_string(x=colnames(training_data)[ii]) ) +
    geom_histogram() +
    scale_y_continuous(name="Frequência Absoluta") +
    theme_classic()+
    theme(axis.text.x = element_text(angle = 60, hjust = 1))
  print(gg);
}

#Matriz de Correlação entre as variáveis
var_quant <- training_data[quantitativas] #Seleciona as variáveis quantitativas de training_data
cor_quant <- cor(var_quant) #cria a matriz de correlação

corrplot(cor_quant, method = "circle") #plota a matriz de correlação com círculos

```