

Análisis vinatería

Bustamante Luciano Emanuel

introducción

Se tomo una empresa ficticia de vinatería, se compra vinos a productores locales y se vende a larga escala en el mercado. Uno de los problemas que se encuentra, es que los vinos no se estan vendiendo y no estan generando satisfacción al cliente, realizo este análisis para saber cual es un punto fuerte de la venta del mismo.

Código:

```
#carga las librerias  
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.0.5
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(DataExplorer)
```

```
## Warning: package 'DataExplorer' was built under R version 4.0.5
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(proto)
```

```
## Warning: package 'proto' was built under R version 4.0.5
```

```
library(gsubfn)
```

```
## Warning: package 'gsubfn' was built under R version 4.0.5
```

```
library(sqldf)
```

```
## Warning: package 'sqldf' was built under R version 4.0.5
```

```
## Loading required package: RSQLite
```

```
## Warning: package 'RSQLite' was built under R version 4.0.5
```

```
library(RSQLite)
```

```
#carga los datos que voy a utilizar  
vinos_tintos<- read_csv("vinos_tintos.csv")
```

```
##  
## -- Column specification -----  
## cols(  
##   'fixed acidity' = col_double(),  
##   'volatile acidity' = col_double(),  
##   'citric acid' = col_double(),  
##   'residual sugar' = col_double(),  
##   chlorides = col_double(),  
##   'free sulfur dioxide' = col_double(),  
##   'total sulfur dioxide' = col_double(),  
##   density = col_double(),  
##   pH = col_double(),  
##   sulphates = col_double(),  
##   alcohol = col_double(),  
##   success = col_double(),  
##   country = col_character(),  
##   pricing = col_character()  
## )
```

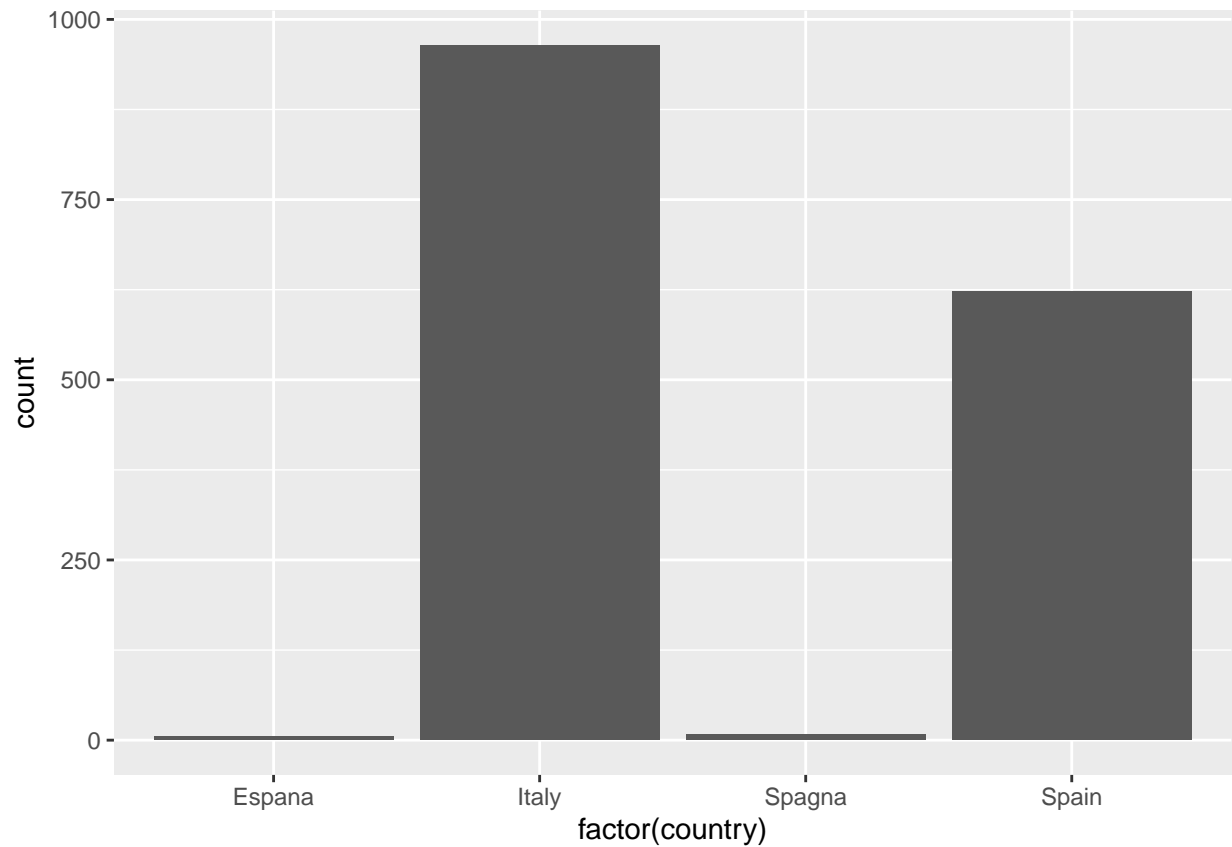
```
head(vinos_tintos)
```

```
## # A tibble: 6 x 14  
##   'fixed acidity' 'volatile acidity' 'citric acid' 'residual sugar' chlorides  
##           <dbl>           <dbl>         <dbl>         <dbl>      <dbl>  
## 1             7.4             0.7           0             1.9      0.076  
## 2             7.8             0.88          0             2.6      0.098  
## 3             7.8             0.76          0.04           2.3      0.092  
## 4            11.2             0.28          0.56           1.9      0.075  
## 5             7.4             0.7           0             1.9      0.076  
## 6             7.4             0.66          0             1.8      0.075  
## # ... with 9 more variables: free sulfur dioxide <dbl>,  
## #   total sulfur dioxide <dbl>, density <dbl>, pH <dbl>, sulphates <dbl>,  
## #   alcohol <dbl>, success <dbl>, country <chr>, pricing <chr>
```

```
#conozco el formato de mi base de datos, encontrando que la mayoría son numericos y algunos son objetos.
str(vinos_tintos)
```

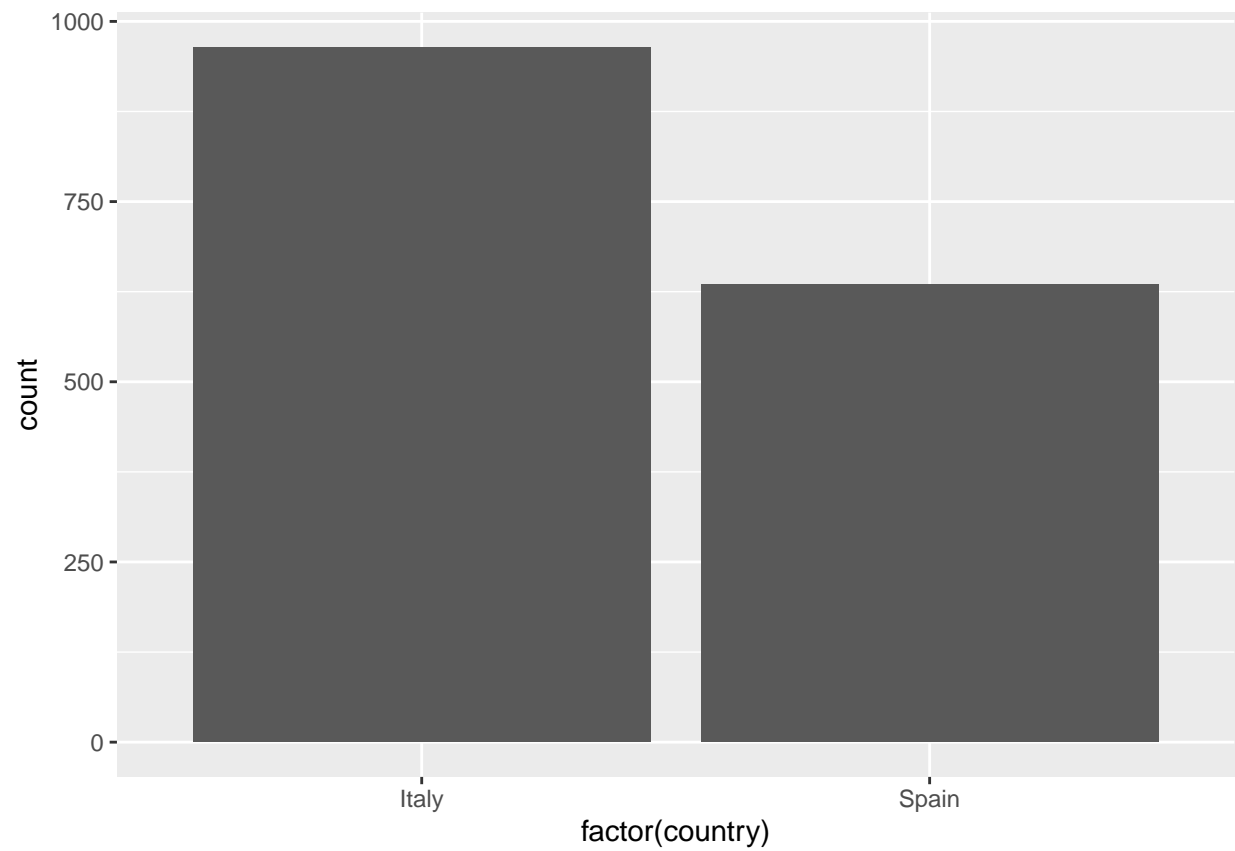
```
## spec_tbl_df[,14] [1,599 x 14] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ fixed acidity      : num [1:1599] 7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 7.5 ...
## $ volatile acidity   : num [1:1599] 0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.5 ...
## $ citric acid        : num [1:1599] 0 0 0.04 0.56 0 0 0.06 0 0.02 0.36 ...
## $ residual sugar     : num [1:1599] 1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 6.1 ...
## $ chlorides          : num [1:1599] 0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065 0.073 0.071 ...
## $ free sulfur dioxide: num [1:1599] 11 25 15 17 11 13 15 15 9 17 ...
## $ total sulfur dioxide: num [1:1599] 34 67 54 60 34 40 59 21 18 102 ...
## $ density            : num [1:1599] 0.998 0.997 0.997 0.998 0.998 ...
## $ pH                 : num [1:1599] 3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.35 ...
## $ sulphates          : num [1:1599] 0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.8 ...
## $ alcohol            : num [1:1599] 0.094 0.098 0.098 0.098 0.094 0.094 0.094 0.1 0.095 0.105 ...
## $ success            : num [1:1599] 49 52 52 72 56 53 57 70 68 53 ...
## $ country            : chr [1:1599] "Spain" "Italy" "Italy" "Italy" ...
## $ pricing            : chr [1:1599] "Medium" "Expensive" "Expensive" "Budget" ...
## - attr(*, "spec")=
## .. cols(
## ..   'fixed acidity' = col_double(),
## ..   'volatile acidity' = col_double(),
## ..   'citric acid' = col_double(),
## ..   'residual sugar' = col_double(),
## ..   chlorides = col_double(),
## ..   'free sulfur dioxide' = col_double(),
## ..   'total sulfur dioxide' = col_double(),
## ..   density = col_double(),
## ..   pH = col_double(),
## ..   sulphates = col_double(),
## ..   alcohol = col_double(),
## ..   success = col_double(),
## ..   country = col_character(),
## ..   pricing = col_character()
## .. )
```

```
# quiero saber que cuantos paises se encuentran dentro de mis datos, vemos anomalias
graficobarras <- ggplot( data = vinos_tintos,
  mapping = aes(x= factor(country))) +
  geom_bar()
graficobarras
```

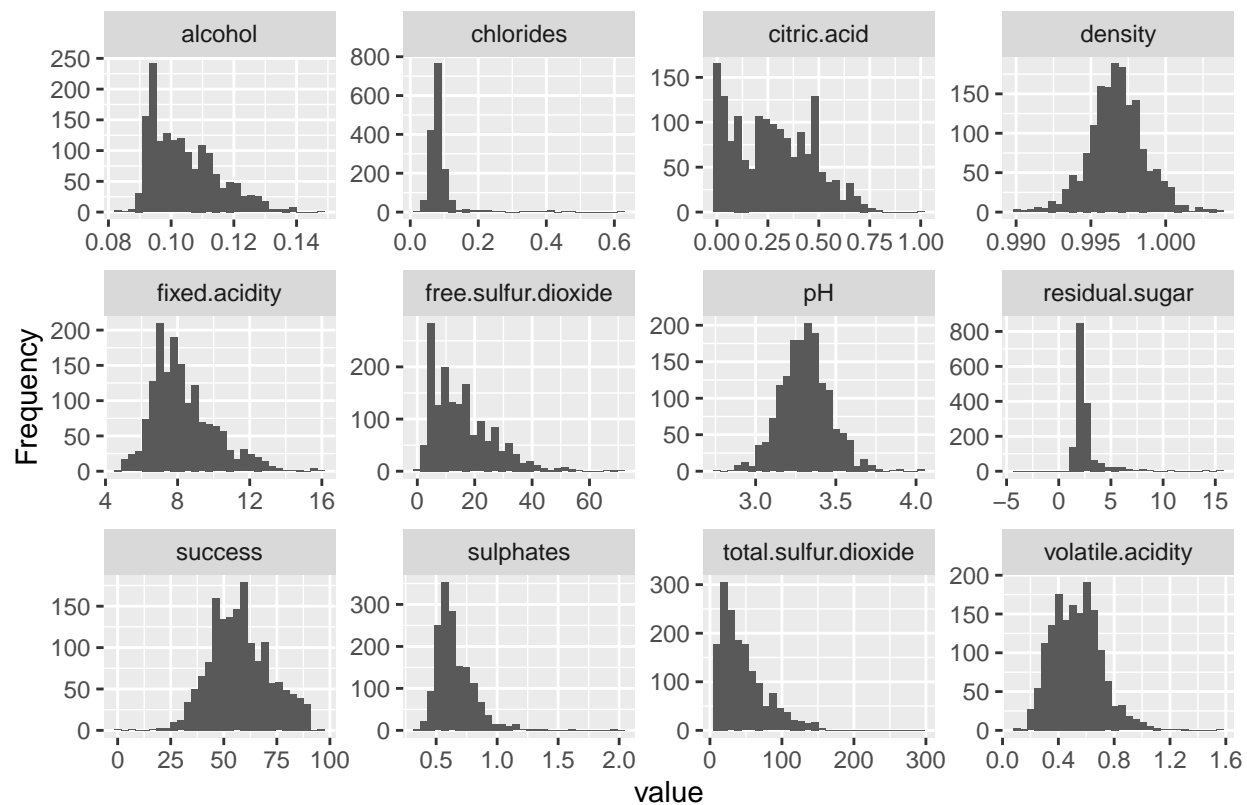


```
#observo que los datos estan mal ordenados, voy a unir espana y spagna con spain usando un bucle
for ( i in 1:length(vinos_tintos$country)) {
  if (vinos_tintos$country[i] == 'Espana' || vinos_tintos$country[i] == 'Spagna') {
    vinos_tintos$country[i] = 'Spain'
  }
}
```

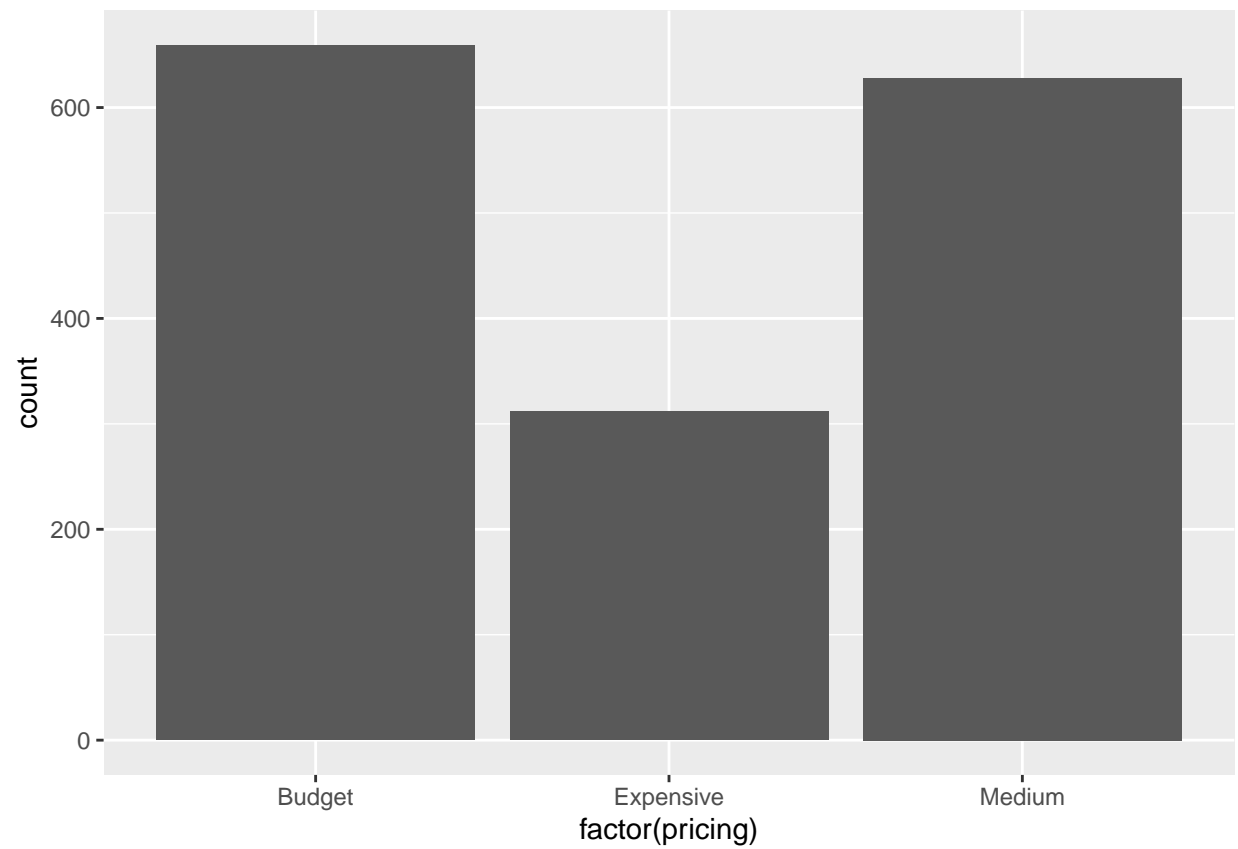
```
#vuelvo a graficar ya con los datos acomodados
graficobarras <- ggplot( data = vinos_tintos,
  mapping = aes(x= factor(country))) +
  geom_bar()
graficobarras
```



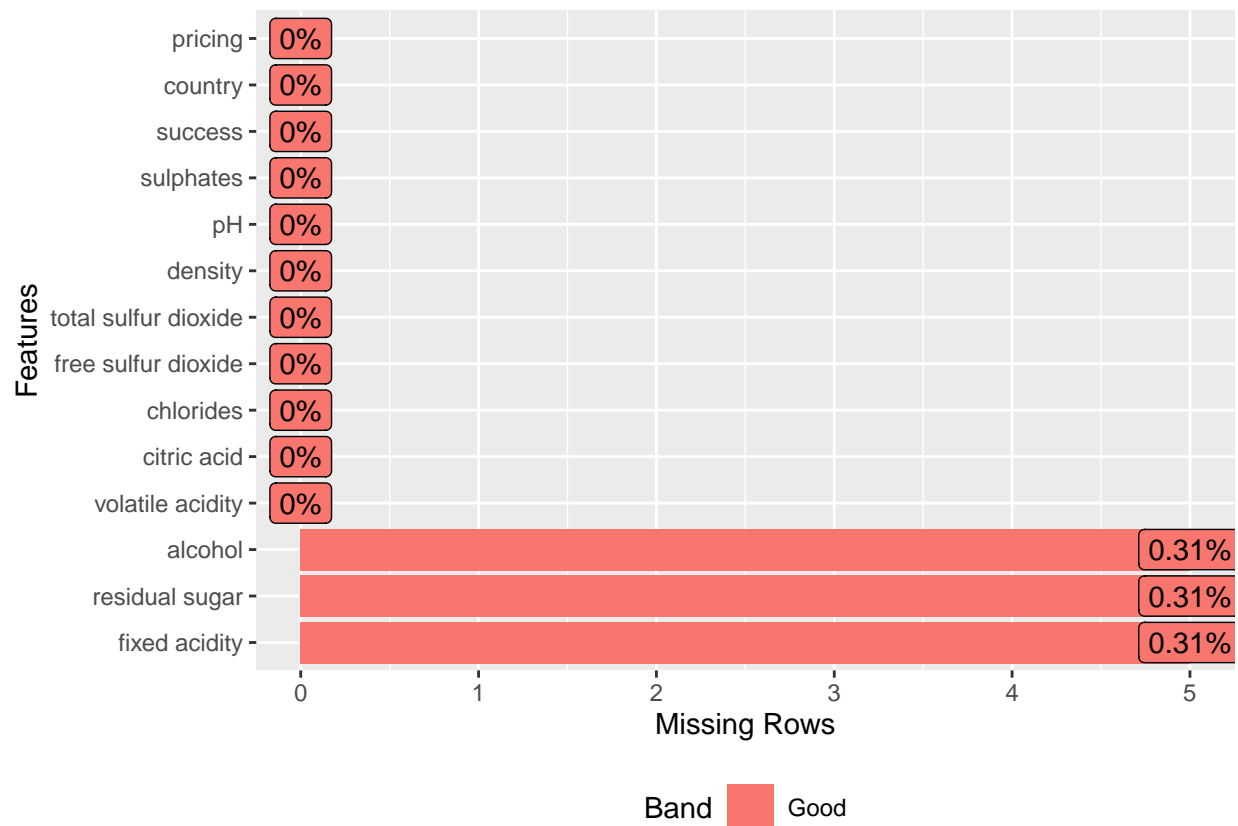
```
attach(vinos_tintos)
#histograma de todos los componentes
plot_histogram(vinos_tintos)
```



```
#grafico de barras de los precios
graficobarrasprecio <- ggplot( data = vinos_tintos,
  mapping = aes(x= factor(pricing))) +
  geom_bar()
graficobarrasprecio
```

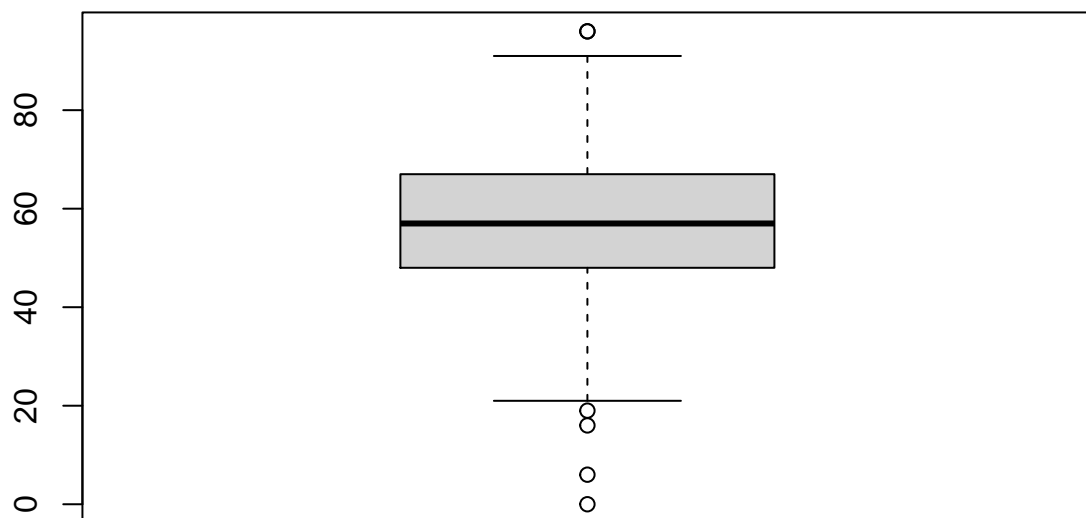


```
# quiero saber cual es el porcentaje de valores faltantes  
plot_missing(vinos_tintos)
```

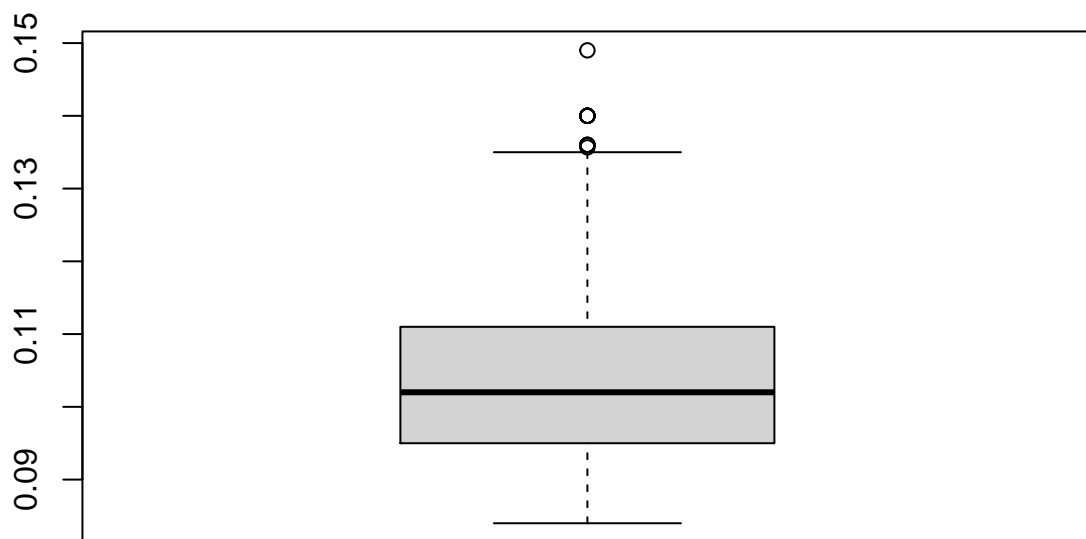


```
#elimino los datos faltantes
vinos_tintos <- na.omit(vinos_tintos)
```

```
#boxplot de alcohol y success
#la mayoria en promedio esta cerca del 60% de alcohol
boxplot(success)
```

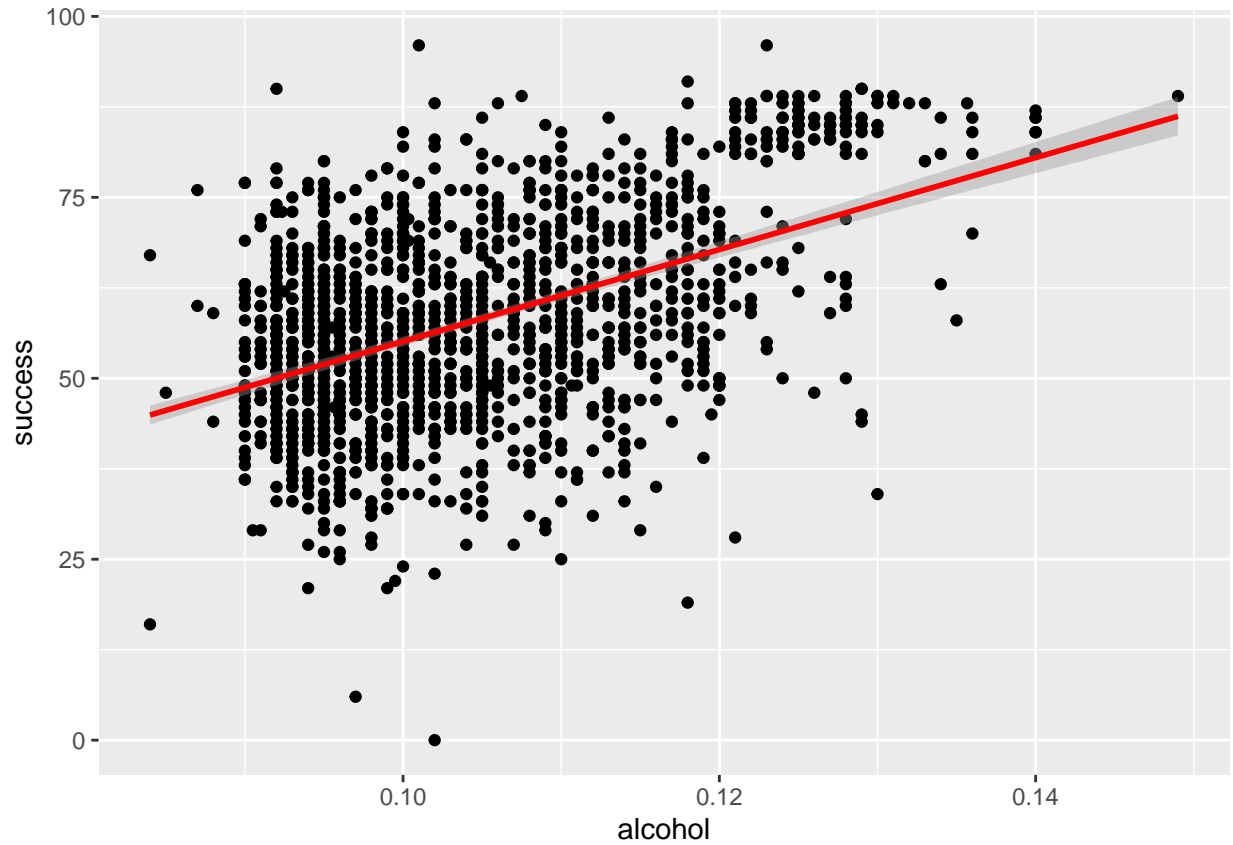



```
boxplot(alcohol)
```



```
# grafico de regresión y dispersión  
ggplot( data = vinos_tintos) +  
  aes(x = alcohol, y = success) +  
  geom_point() +  
  geom_smooth( method = 'lm', col = 'red')
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



Conclusión:

-Podemos deducir, con nuestro análisis, que uno de los factores determinantes en el precio y éxito es el alcohol, ya que tiene una gran tendencia a crecer. Por lo tanto, más alcohol, mejor serán mis ventas.