PR2-Tipologia

Octavi Castro Nuez

27 de desembre de 2017

Contents

L'objectiu d'aquesta activitat serà el tractament d'un dataset, que pot ser el creat a la pràctica 1 o bé qualsevol dataset lliure disponible a Kaggle (https://www.kaggle.com). Les diferents tasques a realitzar (i justificar) són les següents:

1 Descripció del dataset. Perquè és important i quina pregunta/problema pretèn respondre?

Per aquesta activitat triem un dataset de kaggle, concretament, un sobre vins que es pot trobar a https://www.kaggle.com/zynicide/wine-reviews

A l'adreça anterior trobarem dos datasets, enlloc d'un, i treballarem amb els dos, per poder tenir un major nombre de mostres de vins. El primer dataset que trobem conté més de 150 mil vins i no conté tots els camps. El segon dataset conté una mica menys de 130 mil mostres i disposa de tres camps més que l'anterior.

Aquests datasets contenen informació sobre vins que han obtingut una puntuació entre 80 i 100 punts (el màxim és 100 punts). Aquestes dades van ser obtingudes mitjançant scraping de WineEnthusiast (http://www.winemag.com/?s=&drink_type=wine) durant la setmana del 15 de juny de 2017.

Passem a veure la llista d'atributs:

- Points: el nombre de punts obtinguts pel vi, va de 1 fins a 100, però aquí només hi ha vins amb una puntuació de 80 o més.
- Variety: el tipus de raïm que s'utilitza per elaborar el vi
- Descripction: unes poques frases del tastador del vi descrivint el tast.
- Country: el país d'on prové el vi.

- Province: la província o estat d'on prové el vi. (Comentar que Province es refereix més aviat a la zona on es produeix el vi o a la seva denominació d'origen, ja que si la revisem podrem veure que per Country = Spain tenim una provincia anomenada Northern Spain que correspondria a les tres comunitats automes que conformen la D.O. Rioja.)
- Region 1: l'àrea vinícola d'una província o estat.
- Region 2: de vegades hi ha una regió més específica de l'àrea vinícola, però aquest camp pot estar en blanc.
- Winery: el celler que ha fet el vi.
- Designation: la vinya dins del celler d'on procedeixen els raïms que han fet el vi.
- Price: el cost per una ampolla del vi (en dollars).
- Taster Name: el nom de la persona que va fer el tast i la ressenya del vi.
- Taster Twitter Handle: compte a Twitter del tastador del vi.
- Title: El títol del vi i en molts casos la data de la verema.

Els tres últims camps només es troben presents en el segon dataset.

En aquests datasets trobem força informació sobre vins amb una bona puntuació, i del qual podem veure alguns estudis fets. En el nostre cas pretendrem respondre a la pregunta següent:

Quina zona m'ofereix la millor relació qualitat-preu per a una varietat concreta?

```
# els valors absents venen indicats per un camp en blanc.
# llegim el primer dataset.
wine.150 <- read.csv("./csv/winemag-data_first150k.csv", na.strings = "")
# llegim el segon dataset.
wine.130 <- read.csv("./csv/winemag-data-130k-v2.csv", na.strings = "")
```

2 Neteja de les dades.

Examinem les dades dels datasets.

En el primer dataset tenim 150930 mostres i un total de 11 camps.

Amb els factors següents:

```
str(wine.150)
```

```
## 'data.frame':    150930 obs. of  11 variables:
##  $ X          : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ country    : Factor w/ 48 levels "Albania","Argentina",..: 47 41 47 47 16 41 41 41 47 47 ...
##  $ description: Factor w/ 97821 levels ". Big, lively and very intense, this powerful Amarone opens w
##  $ designation: Factor w/ 30621 levels "¡Adentro! Red",..: 17369 4413 25554 22403 14344 19205 23925 4
##  $ points     : int  96 96 96 96 95 95 95 95 95 95 ...
##  $ price      : num  235 110 90 65 66 73 65 110 65 60 ...
##  $ province   : Factor w/ 455 levels "Achaia","Aconcagua Costa",..: 52 275 52 283 315 275 275 275 283
##  $ region_1   : Factor w/ 1236 levels "Abruzzo","Adelaida District",..: 739 1071 529 1223 67 1071 107
##  $ region_2   : Factor w/ 18 levels "California Other",..: 8 NA 14 18 NA NA NA NA 18 14 ...
##  $ variety    : Factor w/ 632 levels "Agiorgitiko",..: 71 550 470 403 423 550 550 550 403 403 ...
##  $ winery     : Factor w/ 14810 levels ":Nota Bene","'37 Cellars",..: 7305 1240 9050 11038 5106 10203
```

En el segon dataset tenim 129971 mostres i un total de 14 camps.

Amb els factors següents:

```
str(wine.130)
```

```
## 'data.frame':    129971 obs. of  14 variables:
##  $ X                    : int  0 1 2 3 4 5 6 7 8 9 ...
```

```
##  $ country              : Factor w/ 43 levels "Argentina","Armenia",..: 23 32 43 43 43 38 23 16 18 1
##  $ description          : Factor w/ 119955 levels ". A delightfully intriguing "White Burgundy" blen
##  $ designation          : Factor w/ 37979 levels "??? Vineyard",..: 36976 2352 NA 28123 36715 1996 3
##  $ points               : int  87 87 87 87 87 87 87 87 87 87 ...
##  $ price                : num  NA 15 14 13 65 15 16 24 12 27 ...
##  $ province             : Factor w/ 425 levels "Achaia","Aconcagua Costa",..: 334 110 269 220 269 263
##  $ region_1             : Factor w/ 1229 levels "Abruzzo","Adelaida District",..: 425 NA 1218 550 12
##  $ region_2             : Factor w/ 17 levels "California Other",..: NA NA 17 NA 17 NA NA NA NA NA .
##  $ taster_name          : Factor w/ 19 levels "Alexander Peartree",..: 10 16 15 1 15 13 10 16 2 16 .
##  $ taster_twitter_handle: Factor w/ 15 levels "@AnneInVino",..: 5 11 8 NA 8 13 5 11 NA 11 ...
##  $ title                : Factor w/ 118840 levels ":Nota Bene 2005 Una Notte Red (Washington)",..: 7
##  $ variety              : Factor w/ 707 levels "Abouriou","Agiorgitiko",..: 692 452 438 481 442 593
##  $ winery               : Factor w/ 16757 levels ":Nota Bene","1+1=3",..: 11641 12988 13054 14432 14
```

Com ja havíem comentat el segon dataset conté un major nombre de camps, per tant, haurem d'igualar-los per a poder unir-los.

```
# aprofitem per eliminar el primer camp que son les row.names
wine.t <- rbind(wine.150[,-1], wine.130[, -c(1,10,11,12)])
str(wine.t)
```

```
## 'data.frame':    280901 obs. of  10 variables:
##  $ country    : Factor w/ 50 levels "Albania","Argentina",..: 47 41 47 47 16 41 41 41 47 47 ...
##  $ description: Factor w/ 169430 levels ". Big, lively and very intense, this powerful Amarone opens
##  $ designation: Factor w/ 47239 levels "¡Adentro! Red",..: 17369 4413 25554 22403 14344 19205 23925 4
##  $ points     : int  96 96 96 96 95 95 95 95 95 95 ...
##  $ price      : num  235 110 90 65 66 73 65 110 65 60 ...
##  $ province   : Factor w/ 490 levels "Achaia","Aconcagua Costa",..: 52 275 52 283 315 275 275 275 283
##  $ region_1   : Factor w/ 1332 levels "Abruzzo","Adelaida District",..: 739 1071 529 1223 67 1071 107
##  $ region_2   : Factor w/ 18 levels "California Other",..: 8 NA 14 18 NA NA NA NA 18 14 ...
##  $ variety    : Factor w/ 756 levels "Agiorgitiko",..: 71 550 470 403 423 550 550 550 403 403 ...
##  $ winery     : Factor w/ 19186 levels ":Nota Bene","'37 Cellars",..: 7305 1240 9050 11038 5106 10203
```

2.1 Selecció de les dades d'interès a analitzar. Quins són els camps més rellevants per tal de respondre al problema?

Abans de procedir amb aquest apartat passarem a eliminar els elements repetits que pugui contenir el nostre dataset final.

```
wine.t <- wine.t[!duplicated(wine.t), ]
str(wine.t)
```

```
## 'data.frame':    170531 obs. of  10 variables:
##  $ country    : Factor w/ 50 levels "Albania","Argentina",..: 47 41 47 47 16 41 41 41 47 47 ...
##  $ description: Factor w/ 169430 levels ". Big, lively and very intense, this powerful Amarone opens
##  $ designation: Factor w/ 47239 levels "¡Adentro! Red",..: 17369 4413 25554 22403 14344 19205 23925 4
##  $ points     : int  96 96 96 96 95 95 95 95 95 95 ...
##  $ price      : num  235 110 90 65 66 73 65 110 65 60 ...
##  $ province   : Factor w/ 490 levels "Achaia","Aconcagua Costa",..: 52 275 52 283 315 275 275 275 283
##  $ region_1   : Factor w/ 1332 levels "Abruzzo","Adelaida District",..: 739 1071 529 1223 67 1071 107
##  $ region_2   : Factor w/ 18 levels "California Other",..: 8 NA 14 18 NA NA NA NA 18 14 ...
##  $ variety    : Factor w/ 756 levels "Agiorgitiko",..: 71 550 470 403 423 550 550 550 403 403 ...
##  $ winery     : Factor w/ 19186 levels ":Nota Bene","'37 Cellars",..: 7305 1240 9050 11038 5106 10203
```

Per a respondre la pregunta que plantegem en l'apartat 1 considerem que els camps rellevants són els següents:

country, points, price, province, variety, winery

```
# establim els índexs de les columnes a eliminar
indexs <- c(2,3,7,8)
wine.a <- wine.t[ ,-indexs]
dim(wine.a)
```

```
## [1] 170531      6
```

```
names(wine.a)
```

```
## [1] "country"  "points"   "price"    "province" "variety"  "winery"
```

2.2 Les dades contenen zeros o elements buits? I valors extrems? Com gestionaries cadascun d'aquests casos?

```
# mostrem les variables que contenen buits i la quantitat d'elements buits que tenen
vbles.buits <- names(wine.a)[!complete.cases(t(wine.a))]
sapply(wine.a[vbles.buits], function(x) sum(is.na(x)))
```

```
##  country    price province  variety
##       60    12841       60        1
```

Veiem que tant country com province contenen el mateix nombre d'elements buits i que winery no en conté cap, per tant, podem intentar completar les files a partir d'aquest camp complert. Tot i que, primer haurem de normalitzar els camps per homogeneitzar-los i evitar, així, errors d'escritura.

Pel que fa a preu tenim diverses opcions:

1. Intentar aconseguir els preus originals d'internet.
2. Mirar d'assignar valor a aquests camps per mitjà d'algun algorisme, per exemple kNN.
3. Eliminar les files amb camps buits.

Tot i que, la primera opció seria la idònea tenim un nombre massa elevat de valors faltants, per la qual cosa optarem per fer una anàlisi per cada una de les altres dues opcions i compararem el resultat.

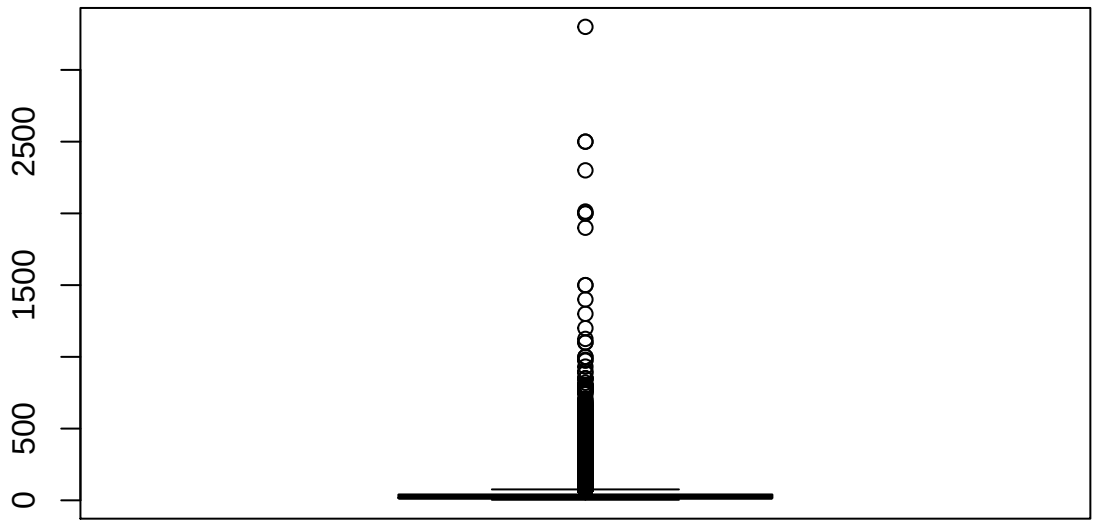
Per últim, veiem que només tenim una mostra sense variety, mirem la web original i la web de la bodega, però no obtenim més informació al respecte. Però si mirem el camp description d'aquest vi

'r wine.a[is.na(wine.a$variety), "description"]

veurem que es tracta d'un Petite Syrah, per tant, assignarem aquesta varietat al camp variety.

Pel que fa als valors extrems comprovarem si price, l'únic camp numèric que pot tenir-ne, en conté.

```r
boxplot(wine.a$price)
```



En el boxplot veiem que aquest camp té un gran nombre de valors extrems, però no ens dona massa informació al respecte.

Per tant, anem a veure'ls numèricament.

Tenim un total de 9568 mostres catalogades com a valors extrems amb un total de 342 valors diferents, que van des de'l 77 fins al 3300.

Tots aquests valors entren dins del rang de preus del vi. De totes formes donarem un cop d'ull a aquells que tinguin preus de 4 xifres per si hi hagués hagut un error a l'hora de ficar el preu amb els decimals.

```r
wine.a[which(wine.a[,3] >= 1000),]
```

```
##         country points price   province                   variety
## 10652   Austria     94  1100     Wachau           Grüner Veltliner
## 13319        US     91  2013 California                 Chardonnay
## 26297    France    100  1400  Champagne                 Chardonnay
## 34921    France     99  2300   Bordeaux   Bordeaux-style Red Blend
## 34923    France     98  1900   Bordeaux   Bordeaux-style Red Blend
## 34928    France     97  1100   Bordeaux   Bordeaux-style Red Blend
## 34940    France     96  1300   Bordeaux   Bordeaux-style Red Blend
## 34943    France     96  1200   Bordeaux   Bordeaux-style Red Blend
## 35532    France     94  1000   Bordeaux Bordeaux-style White Blend
## 166771   France     96  2500   Bordeaux   Bordeaux-style Red Blend
## 187462 Portugal     97  1000       Port                       Port
## 216283   France     97  2000   Bordeaux   Bordeaux-style Red Blend
```

```
## 231221   France     88  3300   Bordeaux   Bordeaux-style Red Blend
## 249311   France     96  2500   Burgundy                Pinot Noir
## 262684   France    100  1500   Bordeaux   Bordeaux-style Red Blend
## 262686   France    100  1500   Bordeaux   Bordeaux-style Red Blend
## 264495   France     96  2000   Burgundy                Pinot Noir
## 264512   France     94  1125   Burgundy                Pinot Noir
##                             winery
## 10652                 Emmerich Knoll
## 13319                          Blair
## 26297                           Krug
## 34921                 Château Latour
## 34923                Château Margaux
## 34928  Château La Mission Haut-Brion
## 34940      Château Mouton Rothschild
## 34943             Château Haut-Brion
## 35532  Château La Mission Haut-Brion
## 166771                Château Pétrus
## 187462              W. & J. Graham's
## 216283                Château Pétrus
## 231221      Château les Ormes Sorbet
## 249311 Domaine du Comte Liger-Belair
## 262684     Château Lafite Rothschild
## 262686          Château Cheval Blanc
## 264495 Domaine du Comte Liger-Belair
## 264512 Domaine du Comte Liger-Belair
```

Podem destacar dues coses d'aquest llistat.

La primera seria que molts dels vins més cars provenen de Bordeaux a França que sabem és una regió amb molta fama i, per tant, és habitual veure vins amb preus elevats.

La segona cosa a destacar és que tots aquests vins tenen més de 90 punts, l'excepció és el vi amb el preu més elevat.

Després de comprovar els preus a internet veiem que el preu del vi més car és una errada ja que podem trobar-lo per uns 30$, com podem comprovar a http://www.hachette-vins.com/guide-vins/les-vins/ch-les-ormes-sorbet-2013-2017/201706208/ o a https://www.chateau.fr/chateau-les-ormes-sorbet-2013-cbo-12x75cl-rouge.html.

Per tant, procedirem a arranjar el preu i a deixar-lo en 33 dollar, enlloc dels 3300$ que actualment té.

```
wine.a[which(wine.a["price"] == 3300), "price"] <- 33.0
```

Per a la resta de vins comprovem que el preu és correcte i donarem aquest punt per finalitzat.

Abans de continuar és convenient comprovar que les dades siguin del tipus corresponent i normalitzar/estandarditzar.

```
res <- sapply(wine.a, class)
kable(data.frame(variables=names(res), classe=as.vector(res)))
```

variables	classe
country	factor
points	integer
price	numeric
province	factor
variety	factor
winery	factor

variables	classe

L'únic tipus que haurem de canviar és el de points, ja que está representada per valors sencers i ens interesarà possarla com numeric per si tenim que fer la mitja o un altre càlcul respecte les variables factor.

```
wine.a$points <- as.numeric(wine.a$points)
res <- sapply(wine.a, class)
kable(data.frame(variables=names(res), classe=as.vector(res)))
```

variables	classe
country	factor
points	numeric
price	numeric
province	factor
variety	factor
winery	factor

Ara que ja tenim els tipus de variable correctament assignats procedim a normalitzar/estandarditzar les variables factor.

```
txtvar <- c("country", "province", "variety", "winery")
accents <- c("áéíóúàèìòùâêîôûäëïöüãõű")
noaccents <- c("aeiouaeiouaeiouaeiouaou")
puntua <- c("-_")
nopuntua <- ("  ")
f.origin = f.blancs = f.minus = f.accents = f.puntua = 0
j <- 1
f.puntua <- 1
for(i in txtvar) {
  f.origin[j] <- nlevels(wine.a[,i])
  # traiem espais en blanc al principi i final del text
  wine.a[,i] <- as.factor(trimws(wine.a[,i], "both"))
  f.blancs[j] <- nlevels(wine.a[,i])
  # possem tot el text en minúscula
  wine.a[,i] <- as.factor(tolower(wine.a[,i]))
  f.minus[j] <- nlevels(wine.a[,i])
  # eliminem accents
    wine.a[,i] <- as.factor(chartr(accents, noaccents, wine.a[,i]))
  f.accents[j] <- nlevels(wine.a[,i])
  wine.a[,i] <- as.factor((chartr(puntua, nopuntua, wine.a[,i])))
  wine.a[,i] <- as.factor(gsub("\\.", "", wine.a[,i]))
  wine.a[,i] <- as.factor(gsub("\\,", "", wine.a[,i]))
  wine.a[,i] <- as.factor(gsub("\\:", "", wine.a[,i]))
  wine.a[,i] <- as.factor(gsub("\\;", "", wine.a[,i]))
  wine.a[,i] <- as.factor(gsub("\\'", "", wine.a[,i]))
  f.puntua[j] <- nlevels(wine.a[,i])
  j <- j + 1
}
kable(data.frame(variables=txtvar, original=f.origin, sense.blancs=f.blancs, en.minuscules=f.minus, sens
```

variables	original	sense.blancs	en.minuscules	sense.accents	sense.puntuacions
country	50	50	50	50	50

variables	original	sense.blancs	en.minuscules	sense.accents	sense.puntuacions
province	490	490	490	490	490
variety	756	756	756	756	756
winery	19186	19186	19158	19119	19086

Un cop normalitzades les dades passarem a assignar el valor "petite shyrah" a l'exemple sense variety. Però, abans comprovarem que aquest valor existeixi per no crear un nou factor.

```
varietats <- grep("s[i|y]rah", wine.a$variety)
sort(unique(wine.a[varietats, "variety"]))
```

```
##  [1] cabernet sauvignon syrah cabernet syrah
##  [3] carignan syrah           carmenere syrah
##  [5] garnacha syrah           grenache syrah
##  [7] malbec syrah             merlot syrah
##  [9] monastrell syrah         mourvedre syrah
## [11] petite sirah             petite syrah
## [13] pinot noir syrah         sangiovese syrah
## [15] syrah                    syrah bonarda
## [17] syrah cabernet           syrah cabernet franc
## [19] syrah cabernet sauvignon syrah carignan
## [21] syrah grenache           syrah grenache viognier
## [23] syrah malbec             syrah merlot
## [25] syrah mourvedre          syrah petit verdot
## [27] syrah petite sirah       syrah tempranillo
## [29] syrah viognier           tannat syrah
## [31] tempranillo syrah
## 756 Levels: abouriou agiorgitiko aglianico aidani airen ... zweigelt
```

Veiem que aquesta varietat es presenta amb diferents noms. El mateix ens passarà amb altres varietats com podem comprovar en https://vivancoculturadevino.es/blog/2015/07/17/variedades-de-uva/ o en https://turismodevino.com/saber-de-vino/tipos-de-uva-en-el-vino/

Reassignarem algunes varietat, tot i que, per no allargar més la neteja (i la pràctica) només juntarem les que veiem són formes diferents d'escriure una mateixa varietat, com per exemple shirah i shyrah. Però deixarem aquelles que tot i ser la mateixa varietat rebin diferents noms en diferents Denominacions d'Origen (DO), com per exemple shiraz, que és el nom australià de la varietat syrah com podem veure a https://www.leaf.tv/articles/what-is-a-shiraz-wine/.

Pel que fa als vins formats per més d'una varietat mantindrem l'ordre, és a dir, si tenim les varietats syrah tempranillo (mostra [28]) la considerarem diferent a tempranillo syrah (mostra [31]), ja que indica que la varietat dominant en el vi és la primera i, per tant, el vi tindrà propietats/qualitats diferents.

Després d'examinar les varietats actuals durem a terme els canvis següents:

aragones, aragonez = aragones

assyrtico, assyrtiko = assyrtiko

carignan, carignane, carignano = carignan

chardonel, chardonelle = chardonel

durella, durello = durella

insolia, inzolia = inzolia

malagousia, malagouzia = malagouzia

malvasia, mavazija = malvasia

moscatel, muscatel = moscatel

moschofilero, moscofilero = moschofilero

muscadel, muscadelle = muscadel

muscat blanc a petits grains, muscat blanc a petit grain = muscat blanc a petit grain

muscat, muskat = muskat

petit verdot, petite verdot = petite verdot

pinot bianco, pinot blanc = pinot blanc

pinot nero, pinot noir = pinto noir

pinot grigio, pinot gris = pinot gris

sirah, syrah = syrah

tinta de toro, tinta del toro = tinta de toro

tinta fina, tinto fino = tinta fina

tinta del pais, tinto del pais = tinta del pais

tocai, tokay = tokay

vranac, vranec = vranac

```
o.variety <- c("aragonez", "assyrtico", "chardonelle", "durello", "insolia", "malagousia", "malvazija",

n.variety <- c("aragones", "assyrtiko", "chardonel", "durella", "inzolia", "malagouzia", "malvasia", "mu

# ho farem en dos vegades
# primer els que no presenten modficacions
for(n in 1:length(o.variety)) {
  wine.a[which(wine.a[,"variety"] == o.variety[n]), "variety"] <- as.factor(n.variety[n])
}
# segon els que sí en presenten
om.variety <- c("carignan[e|o]", "muscat", "pinot grigio", "sirah", "tocai")
nm.variety <- c("carignan", "muskat", "pinot gris", "syrah", "tokay")
for(n in 1:length(om.variety)) {
  indexs<-grep(om.variety[n], wine.a$variety)
  wine.a[indexs, "variety"] <- as.factor(nm.variety[n])
}
```

Per a variety el nombre de factors actual és 717. Ha disminuit en 39.

Per últim, podem comprovar visualment que per a country no hi ha errors.

```
sort(unique(factor(wine.a[,"country"])))
```

```
##  [1] albania                 argentina               armenia
##  [4] australia               austria                 bosnia and herzegovina
##  [7] brazil                  bulgaria                canada
## [10] chile                   china                   croatia
## [13] cyprus                  czech republic          egypt
## [16] england                 france                  georgia
## [19] germany                 greece                  hungary
## [22] india                   israel                  italy
```

```
## [25] japan                  lebanon               lithuania
## [28] luxembourg             macedonia             mexico
## [31] moldova                montenegro            morocco
## [34] new zealand            peru                  portugal
## [37] romania                serbia                slovakia
## [40] slovenia               south africa          south korea
## [43] spain                  switzerland           tunisia
## [46] turkey                 ukraine               uruguay
## [49] us                     us france
## 50 Levels: albania argentina armenia australia ... us france
```

I ara passariem a assignar valors als camps amb NA.

```
# assignem el valor petite syrah a la mostra sense variety
wine.a[is.na(wine.a$variety), "variety"] <- as.factor("petite syrah")

# per a country i province primer comprovarem que els NA corresponent a les mateixes mostres
wine.nacountry <- wine.a[is.na(wine.a$country), ]
wine.naprovince <- wine.a[is.na(wine.a$province), ]
identical(wine.nacountry, wine.naprovince)
```

```
## [1] TRUE
```

Efectivament són iguals així que els tractarem conjuntament

```
# obtenim les mostres que no tenen NA a country ni province
wine.nonacp <- wine.a[!is.na(wine.a$country), ]
# recorrem totes les mostres amb NA
for(i in 1:(nrow(wine.nacountry))) {
  # obtenim les mostres amb el mateix winery
  wine.prov <- wine.nonacp[wine.nonacp$winery == wine.nacountry$winery[i],]
  # si només tenim un province
  if(is.na(unique(wine.prov$province)[2])) {
    # l'assignem a la province NA
    wine.a[wine.a$winery == wine.nacountry$winery[i], "province"] <- unique(wine.prov$province)[1]
  }
  # si només tenim un country
  if(is.na(unique(wine.prov$country)[2])) {
    # l'assignem al country NA
    wine.a[wine.a$winery == wine.nacountry$winery[i], "country"] <- unique(wine.prov$country)[1]
  }
}

#comprovem novament els valors buits
vbles.buits <- names(wine.a)[!complete.cases(t(wine.a))]
sapply(wine.a[vbles.buits], function(x) sum(is.na(x)))
```

```
##  country    price province
##       27    12841       47
```

Veiem que ara tenim 27 country no identificats i 47 province. Passem a eliminar les mostres sense country (recordem que les mostres sense country eren les mateixes que sense province).

```
temp <- which(is.na(wine.a[,"country"]))
wine.a <- wine.a[-temp, ]
summary(wine.a)
```

```
##      country          points           price               province
```

```
##  us      :71754   Min.   : 80.00   Min.   :   4.00   california:49516
##  france  :27227   1st Qu.: 86.00   1st Qu.:  16.00   washington:11353
##  italy   :25016   Median : 88.00   Median :  25.00   bordeaux  : 7736
##  spain   : 8833   Mean   : 88.24   Mean   :  34.64   tuscany   : 7462
##  portugal: 6914   3rd Qu.: 90.00   3rd Qu.:  40.00   oregon    : 6106
##  chile   : 6138   Max.   :100.00   Max.   :2500.00   (Other)   :88311
##  (Other) :24622                    NA's   :12841     NA's      :   20
##                       variety                         winery
##  pinot noir                  :17093   williams selyem     :   306
##  chardonnay                  :15709   testarossa          :   286
##  cabernet sauvignon          :13359   chateau ste michelle:   269
##  red blend                   :11276   dfj vinhos          :   268
##  bordeaux style red blend: 9035   wines & winemakers  :   249
##  syrah                       : 6934   louis latour        :   245
##  (Other)                     :97098   (Other)             :168881
```

3 Anàlisi de les dades.

3.1 Selecció dels grups de dades que es volen analitzar/comparar.

Per a respondre la pregunta de l'apartat 1 ens quedarem amb els camps country, points, price i variety. Descartem winery ja que aquesta rarament es troba a les ampolles i province conté un major nombre de NA que country, a més a més, considerem més fàcil (especialment a la Xina on els vins es presenten ordenats per països) saber de quin pais és una ampolla de vi.

Per obtenir la relació qualitat/preu emprarem points/price les quals relaciorem amb province i variety per mirar de respondre la pregunta plantejada. Per tant, un altre grup de dades que podem analitzar son les varietats per pais.

Uns altres grups que podriem analitzar seria puntuació i preus per pais o qualitat per pais.

Generem el data frame final.

```
# obtenim el data frame final amb camps buits a price
wine.nafinal <- wine.a[, -c(4,6)]
dim(wine.nafinal)
```

```
## [1] 170504      4
```

```
# imputació de valors basada en kNN i distància de Gover
# descartem aquesta opció ja que el temps que tarda és massa gran
# ---- wine.knnfinal <- kNN(wine.nafinal) ----- OPCIÓ DESCARTADA

# generem un data frame sense els valors NA de price
temp <- which(is.na(wine.nafinal[,"price"]))
wine.final <- wine.nafinal[-temp, ]
summary(wine.final)
```

```
##      country          points           price
##  us      :71403   Min.   : 80.00   Min.   :   4.00
##  italy   :21263   1st Qu.: 86.00   1st Qu.:  16.00
##  france  :21006   Median : 88.00   Median :  25.00
##  spain   : 8714   Mean   : 88.19   Mean   :  34.64
##  chile   : 6072   3rd Qu.: 90.00   3rd Qu.:  40.00
##  portugal: 5778   Max.   :100.00   Max.   :2500.00
```

```
##  (Other) :23427
##                variety
##  pinot noir         :16401
##  chardonnay         :14878
##  cabernet sauvignon:13221
##  red blend          :10648
##  syrah              : 6817
##  sauvignon blanc    : 6636
##  (Other)            :89062
```

El resum, entre altres coses, ens mostra com no tenim camps buits.

3.2 Comprovació de la normalitat i homogeneïtat de la variància. Si és necessari (i possible), aplicar transformacions que normalitzin les dades.

```
# comprovem la normalitat de la variància amb el test de Levene per a la qualitat/preu
with(wine.final, tapply(price, country, var, na.rm=TRUE))
```

```
##               albania             argentina               armenia
##                    NA            505.1083556             0.5000000
##             australia                austria bosnia and herzegovina
##          1917.3854017            707.3120165             0.3333333
##                brazil               bulgaria                canada
##           118.0045249             87.8807947           437.9626697
##                 chile                 china               croatia
##           429.6291564            100.3333333           158.3416361
##                cyprus          czech republic                 egypt
##            12.5087719             93.9780220                    NA
##               england                france               georgia
##           233.5581921          5066.9898808            59.2652821
##               germany                greece               hungary
##          3524.8651971            147.7135387          4948.6127622
##                 india                israel                 italy
##            16.0277778            353.4503545          1401.0666280
##                 japan                lebanon             lithuania
##                    NA            256.1858304             0.0000000
##              luxembourg              macedonia                mexico
##           113.8392857              9.5065359           265.7847293
##               moldova             montenegro                morocco
##            76.0116200                     NA            48.5538462
##            new zealand                   peru              portugal
##           236.5906883            186.7291667          1554.9805044
##               romania                 serbia              slovakia
##           716.6802316             80.8909091             0.5000000
##              slovenia           south africa            south korea
##           216.0224660            327.2872398            12.5000000
##                 spain            switzerland               tunisia
##          1189.8543292           4237.4761905                    NA
##                turkey                ukraine               uruguay
##           239.3993585              6.1102941           315.9985098
##                    us              us france
##           697.5766916                     NA
```

```r
leveneTest(price~country, data = wine.final, center = "median")
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##           Df F value    Pr(>F)
## group     47  56.752 < 2.2e-16 ***
##       157615
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Com ja haviem suposat no tenim homogeneïtat, ja que el p-valor és menor a 0.05. Això és deu al preu elevat d'alguns vins, recordem que tenim més de 9500 valors extrems.

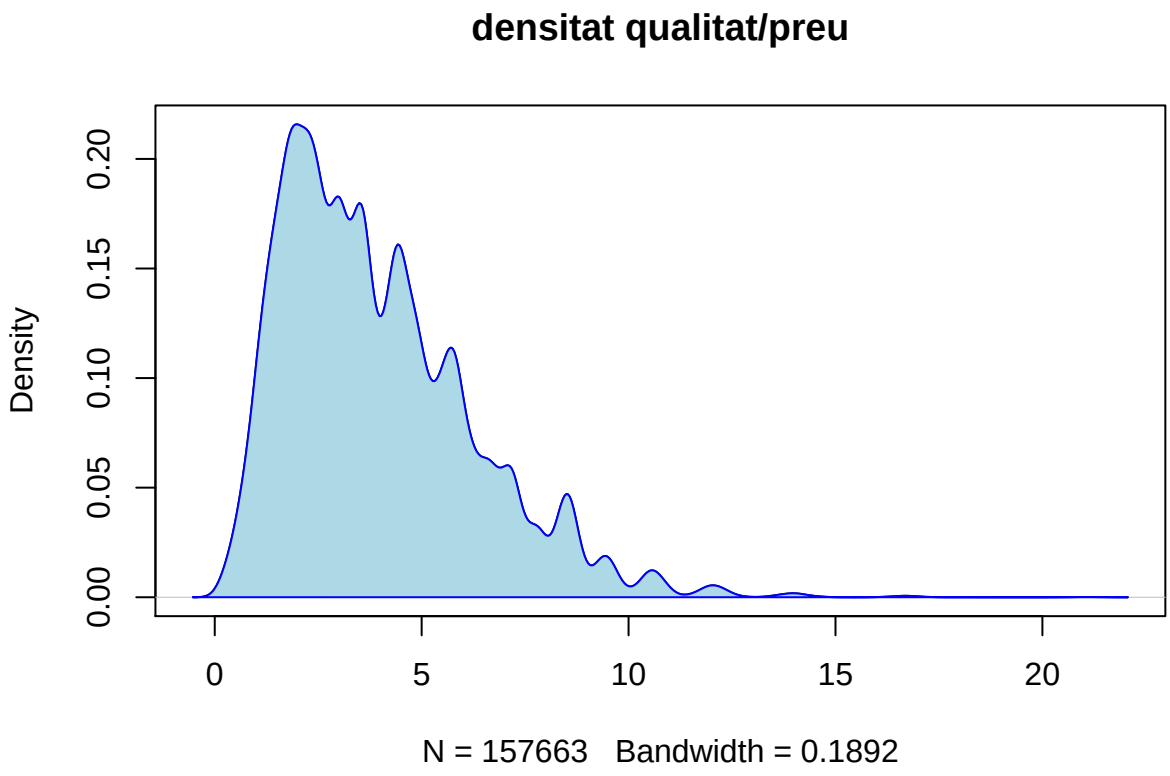
No normalitzarem, atés que el que estem buscant és una relació qualitat-preu i, per tant, volem mantenir la relació actual.

3.3 Aplicació de proves estadístiques (tantes com sigui possible) per comparar els grups de dades.

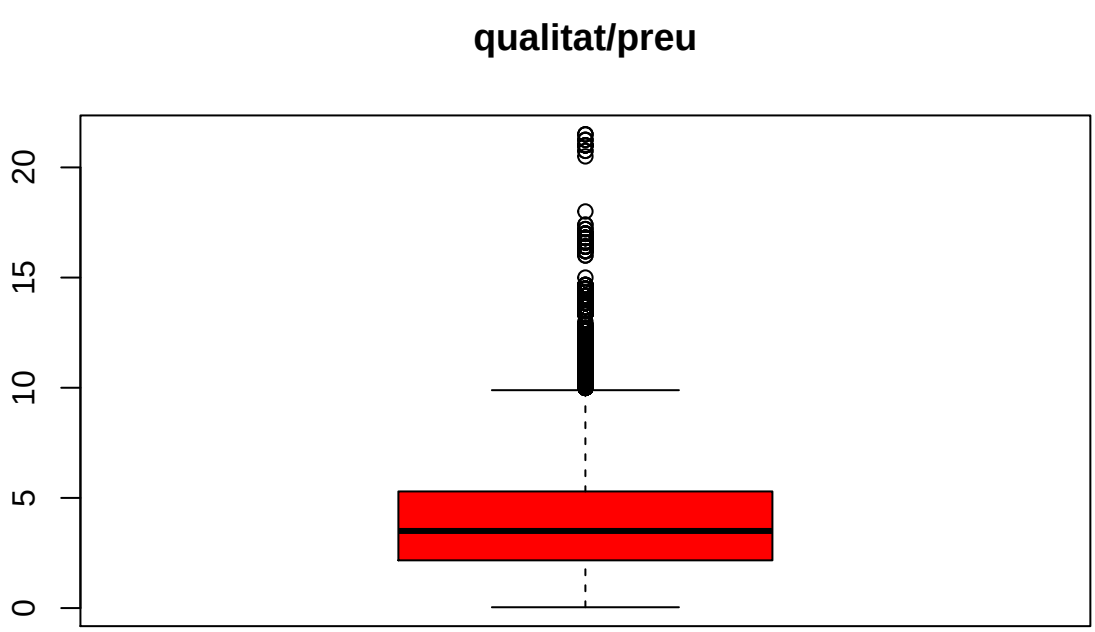
4 Representació dels resultats a partir de taules i gràfiques.

```r
# en el nostre cas definirem la relació qualitat-preu com el nombre de punts dividit entre el preu
qual.preu <- wine.final$points/wine.final$price

plot(density(qual.preu), main="densitat qualitat/preu")
polygon(density(qual.preu), col="light blue", border="blue")
```




```
boxplot(qual.preu, main="qualitat/preu", col="red")
```




```
summary(qual.preu)
```

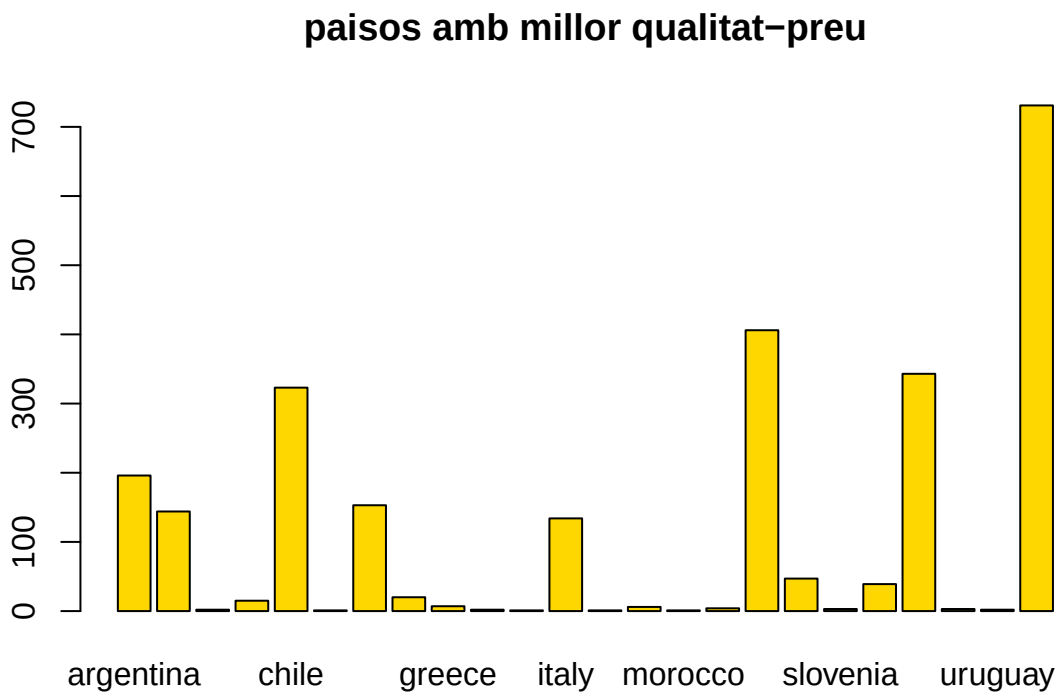
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0384  2.1670  3.5000  3.9410  5.2940 21.5000
```

A partir d'aquestes gràfiques podem veure que els vins amb major qualitat-preu són aquells que és consideren valors extrems.

Relacionem ara aquests valors amb els països d'origen i el tipus de raïm.

```
#qual.preu.extrem <- boxplot.stats(qual.preu)$out
# obtenim els indexs dels valors extrems
qual.preu.iextrem  <- which(qual.preu  %in%  boxplot.stats(qual.preu)$out)
# i obtenim els paisos als que pertanyen aquests vins
country.extrem <- droplevels(wine.final[qual.preu.iextrem, "country"])

# els mostrem gràfica i numèricament
plot(country.extrem, main= "paisos amb millor qualitat-preu", col="gold")
```




```
sort(table(country.extrem), decreasing = TRUE)
```

```
## country.extrem
##           us     portugal        spain        chile    argentina
##          731          406          343          323          196
##       france    australia        italy      romania south africa
##          153          144          134           47           39
##      germany     bulgaria       greece      moldova  new zealand
##           20           15            7            6            4
##     slovenia      ukraine      austria      hungary      uruguay
##            3            3            2            2            2
##        china       israel       mexico      morocco
##            1            1            1            1
```

Ens queda ara obtenir les varietats de raim que ens proporcionaran millor qualitat-preu.

```
# primerament i igual que amb països obtenim les varietats d'aquests vins
variety.extrem <- droplevels(wine.final[qual.preu.iextrem, "variety"])

# els mostrem gràfica i numèricament
plot(variety.extrem, main= "varietats amb millor qualitat-preu", col="dark green")
```

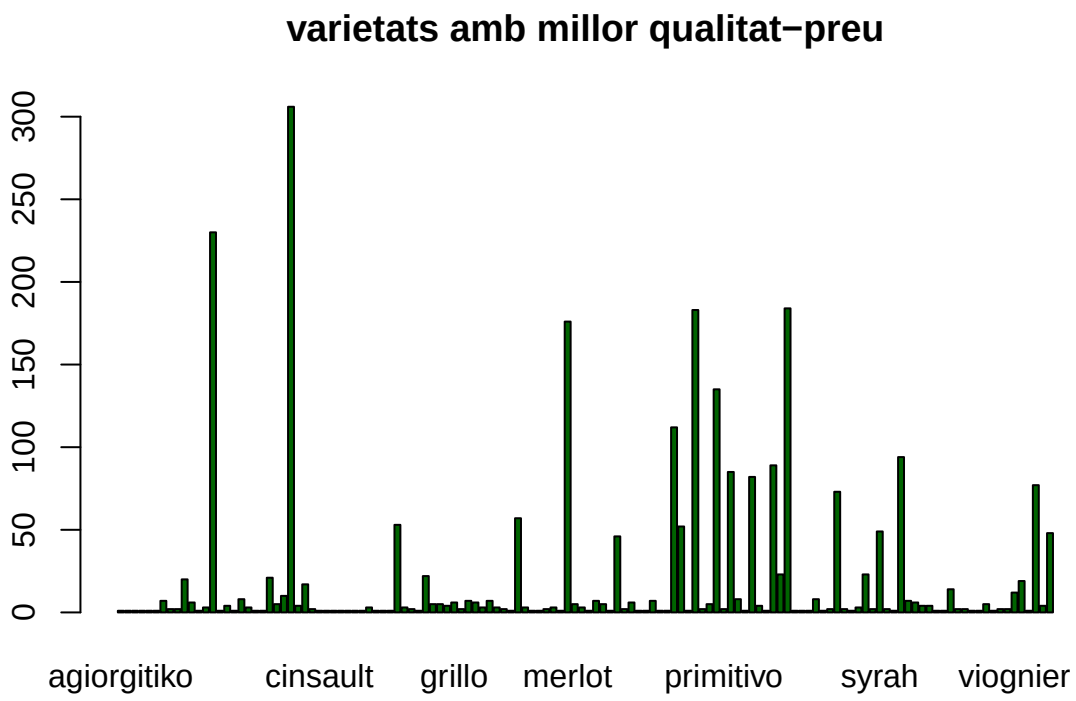
```
## Warning in axis(if (horiz) 2 else 1, at = at.l, labels = names.arg, lty =
## axis.lty, : conversion failure on 'feteasc\u0103 regal\u0103' in 'mbcsToSbcs': dot
## substituted for <c7>
```

```
## Warning in axis(if (horiz) 2 else 1, at = at.l, labels = names.arg, lty =
## axis.lty, : conversion failure on 'feteasc\u0103 regal\u0103' in 'mbcsToSbcs': dot
```

```
## substituted for <8e>
```

```
## Warning in axis(if (horiz) 2 else 1, at = at.l, labels = names.arg, lty =
## axis.lty, : conversion failure on 'fetească regală' in 'mbcsToSbcs': dot
## substituted for <c7>
```

```
## Warning in axis(if (horiz) 2 else 1, at = at.l, labels = names.arg, lty =
## axis.lty, : conversion failure on 'fetească regală' in 'mbcsToSbcs': dot
## substituted for <8e>
```




```
sort(table(variety.extrem), decreasing = TRUE)
```

```
## variety.extrem
##                       chardonnay             cabernet sauvignon
##                              306                            230
##                  sauvignon blanc                portuguese red
##                              184                            183
##                           merlot              portuguese white
##                              176                            135
##                       pinot gris                   tempranillo
##                              112                             94
##                             rose                     red blend
##                               89                             85
##                         riesling                   white blend
##                               82                             77
##                           shiraz                        malbec
##                               73                             57
##                         garnacha                    pinot noir
```

```
##                                53                                52
##                             syrah                          zinfandel
##                                49                                48
##                           moscato                         sangiovese
##                                46                                23
##                    sparkling blend                   gewurztraminer
##                                23                                22
##                          carmenere              bordeaux style red blend
##                                21                                20
##                             viura                      chenin blanc
##                                19                                17
##                         torrontes                          viognier
##                                14                                12
##                    champagne blend            cabernet sauvignon merlot
##                                10                                 8
##              rhone style red blend                  semillon chardonnay
##                                 8                                 8
##                            arinto                johannisberg riesling
##                                 7                                 7
##                          loureiro                       monastrell
##                                 7                                 7
##                       nero davola                  tempranillo blend
##                                 7                                 7
##         bordeaux style white blend                           grillo
##                                 6                                 6
##                          lambrusco                          muskat
##                                 6                                 6
##     tempranillo cabernet sauvignon                         castelao
##                                 6                                 5
##                             glera                          grenache
##                                 5                                 5
##                    merlot cabernet                    montepulciano
##                                 5                                 5
##               portuguese sparkling                           verdejo
##                                 5                                 5
##        cabernet sauvignon carmenere                 chardonnay viognier
##                                 4                                 4
##                     grenache syrah                           rosado
##                                 4                                 4
##               tempranillo garnacha                 tempranillo merlot
##                                 4                                 4
##                     white riesling                    cabernet merlot
##                                 4                                 3
##          cabernet sauvignon shiraz                         fume blanc
##                                 3                                 3
##                    garnacha blanca                          lemberger
##                                 3                                 3
##                            macabeo                    malbec bonarda
##                                 3                                 3
##                             mencia             merlot cabernet sauvignon
##                                 3                                 3
##                 shiraz tempranillo                             bical
##                                 3                                 2
##                              bobal               chenin blanc chardonnay
```

```
##                                 2                                 2
##                     garnacha syrah                          inzolia
##                                 2                                 2
##                   macabeo moscatel                            melon
##                                 2                                 2
##                     muller thurgau                  portuguese rose
##                                 2                                 2
##                           primitivo                           sherry
##                                 2                                 2
##          shiraz cabernet sauvignon                         symphony
##                                 2                                 2
##                     syrah cabernet                 touriga nacional
##                                 2                                 2
## touriga nacional cabernet sauvignon                       vermentino
##                                 2                                 2
##                           vernaccia                      agiorgitiko
##                                 2                                 1
##                           aglianico                           airen
##                                 1                                 1
##                             albana                alicante bouschet
##                                 1                                 1
##                           aragones                  cabernet blend
##                                 1                                 1
##  cabernet sauvignon and tinta roriz        cabernet sauvignon malbec
##                                 1                                 1
##           cabernet sauvignon syrah                         carineña
##                                 1                                 1
##                           cinsault                       colombard
##                                 1                                 1
##        corvina rondinella molinara                       dornfelder
##                                 1                                 1
##                       fernao pires                         feteasca
##                                 1                                 1
##                   fetească regală                             g s m
##                                 1                                 1
##                             gamay                        garganega
##                                 1                                 1
##               garnacha tempranillo             malagouzia chardonnay
##                                 1                                 1
##                     malbec syrah                malbec tempranillo
##                                 1                                 1
##                         meritage                     merlot malbec
##                                 1                                 1
##                     morio muskat                    muskat ottonel
##                                 1                                 1
##                         negroamaro                          pigato
##                                 1                                 1
##                       pinot blanc                         pinotage
##                                 1                                 1
##            rhone style white blend                          rosato
##                                 1                                 1
##           sauvignon blanc semillon                         savatiano
##                                 1                                 1
##                           semillon           semillon sauvignon blanc
```

```
##                                 1                                 1
##                     shiraz pinotage                     syrah grenache
##                                 1                                 1
##                       tinta de toro                       tinta roriz
##                                 1                                 1
##                           trajadura                         valdiguie
##                                 1                                 1
##                       verdejo viura                   viura chardonnay
##                                 1                                 1
```

I, finalment, mostrem una taula amb la relació de les varietats segons els seu país.

```
# mostrem-les ara conjuntament
table(variety.extrem, country.extrem)
```

```
##                                     country.extrem
## variety.extrem                       argentina australia austria bulgaria
##   agiorgitiko                                0         0       0        0
##   aglianico                                  0         0       0        0
##   airen                                      0         0       0        0
##   albana                                     0         0       0        0
##   alicante bouschet                          0         0       0        0
##   aragones                                   0         0       0        0
##   arinto                                     0         0       0        0
##   bical                                      0         0       0        0
##   bobal                                      0         0       0        0
##   bordeaux style red blend                   0         0       0        0
##   bordeaux style white blend                 0         0       0        0
##   cabernet blend                             0         0       0        0
##   cabernet merlot                            0         0       0        0
##   cabernet sauvignon                        28        14       0        4
##   cabernet sauvignon and tinta roriz         0         0       0        0
##   cabernet sauvignon carmenere               0         0       0        0
##   cabernet sauvignon malbec                  1         0       0        0
##   cabernet sauvignon merlot                  1         1       1        0
##   cabernet sauvignon shiraz                  0         3       0        0
##   cabernet sauvignon syrah                   0         0       0        0
##   carineña                                   0         0       0        0
##   carmenere                                  0         0       0        0
##   castelao                                   0         0       0        0
##   champagne blend                            0         0       0        0
##   chardonnay                                31        53       0        1
##   chardonnay viognier                        2         0       0        0
##   chenin blanc                               0         0       0        0
##   chenin blanc chardonnay                    0         0       0        0
##   cinsault                                   0         0       0        0
##   colombard                                  0         0       0        0
##   corvina rondinella molinara                0         0       0        0
##   dornfelder                                 0         0       0        0
##   fernao pires                               0         0       0        0
##   feteasca                                   0         0       0        0
##   fetească regală                            0         0       0        0
##   fume blanc                                 0         0       0        0
##   g s m                                      0         1       0        0
##   gamay                                      0         0       0        0
```

##	garganega	0	0	0	0
##	garnacha	0	0	0	0
##	garnacha blanca	0	0	0	0
##	garnacha syrah	0	0	0	0
##	garnacha tempranillo	0	0	0	0
##	gewurztraminer	0	0	0	0
##	glera	0	0	0	0
##	grenache	0	0	0	0
##	grenache syrah	0	0	0	0
##	grillo	0	0	0	0
##	inzolia	0	0	0	0
##	johannisberg riesling	0	0	0	0
##	lambrusco	0	0	0	0
##	lemberger	0	0	0	0
##	loureiro	0	0	0	0
##	macabeo	0	0	0	0
##	macabeo moscatel	0	0	0	0
##	malagouzia chardonnay	0	0	0	0
##	malbec	48	1	0	0
##	malbec bonarda	3	0	0	0
##	malbec syrah	1	0	0	0
##	malbec tempranillo	1	0	0	0
##	melon	0	0	0	0
##	mencia	0	0	0	0
##	meritage	0	0	0	1
##	merlot	15	13	0	4
##	merlot cabernet	0	0	0	0
##	merlot cabernet sauvignon	0	0	0	0
##	merlot malbec	1	0	0	0
##	monastrell	0	0	0	0
##	montepulciano	0	0	0	0
##	morio muskat	0	0	0	0
##	moscato	6	3	0	0
##	muller thurgau	0	0	0	0
##	muskat	0	1	0	1
##	muskat ottonel	0	0	1	0
##	negroamaro	0	0	0	0
##	nero davola	0	0	0	0
##	pigato	0	0	0	0
##	pinot blanc	0	0	0	0
##	pinot gris	8	3	0	0
##	pinot noir	1	2	0	0
##	pinotage	0	0	0	0
##	portuguese red	0	0	0	0
##	portuguese rose	0	0	0	0
##	portuguese sparkling	0	0	0	0
##	portuguese white	0	0	0	0
##	primitivo	0	0	0	0
##	red blend	5	1	0	0
##	rhone style red blend	0	1	0	0
##	rhone style white blend	0	0	0	0
##	riesling	0	8	0	1
##	rosado	0	0	0	0
##	rosato	0	0	0	0

```
##   rose                                  2       0       0       0
##   sangiovese                            0       0       0       0
##   sauvignon blanc                       5       5       0       2
##   sauvignon blanc semillon              1       0       0       0
##   savatiano                             0       0       0       0
##   semillon                              0       0       0       0
##   semillon chardonnay                   0       5       0       0
##   semillon sauvignon blanc              0       1       0       0
##   sherry                                0       0       0       0
##   shiraz                               11      23       0       0
##   shiraz cabernet sauvignon             0       2       0       0
##   shiraz pinotage                       0       0       0       0
##   shiraz tempranillo                    0       0       0       0
##   sparkling blend                       1       1       0       0
##   symphony                              0       0       0       0
##   syrah                                 1       0       0       0
##   syrah cabernet                        0       0       0       0
##   syrah grenache                        0       0       0       0
##   tempranillo                           5       0       0       0
##   tempranillo blend                     0       0       0       0
##   tempranillo cabernet sauvignon        0       0       0       0
##   tempranillo garnacha                  0       0       0       0
##   tempranillo merlot                    0       0       0       0
##   tinta de toro                         0       0       0       0
##   tinta roriz                           0       0       0       0
##   torrontes                            14       0       0       0
##   touriga nacional                      0       0       0       0
##   touriga nacional cabernet sauvignon   0       0       0       0
##   trajadura                             0       0       0       0
##   valdiguie                             0       0       0       0
##   verdejo                               0       0       0       0
##   verdejo viura                         0       0       0       0
##   vermentino                            0       0       0       0
##   vernaccia                             0       0       0       0
##   viognier                              3       0       0       0
##   viura                                 0       0       0       0
##   viura chardonnay                      0       0       0       0
##   white blend                           1       2       0       1
##   white riesling                        0       0       0       0
##   zinfandel                             0       0       0       0
##                                country.extrem
## variety.extrem                  chile china france germany greece
##   agiorgitiko                       0     0      0       0      1
##   aglianico                         0     0      0       0      0
##   airen                             0     0      0       0      0
##   albana                            0     0      0       0      0
##   alicante bouschet                 0     0      0       0      0
##   aragones                          0     0      0       0      0
##   arinto                            0     0      0       0      0
##   bical                             0     0      0       0      0
##   bobal                             0     0      0       0      0
##   bordeaux style red blend          0     0     18       0      0
##   bordeaux style white blend        0     0      6       0      0
##   cabernet blend                    0     0      0       0      0
```

```
##   cabernet merlot                      0     0      0       0      0
##   cabernet sauvignon                  73     0     12       0      0
##   cabernet sauvignon and tinta roriz   0     0      0       0      0
##   cabernet sauvignon carmenere         4     0      0       0      0
##   cabernet sauvignon malbec            0     0      0       0      0
##   cabernet sauvignon merlot            0     0      1       0      0
##   cabernet sauvignon shiraz            0     0      0       0      0
##   cabernet sauvignon syrah             1     0      0       0      0
##   carineña                             0     0      0       0      0
##   carmenere                           21     0      0       0      0
##   castelao                             0     0      0       0      0
##   champagne blend                      0     0      0       0      0
##   chardonnay                          61     0     17       0      0
##   chardonnay viognier                  1     0      1       0      0
##   chenin blanc                         0     0      0       0      0
##   chenin blanc chardonnay              0     0      0       0      0
##   cinsault                             0     0      1       0      0
##   colombard                            0     0      1       0      0
##   corvina rondinella molinara          0     0      0       0      0
##   dornfelder                           0     0      0       1      0
##   fernao pires                         0     0      0       0      0
##   feteasca                             0     0      0       0      0
##   fetească regală                      0     0      0       0      0
##   fume blanc                           0     0      0       0      0
##   g s m                                0     0      0       0      0
##   gamay                                0     0      1       0      0
##   garganega                            0     0      0       0      0
##   garnacha                             0     0      0       0      0
##   garnacha blanca                      0     0      0       0      0
##   garnacha syrah                       0     0      0       0      0
##   garnacha tempranillo                 0     0      0       0      0
##   gewurztraminer                       2     0      0       0      0
##   glera                                0     0      0       0      0
##   grenache                             0     0      0       0      0
##   grenache syrah                       0     0      1       0      0
##   grillo                               0     0      0       0      0
##   inzolia                              0     0      0       0      0
##   johannisberg riesling                0     0      0       0      0
##   lambrusco                            0     0      0       0      0
##   lemberger                            0     0      0       0      0
##   loureiro                             0     0      0       0      0
##   macabeo                              0     0      0       0      0
##   macabeo moscatel                     0     0      0       0      0
##   malagouzia chardonnay                0     0      0       0      1
##   malbec                               7     0      1       0      0
##   malbec bonarda                       0     0      0       0      0
##   malbec syrah                         0     0      0       0      0
##   malbec tempranillo                   0     0      0       0      0
##   melon                                0     0      2       0      0
##   mencia                               0     0      0       0      0
##   meritage                             0     0      0       0      0
##   merlot                              42     0     11       0      0
##   merlot cabernet                      0     0      0       0      0
##   merlot cabernet sauvignon            0     0      0       0      0
```

```
##   merlot malbec                          0     0     0     0     0
##   monastrell                             0     0     0     0     0
##   montepulciano                          0     0     0     0     0
##   morio muskat                           0     0     0     0     0
##   moscato                                4     0     1     0     0
##   muller thurgau                         0     0     0     1     0
##   muskat                                 0     0     0     0     0
##   muskat ottonel                         0     0     0     0     0
##   negroamaro                             0     0     0     0     0
##   nero davola                            0     0     0     0     0
##   pigato                                 0     0     0     0     0
##   pinot blanc                            0     0     1     0     0
##   pinot gris                             1     0     0     1     0
##   pinot noir                            11     0     5     0     0
##   pinotage                               0     0     0     0     0
##   portuguese red                         0     0     0     0     0
##   portuguese rose                        0     0     0     0     0
##   portuguese sparkling                   0     0     0     0     0
##   portuguese white                       0     0     0     0     0
##   primitivo                              0     0     0     0     0
##   red blend                              2     0     9     0     2
##   rhone style red blend                  0     0     7     0     0
##   rhone style white blend                0     0     1     0     0
##   riesling                               2     0     0    17     0
##   rosado                                 0     0     0     0     0
##   rosato                                 0     0     0     0     0
##   rose                                   2     0     7     0     0
##   sangiovese                             0     0     0     0     0
##   sauvignon blanc                       72     0    14     0     0
##   sauvignon blanc semillon               0     0     0     0     0
##   savatiano                              0     0     0     0     1
##   semillon                               0     0     0     0     0
##   semillon chardonnay                    0     0     0     0     0
##   semillon sauvignon blanc               0     0     0     0     0
##   sherry                                 0     0     0     0     0
##   shiraz                                 2     0     3     0     0
##   shiraz cabernet sauvignon              0     0     0     0     0
##   shiraz pinotage                        0     0     0     0     0
##   shiraz tempranillo                     0     0     0     0     0
##   sparkling blend                        0     0    10     0     0
##   symphony                               0     0     0     0     0
##   syrah                                 11     0     9     0     0
##   syrah cabernet                         2     0     0     0     0
##   syrah grenache                         0     0     0     0     0
##   tempranillo                            0     0     0     0     0
##   tempranillo blend                      0     0     0     0     0
##   tempranillo cabernet sauvignon         0     0     0     0     0
##   tempranillo garnacha                   0     0     0     0     0
##   tempranillo merlot                     0     0     0     0     0
##   tinta de toro                          0     0     0     0     0
##   tinta roriz                            0     0     0     0     0
##   torrontes                              0     0     0     0     0
##   touriga nacional                       0     0     0     0     0
##   touriga nacional cabernet sauvignon    0     0     0     0     0
```

```
##   trajadura                            0     0      0       0      0
##   valdiguie                            0     0      0       0      0
##   verdejo                              0     0      0       0      0
##   verdejo viura                        0     0      0       0      0
##   vermentino                           0     0      0       0      0
##   vernaccia                            0     0      0       0      0
##   viognier                             0     0      0       0      0
##   viura                                0     0      0       0      0
##   viura chardonnay                     0     0      0       0      0
##   white blend                          2     1     13       0      2
##   white riesling                       0     0      0       0      0
##   zinfandel                            0     0      0       0      0
##                                  country.extrem
## variety.extrem                    hungary israel italy mexico moldova
##   agiorgitiko                           0      0     0      0       0
##   aglianico                             0      0     1      0       0
##   airen                                 0      0     0      0       0
##   albana                                0      0     1      0       0
##   alicante bouschet                     0      0     0      0       0
##   aragones                              0      0     0      0       0
##   arinto                                0      0     0      0       0
##   bical                                 0      0     0      0       0
##   bobal                                 0      0     0      0       0
##   bordeaux style red blend              0      0     0      0       0
##   bordeaux style white blend            0      0     0      0       0
##   cabernet blend                        0      0     0      0       0
##   cabernet merlot                       0      0     0      0       0
##   cabernet sauvignon                    0      0     1      0       1
##   cabernet sauvignon and tinta roriz    0      0     0      0       0
##   cabernet sauvignon carmenere          0      0     0      0       0
##   cabernet sauvignon malbec             0      0     0      0       0
##   cabernet sauvignon merlot             0      0     0      0       1
##   cabernet sauvignon shiraz             0      0     0      0       0
##   cabernet sauvignon syrah              0      0     0      0       0
##   carineña                              0      0     0      0       0
##   carmenere                             0      0     0      0       0
##   castelao                              0      0     0      0       0
##   champagne blend                       0      0     0      0       0
##   chardonnay                            0      1     7      0       0
##   chardonnay viognier                   0      0     0      0       0
##   chenin blanc                          0      0     0      0       0
##   chenin blanc chardonnay               0      0     0      0       0
##   cinsault                              0      0     0      0       0
##   colombard                             0      0     0      0       0
##   corvina rondinella molinara           0      0     1      0       0
##   dornfelder                            0      0     0      0       0
##   fernao pires                          0      0     0      0       0
##   feteasca                              0      0     0      0       0
##   fetească regală                       0      0     0      0       0
##   fume blanc                            0      0     0      0       0
##   g s m                                 0      0     0      0       0
##   gamay                                 0      0     0      0       0
##   garganega                             0      0     1      0       0
##   garnacha                              0      0     0      0       0
```

##	garnacha blanca	0	0	0	0	0
##	garnacha syrah	0	0	0	0	0
##	garnacha tempranillo	0	0	0	0	0
##	gewurztraminer	0	0	0	0	0
##	glera	0	0	5	0	0
##	grenache	0	0	0	0	0
##	grenache syrah	0	0	0	0	0
##	grillo	0	0	6	0	0
##	inzolia	0	0	2	0	0
##	johannisberg riesling	0	0	0	0	0
##	lambrusco	0	0	6	0	0
##	lemberger	0	0	0	0	0
##	loureiro	0	0	0	0	0
##	macabeo	0	0	0	0	0
##	macabeo moscatel	0	0	0	0	0
##	malagouzia chardonnay	0	0	0	0	0
##	malbec	0	0	0	0	0
##	malbec bonarda	0	0	0	0	0
##	malbec syrah	0	0	0	0	0
##	malbec tempranillo	0	0	0	0	0
##	melon	0	0	0	0	0
##	mencia	0	0	0	0	0
##	meritage	0	0	0	0	0
##	merlot	0	0	3	0	0
##	merlot cabernet	0	0	0	0	0
##	merlot cabernet sauvignon	0	0	0	0	0
##	merlot malbec	0	0	0	0	0
##	monastrell	0	0	0	0	0
##	montepulciano	0	0	5	0	0
##	morio muskat	0	0	0	0	0
##	moscato	0	0	5	0	0
##	muller thurgau	0	0	0	0	0
##	muskat	0	0	0	0	0
##	muskat ottonel	0	0	0	0	0
##	negroamaro	0	0	1	0	0
##	nero davola	0	0	7	0	0
##	pigato	0	0	1	0	0
##	pinot blanc	0	0	0	0	0
##	pinot gris	1	0	29	0	1
##	pinot noir	0	0	0	0	0
##	pinotage	0	0	0	0	0
##	portuguese red	0	0	0	0	0
##	portuguese rose	0	0	0	0	0
##	portuguese sparkling	0	0	0	0	0
##	portuguese white	0	0	0	0	0
##	primitivo	0	0	2	0	0
##	red blend	1	0	13	0	2
##	rhone style red blend	0	0	0	0	0
##	rhone style white blend	0	0	0	0	0
##	riesling	0	0	0	0	0
##	rosado	0	0	0	0	0
##	rosato	0	0	1	0	0
##	rose	0	0	2	0	0
##	sangiovese	0	0	23	0	0

```
##   sauvignon blanc                          0      0     0      0       1
##   sauvignon blanc semillon                 0      0     0      0       0
##   savatiano                                0      0     0      0       0
##   semillon                                 0      0     0      0       0
##   semillon chardonnay                      0      0     0      0       0
##   semillon sauvignon blanc                 0      0     0      0       0
##   sherry                                   0      0     0      0       0
##   shiraz                                   0      0     0      0       0
##   shiraz cabernet sauvignon                0      0     0      0       0
##   shiraz pinotage                          0      0     0      0       0
##   shiraz tempranillo                       0      0     0      0       0
##   sparkling blend                          0      0     0      0       0
##   symphony                                 0      0     0      0       0
##   syrah                                    0      0     3      1       0
##   syrah cabernet                           0      0     0      0       0
##   syrah grenache                           0      0     0      0       0
##   tempranillo                              0      0     0      0       0
##   tempranillo blend                        0      0     0      0       0
##   tempranillo cabernet sauvignon           0      0     0      0       0
##   tempranillo garnacha                     0      0     0      0       0
##   tempranillo merlot                       0      0     0      0       0
##   tinta de toro                            0      0     0      0       0
##   tinta roriz                              0      0     0      0       0
##   torrontes                                0      0     0      0       0
##   touriga nacional                         0      0     0      0       0
##   touriga nacional cabernet sauvignon      0      0     0      0       0
##   trajadura                                0      0     0      0       0
##   valdiguie                                0      0     0      0       0
##   verdejo                                  0      0     0      0       0
##   verdejo viura                            0      0     0      0       0
##   vermentino                               0      0     0      0       0
##   vernaccia                                0      0     2      0       0
##   viognier                                 0      0     0      0       0
##   viura                                    0      0     0      0       0
##   viura chardonnay                         0      0     0      0       0
##   white blend                              0      0     6      0       0
##   white riesling                           0      0     0      0       0
##   zinfandel                                0      0     0      0       0
##                                  country.extrem
## variety.extrem                    morocco new zealand portugal romania
##   agiorgitiko                           0           0        0       0
##   aglianico                             0           0        0       0
##   airen                                 0           0        0       0
##   albana                                0           0        0       0
##   alicante bouschet                     0           0        1       0
##   aragones                              0           0        1       0
##   arinto                                0           0        7       0
##   bical                                 0           0        2       0
##   bobal                                 0           0        0       0
##   bordeaux style red blend              0           0        0       0
##   bordeaux style white blend            0           0        0       0
##   cabernet blend                        0           0        0       0
##   cabernet merlot                       0           0        0       0
##   cabernet sauvignon                    0           0        0       6
```

```
##   cabernet sauvignon and tinta roriz       0          0        1        0
##   cabernet sauvignon carmenere             0          0        0        0
##   cabernet sauvignon malbec                0          0        0        0
##   cabernet sauvignon merlot                0          0        0        0
##   cabernet sauvignon shiraz                0          0        0        0
##   cabernet sauvignon syrah                 0          0        0        0
##   carineña                                 0          0        0        0
##   carmenere                                0          0        0        0
##   castelao                                 0          0        5        0
##   champagne blend                          0          0        0        0
##   chardonnay                               0          0        1        5
##   chardonnay viognier                      0          0        0        0
##   chenin blanc                             0          0        0        0
##   chenin blanc chardonnay                  0          0        0        0
##   cinsault                                 0          0        0        0
##   colombard                                0          0        0        0
##   corvina rondinella molinara              0          0        0        0
##   dornfelder                               0          0        0        0
##   fernao pires                             0          0        1        0
##   feteasca                                 0          0        0        1
##   fetească regală                          0          0        0        1
##   fume blanc                               0          0        0        0
##   g s m                                    0          0        0        0
##   gamay                                    0          0        0        0
##   garganega                                0          0        0        0
##   garnacha                                 0          0        0        0
##   garnacha blanca                          0          0        0        0
##   garnacha syrah                           0          0        0        0
##   garnacha tempranillo                     0          0        0        0
##   gewurztraminer                           0          0        0        1
##   glera                                    0          0        0        0
##   grenache                                 0          0        0        0
##   grenache syrah                           0          0        0        0
##   grillo                                   0          0        0        0
##   inzolia                                  0          0        0        0
##   johannisberg riesling                    0          0        0        0
##   lambrusco                                0          0        0        0
##   lemberger                                0          0        0        0
##   loureiro                                 0          0        7        0
##   macabeo                                  0          0        0        0
##   macabeo moscatel                         0          0        0        0
##   malagouzia chardonnay                    0          0        0        0
##   malbec                                   0          0        0        0
##   malbec bonarda                           0          0        0        0
##   malbec syrah                             0          0        0        0
##   malbec tempranillo                       0          0        0        0
##   melon                                    0          0        0        0
##   mencia                                   0          0        0        0
##   meritage                                 0          0        0        0
##   merlot                                   0          0        0        4
##   merlot cabernet                          0          0        0        0
##   merlot cabernet sauvignon                0          0        0        0
##   merlot malbec                            0          0        0        0
##   monastrell                               0          0        0        0
```

```
##   montepulciano                          0          0        0       0
##   morio muskat                           0          0        0       0
##   moscato                                0          0        0       4
##   muller thurgau                         0          0        0       0
##   muskat                                 0          0        1       0
##   muskat ottonel                         0          0        0       0
##   negroamaro                             0          0        0       0
##   nero davola                            0          0        0       0
##   pigato                                 0          0        0       0
##   pinot blanc                            0          0        0       0
##   pinot gris                             0          0        0       6
##   pinot noir                             0          0        0       8
##   pinotage                               0          0        0       0
##   portuguese red                         0          0      183       0
##   portuguese rose                        0          0        2       0
##   portuguese sparkling                   0          0        5       0
##   portuguese white                       0          0      135       0
##   primitivo                              0          0        0       0
##   red blend                              0          0        4       0
##   rhone style red blend                  0          0        0       0
##   rhone style white blend                0          0        0       0
##   riesling                               0          1        0       3
##   rosado                                 0          0        0       0
##   rosato                                 0          0        0       0
##   rose                                   0          0       40       0
##   sangiovese                             0          0        0       0
##   sauvignon blanc                        0          3        0       6
##   sauvignon blanc semillon               0          0        0       0
##   savatiano                              0          0        0       0
##   semillon                               0          0        0       0
##   semillon chardonnay                    0          0        0       0
##   semillon sauvignon blanc               0          0        0       0
##   sherry                                 0          0        0       0
##   shiraz                                 1          0        0       0
##   shiraz cabernet sauvignon              0          0        0       0
##   shiraz pinotage                        0          0        0       0
##   shiraz tempranillo                     0          0        0       0
##   sparkling blend                        0          0        0       0
##   symphony                               0          0        0       0
##   syrah                                  0          0        1       0
##   syrah cabernet                         0          0        0       0
##   syrah grenache                         0          0        0       0
##   tempranillo                            0          0        0       0
##   tempranillo blend                      0          0        0       0
##   tempranillo cabernet sauvignon         0          0        0       0
##   tempranillo garnacha                   0          0        0       0
##   tempranillo merlot                     0          0        0       0
##   tinta de toro                          0          0        0       0
##   tinta roriz                            0          0        1       0
##   torrontes                              0          0        0       0
##   touriga nacional                       0          0        2       0
##   touriga nacional cabernet sauvignon    0          0        2       0
##   trajadura                              0          0        1       0
##   valdiguie                              0          0        0       0
```

```
##   verdejo                                0           0        0       0
##   verdejo viura                          0           0        0       0
##   vermentino                             0           0        0       0
##   vernaccia                              0           0        0       0
##   viognier                               0           0        0       0
##   viura                                  0           0        0       0
##   viura chardonnay                       0           0        0       0
##   white blend                            0           0        3       2
##   white riesling                         0           0        0       0
##   zinfandel                              0           0        0       0
##                                  country.extrem
## variety.extrem                    slovenia south africa spain ukraine
##   agiorgitiko                            0            0     0       0
##   aglianico                              0            0     0       0
##   airen                                  0            0     1       0
##   albana                                 0            0     0       0
##   alicante bouschet                      0            0     0       0
##   aragones                               0            0     0       0
##   arinto                                 0            0     0       0
##   bical                                  0            0     0       0
##   bobal                                  0            0     2       0
##   bordeaux style red blend               0            0     0       0
##   bordeaux style white blend             0            0     0       0
##   cabernet blend                         0            0     0       0
##   cabernet merlot                        0            0     0       0
##   cabernet sauvignon                     0            2     4       1
##   cabernet sauvignon and tinta roriz     0            0     0       0
##   cabernet sauvignon carmenere           0            0     0       0
##   cabernet sauvignon malbec              0            0     0       0
##   cabernet sauvignon merlot              0            2     0       0
##   cabernet sauvignon shiraz              0            0     0       0
##   cabernet sauvignon syrah               0            0     0       0
##   carineña                               0            0     1       0
##   carmenere                              0            0     0       0
##   castelao                               0            0     0       0
##   champagne blend                        0            0     4       0
##   chardonnay                             1            2     6       0
##   chardonnay viognier                    0            0     0       0
##   chenin blanc                           0            5     0       0
##   chenin blanc chardonnay                0            2     0       0
##   cinsault                               0            0     0       0
##   colombard                              0            0     0       0
##   corvina rondinella molinara            0            0     0       0
##   dornfelder                             0            0     0       0
##   fernao pires                           0            0     0       0
##   feteasca                               0            0     0       0
##   fetească regală                        0            0     0       0
##   fume blanc                             0            0     0       0
##   g s m                                  0            0     0       0
##   gamay                                  0            0     0       0
##   garganega                              0            0     0       0
##   garnacha                               0            0    53       0
##   garnacha blanca                        0            0     3       0
##   garnacha syrah                         0            0     2       0
```

```
##   garnacha tempranillo                0          0     1       0
##   gewurztraminer                      0          0     0       0
##   glera                               0          0     0       0
##   grenache                            0          0     5       0
##   grenache syrah                      0          0     3       0
##   grillo                              0          0     0       0
##   inzolia                             0          0     0       0
##   johannisberg riesling               0          0     0       0
##   lambrusco                           0          0     0       0
##   lemberger                           0          0     0       0
##   loureiro                            0          0     0       0
##   macabeo                             0          0     3       0
##   macabeo moscatel                    0          0     2       0
##   malagouzia chardonnay               0          0     0       0
##   malbec                              0          0     0       0
##   malbec bonarda                      0          0     0       0
##   malbec syrah                        0          0     0       0
##   malbec tempranillo                  0          0     0       0
##   melon                               0          0     0       0
##   mencia                              0          0     3       0
##   meritage                            0          0     0       0
##   merlot                              0          3     2       1
##   merlot cabernet                     0          0     0       0
##   merlot cabernet sauvignon           0          3     0       0
##   merlot malbec                       0          0     0       0
##   monastrell                          0          0     7       0
##   montepulciano                       0          0     0       0
##   morio muskat                        0          0     0       0
##   moscato                             0          2     5       0
##   muller thurgau                      0          0     0       0
##   muskat                              0          0     0       0
##   muskat ottonel                      0          0     0       0
##   negroamaro                          0          0     0       0
##   nero davola                         0          0     0       0
##   pigato                              0          0     0       0
##   pinot blanc                         0          0     0       0
##   pinot gris                          1          0     0       0
##   pinot noir                          1          0     0       0
##   pinotage                            0          1     0       0
##   portuguese red                      0          0     0       0
##   portuguese rose                     0          0     0       0
##   portuguese sparkling                0          0     0       0
##   portuguese white                    0          0     0       0
##   primitivo                           0          0     0       0
##   red blend                           0          1    19       0
##   rhone style red blend               0          0     0       0
##   rhone style white blend             0          0     0       0
##   riesling                            0          0     0       0
##   rosado                              0          0     4       0
##   rosato                              0          0     0       0
##   rose                                0          0    29       1
##   sangiovese                          0          0     0       0
##   sauvignon blanc                     0          7     3       0
##   sauvignon blanc semillon            0          0     0       0
```

```
##   savatiano                                0            0     0       0
##   semillon                                 0            0     0       0
##   semillon chardonnay                      0            0     0       0
##   semillon sauvignon blanc                 0            0     0       0
##   sherry                                   0            0     2       0
##   shiraz                                   0            5     2       0
##   shiraz cabernet sauvignon                0            0     0       0
##   shiraz pinotage                          0            1     0       0
##   shiraz tempranillo                       0            0     3       0
##   sparkling blend                          0            0     8       0
##   symphony                                 0            0     0       0
##   syrah                                    0            0     2       0
##   syrah cabernet                           0            0     0       0
##   syrah grenache                           0            0     1       0
##   tempranillo                              0            0    89       0
##   tempranillo blend                        0            0     7       0
##   tempranillo cabernet sauvignon           0            0     6       0
##   tempranillo garnacha                     0            0     4       0
##   tempranillo merlot                       0            0     4       0
##   tinta de toro                            0            0     1       0
##   tinta roriz                              0            0     0       0
##   torrontes                                0            0     0       0
##   touriga nacional                         0            0     0       0
##   touriga nacional cabernet sauvignon      0            0     0       0
##   trajadura                                0            0     0       0
##   valdiguie                                0            0     0       0
##   verdejo                                  0            0     5       0
##   verdejo viura                            0            0     1       0
##   vermentino                               0            0     0       0
##   vernaccia                                0            0     0       0
##   viognier                                 0            0     0       0
##   viura                                    0            0    19       0
##   viura chardonnay                         0            0     1       0
##   white blend                              0            3    26       0
##   white riesling                           0            0     0       0
##   zinfandel                                0            0     0       0
##                                  country.extrem
## variety.extrem                    uruguay  us
##   agiorgitiko                           0   0
##   aglianico                             0   0
##   airen                                 0   0
##   albana                                0   0
##   alicante bouschet                     0   0
##   aragones                              0   0
##   arinto                                0   0
##   bical                                 0   0
##   bobal                                 0   0
##   bordeaux style red blend              0   2
##   bordeaux style white blend            0   0
##   cabernet blend                        0   1
##   cabernet merlot                       0   3
##   cabernet sauvignon                    1  83
##   cabernet sauvignon and tinta roriz    0   0
##   cabernet sauvignon carmenere          0   0
```

```
##   cabernet sauvignon malbec                0   0
##   cabernet sauvignon merlot                0   1
##   cabernet sauvignon shiraz                0   0
##   cabernet sauvignon syrah                 0   0
##   carineña                                 0   0
##   carmenere                                0   0
##   castelao                                 0   0
##   champagne blend                          0   6
##   chardonnay                               0 120
##   chardonnay viognier                      0   0
##   chenin blanc                             0  12
##   chenin blanc chardonnay                  0   0
##   cinsault                                 0   0
##   colombard                                0   0
##   corvina rondinella molinara              0   0
##   dornfelder                               0   0
##   fernao pires                             0   0
##   feteasca                                 0   0
##   fetească regală                          0   0
##   fume blanc                               0   3
##   g s m                                    0   0
##   gamay                                    0   0
##   garganega                                0   0
##   garnacha                                 0   0
##   garnacha blanca                          0   0
##   garnacha syrah                           0   0
##   garnacha tempranillo                     0   0
##   gewurztraminer                           0  19
##   glera                                    0   0
##   grenache                                 0   0
##   grenache syrah                           0   0
##   grillo                                   0   0
##   inzolia                                  0   0
##   johannisberg riesling                    0   7
##   lambrusco                                0   0
##   lemberger                                0   3
##   loureiro                                 0   0
##   macabeo                                  0   0
##   macabeo moscatel                         0   0
##   malagouzia chardonnay                    0   0
##   malbec                                   0   0
##   malbec bonarda                           0   0
##   malbec syrah                             0   0
##   malbec tempranillo                       0   0
##   melon                                    0   0
##   mencia                                   0   0
##   meritage                                 0   0
##   merlot                                   1  77
##   merlot cabernet                          0   5
##   merlot cabernet sauvignon                0   0
##   merlot malbec                            0   0
##   monastrell                               0   0
##   montepulciano                            0   0
##   morio muskat                             0   1
```

```
##   moscato                                0  16
##   muller thurgau                         0   1
##   muskat                                 0   3
##   muskat ottonel                         0   0
##   negroamaro                             0   0
##   nero davola                            0   0
##   pigato                                 0   0
##   pinot blanc                            0   0
##   pinot gris                             0  61
##   pinot noir                             0  24
##   pinotage                               0   0
##   portuguese red                         0   0
##   portuguese rose                        0   0
##   portuguese sparkling                   0   0
##   portuguese white                       0   0
##   primitivo                              0   0
##   red blend                              0  26
##   rhone style red blend                  0   0
##   rhone style white blend                0   0
##   riesling                               0  50
##   rosado                                 0   0
##   rosato                                 0   0
##   rose                                   0   6
##   sangiovese                             0   0
##   sauvignon blanc                        0  66
##   sauvignon blanc semillon               0   0
##   savatiano                              0   0
##   semillon                               0   1
##   semillon chardonnay                    0   3
##   semillon sauvignon blanc               0   0
##   sherry                                 0   0
##   shiraz                                 0  26
##   shiraz cabernet sauvignon              0   0
##   shiraz pinotage                        0   0
##   shiraz tempranillo                     0   0
##   sparkling blend                        0   3
##   symphony                               0   2
##   syrah                                  0  21
##   syrah cabernet                         0   0
##   syrah grenache                         0   0
##   tempranillo                            0   0
##   tempranillo blend                      0   0
##   tempranillo cabernet sauvignon         0   0
##   tempranillo garnacha                   0   0
##   tempranillo merlot                     0   0
##   tinta de toro                          0   0
##   tinta roriz                            0   0
##   torrontes                              0   0
##   touriga nacional                       0   0
##   touriga nacional cabernet sauvignon    0   0
##   trajadura                              0   0
##   valdiguie                              0   1
##   verdejo                                0   0
##   verdejo viura                          0   0
```

```
##   vermentino                0   2
##   vernaccia                 0   0
##   viognier                  0   9
##   viura                     0   0
##   viura chardonnay          0   0
##   white blend               0  15
##   white riesling            0   4
##   zinfandel                 0  48
```

5 Resolució del problema. A partir dels resultats obtinguts, quines són les conclusions? Els resultats permeten respondre al problema?

A partir de la taula final podem veure quins paisos ens donen millor qualitat preu per a cada varietat. Observem, també, que tenim varietats que no es troben en cap dels països anteriors i, per tant, no obtenim informació sobre elles.

Amb l'estudi realitzat el responen parcialment, però creiem és un bon inici i amb uns coneixements més amplis del programari segurament podriem obtenir millors resultats.