

```
library(dplyr)
```

```
rladies_global %>%  
  filter(city == 'Your city')
```



# "Clasificación geográfica de hollejos de uva de Mendoza mediante Regresión Logística Multinomial y R"

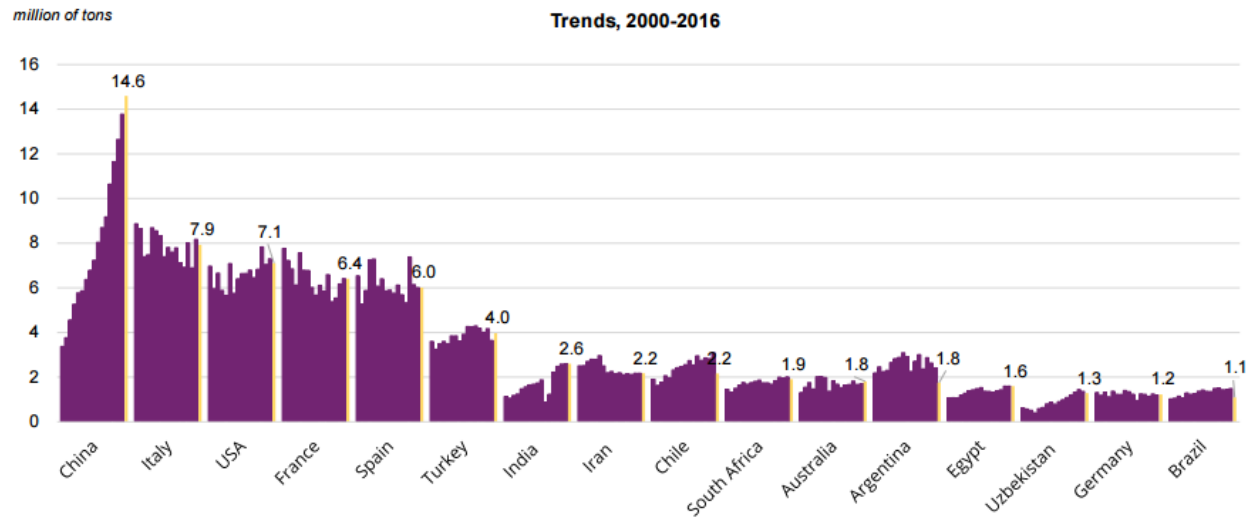
Lic. Brenda V. Canizo



# 1. Para empezar algunas estadísticas...



- Argentina: 12º productor de uvas  
9º productor de vinos



## 👉 A nivel nacional



- Mendoza: 1º productor de uvas y vinos



**PRODUCCIÓN DE UVA SEGÚN PROVINCIA DE ORIGEN  
TODOS LOS DESTINOS  
(ELABORAR VINOS Y MOSTOS, CONSUMO EN FRESCO Y PARA PASAS DE UVA)  
Año 2016 - En quintales métricos**

**TOTAL PAÍS**

PROVINCIAS	ELABORACION	CONSUMO EN FRESCO	PASAS	TOTAL
BUENOS AIRES	1.303	-	-	1.303
CATAMARCA	100.032	-	-	100.032
CHUBUT	181	-	-	181
CORDOBA	3.638	-	-	3.638
ENTRE RIOS	731	-	-	731
JUJUY	236	-	-	236
LA PAMPA	17.460	-	-	17.460
LA RIOJA	707.302	-	10.514	717.816
<b>MENDOZA</b>	<b>10.191.468</b>	<b>2.848</b>	<b>1.736</b>	<b>10.196.052</b>
NEUQUEN	125.720	-	-	125.720
RIO NEGRO	54.717	-	-	54.717
SALTA	211.962	-	-	211.962
SAN JUAN	5.624.149	89.864	438.274	6.152.287
SAN LUIS	1.153	-	-	1.153
TUCUMAN	889	-	-	889
<b>TOTAL</b>	<b>17.040.942</b>	<b>92.712</b>	<b>450.524</b>	<b>17.584.178</b>

# Clasificación Geográfica



Ley 25163 define **Denominación de Origen Controlada (DOC)**.

## CAPITULO IV: DENOMINACION DE ORIGEN CONTROLADA

**ARTICULO 13.** - A los efectos de la presente ley, se entiende por Denominación de Origen Controlada (DOC) el nombre que identifica un producto originario de una región, de una localidad o de área de producción delimitada del territorio nacional, cuyas cualidades o características particulares se deben exclusiva o esencialmente al medio geográfico, abarcando los factores naturales factores humanos.



- ❖ La composición elemental del vino refleja la geoquímica del suelo en el que se cultivan las vides.
- ❖ La mayoría de los elementos minerales no se metabolizan o se transforman durante los procesos enológicos y permanecen casi inalterados.



**Composición elemental**



**Huella Dactilar**



**Análisis estadístico multivariado**



**CLASIFICACIÓN GEOGRÁFICA**

## 2. Manos a la obra...



## Instrumentación

☞ **ICP-MS** Perkin-Elmer SCIEX,  
ELAN DRC-e (Thornhill, Canadá)



☞ **29 elementos analizados:**  $^{47}\text{Ti}$ ,  $^{51}\text{V}$ ,  $^{55}\text{Mn}$ ,  $^{57}\text{Fe}$ ,  $^{59}\text{Co}$ ,  $^{60}\text{Ni}$ ,  $^{63}\text{Cu}$ ,  $^{66}\text{Zn}$ ,  $^{69}\text{Ga}$ ,  $^{75}\text{As}$ ,  $^{85}\text{Rb}$ ,  $^{89}\text{Y}$ ,  $^{90}\text{Zr}$ ,  $^{93}\text{Nb}$ ,  $^{98}\text{Mo}$ ,  $^{107}\text{Ag}$ ,  $^{130}\text{Te}$ ,  $^{133}\text{Cs}$ ,  $^{139}\text{La}$ ,  $^{140}\text{Ce}$ ,  $^{141}\text{Pr}$ ,  $^{142}\text{Nd}$ ,  $^{152}\text{Sm}$ ,  $^{153}\text{Eu}$ ,  $^{158}\text{Gd}$ ,  $^{169}\text{Tm}$ ,  $^{175}\text{Lu}$ ,  $^{205}\text{Tl}$ , and  $^{238}\text{U}$ .



## Muestreo y tratamiento de muestras

Toma de muestras de 35 fincas de:

**GY: Guaymallén**

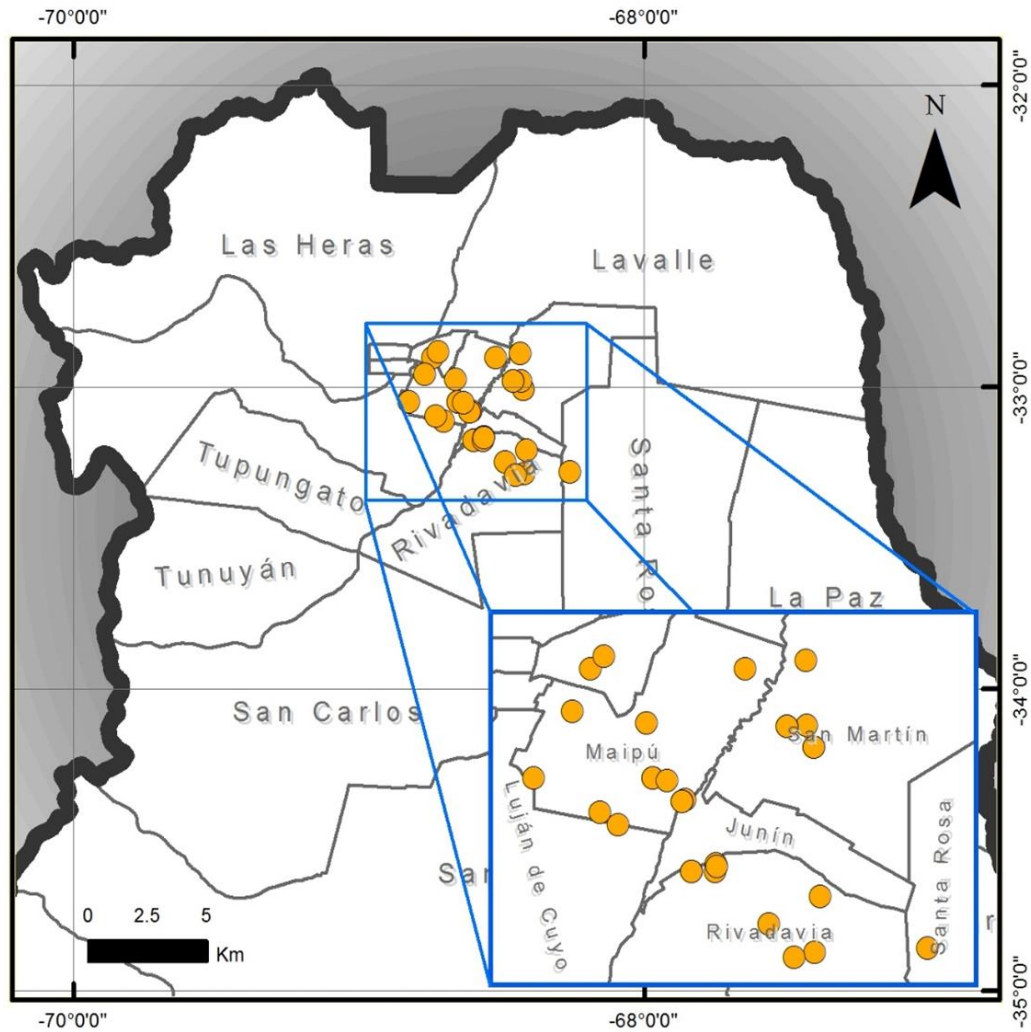
**MP: Maipú**

**RV: Rivadavia**

**JN: Junín**

**SM: San Martín**

- Varietales: Cabernet Sauvignon, Malbec, Bonarda, Aspirant Bouchet (tintos), Chardonnay, Sauvignon Blanc, y Pedro Jiménez (blancos).
- Se separó manualmente hollejos, semillas y pulpa.
- Tratamientos de los hollejos:  
liofilización -> molienda - > digestión -> análisis



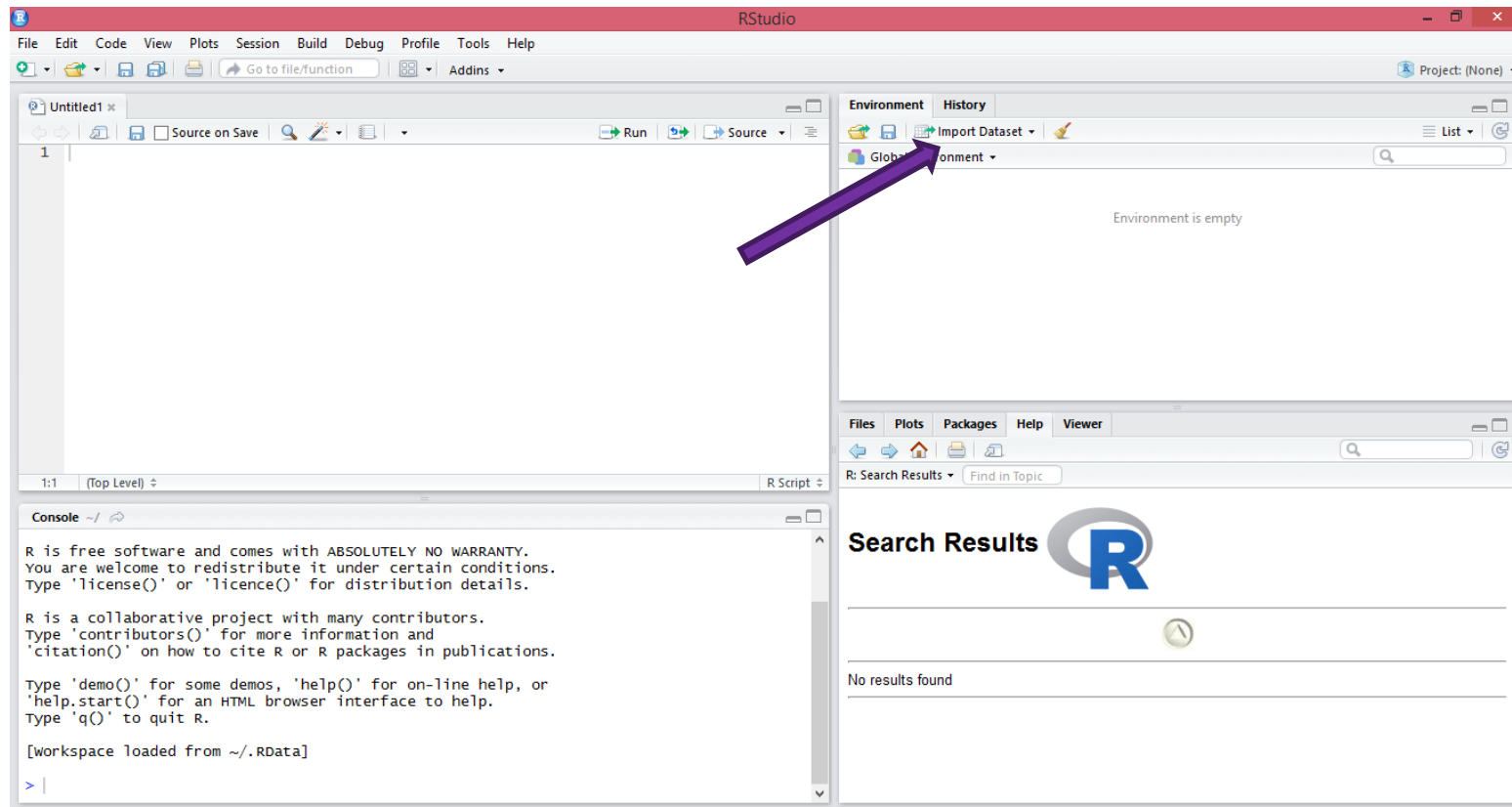


### 3. Procesando datos en R...



## 1

## Carga de los datos



The screenshot displays the RStudio environment. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar below the menu bar contains icons for opening files, saving, running code, and other functions. The main editor window on the left shows a file named 'Untitled1' with a single line of code: '1'. The Environment pane on the right, titled 'Environment', shows a 'Global Environment' and states 'Environment is empty'. A purple arrow points to the 'Import Dataset' button in the Environment pane. The Console pane at the bottom left shows the R startup message: 'R is free software and comes with ABSOLUTELY NO WARRANTY. You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details. R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications. Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. [workspace loaded from ~/.RData]'. The bottom right pane shows 'Search Results' with the R logo and the text 'No results found'.



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

Environment History

Untitled1 x

Import Text Data

File/Url:

C:/Users/Brenda Chan/Desktop/Doctorado en Química/Cursos de Posgrado/8. Programación y Análisis de Datos en R con Aplicaciones Científicas 60hs/trabajo final/hollejos.csv

Browse...

Data Preview:

Departamento (character)	Ti (double)	V (double)	Mn (double)	Fe (double)	Co (double)	Ni (double)	Cu (double)	Zn (double)	Ga (double)	As (double)	Rb (double)	Y (double)	Zr (double)	Nb (double)	Mo (double)
RV	0.0001	0.0001	2.154	3.148	0.0370	9.817	8.249	10.1830	0.0030	0.0770	1.194	0.0030	0.057	0.185	8.061
RV	2.0320	0.0860	3.090	36.266	0.0340	13.675	14.571	12.5610	0.0180	0.1780	1.833	0.0090	0.191	0.253	5.913
RV	2.0460	0.0690	2.991	36.564	0.0310	13.678	14.664	12.1950	0.0200	0.2320	1.748	0.0090	0.176	0.128	4.963
RV	2.3190	0.0690	2.949	37.500	0.0320	14.301	14.388	11.8950	0.0250	0.1930	1.791	0.0110	0.171	0.108	4.931
RV	2.1870	0.0500	2.986	38.024	0.0330	14.261	14.643	12.3520	0.0250	0.1420	1.810	0.0120	0.180	0.100	5.192
RV	0.0001	0.0001	1.893	20.176	0.0170	9.217	15.067	11.5950	0.0070	0.1730	1.429	0.0001	0.142	0.214	5.005
RV	0.5930	0.0001	2.632	30.968	0.0100	9.203	17.191	6.7110	0.0110	0.0830	1.291	0.0090	0.796	0.270	4.160
PV	0.4700	0.0001	2.333	27.685	0.0220	9.546	17.810	6.2410	0.0200	0.0830	1.255	0.0080	0.806	0.058	2.077

Previewing first 50 entries.

Import Options:

Name: hollejos

Skip: 0

☒ First Row as Names

☒ Trim Spaces

☒ Open Data Viewer

Delimiter: Comma

Quotes: Default

Local: Configure...

Escape: None

Comment: Default

NA: Default

Code Preview:

```
library(readr)
hollejos <- read_csv("C:/Users/Brenda Chan/desktop/Doctorado en química/Cursos de Posgrado/8. Programación y Análisis de Datos en R con Aplicaciones Científicas 60hs/trabajo final/hollejos.csv")
view(hollejos)
```

Import Cancel

R is free  
You are w  
Type 'lic  
  
R is a co  
Type 'con  
'citation  
  
Type 'dem  
'help.sta  
Type 'q()  
  
[workspace loaded from ~/.RData]

## 2 Inspección de los datos

```
> head(hollejos)
# A tibble: 6 × 30
  Departamento Ti      V      Mn      Fe      Co      Ni      Cu      Zn      Ga      As
  <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1      RV 0.0001 0.0001 2.154  3.148 0.037  9.817  8.249 10.183 0.003 0.077
2      RV 2.0320 0.0860 3.090 36.266 0.034 13.675 14.571 12.561 0.018 0.178
3      RV 2.0460 0.0690 2.991 36.564 0.031 13.678 14.664 12.195 0.020 0.232
4      RV 2.3190 0.0690 2.949 37.500 0.032 14.301 14.388 11.895 0.025 0.193
5      RV 2.1870 0.0500 2.986 38.024 0.033 14.261 14.643 12.352 0.025 0.142
6      RV 0.0001 0.0001 1.893 20.176 0.017  9.217 15.067 11.595 0.007 0.173
# ... with 19 more variables: Rb <dbl>, Y <dbl>, Zr <dbl>, Nb <dbl>, Mo <dbl>,
# Ag <dbl>, Te <dbl>, Cs <dbl>, La <dbl>, Ce <dbl>, Pr <dbl>, Nd <dbl>, Sm <dbl>,
# Eu <dbl>, Gd <dbl>, Tm <dbl>, Lu <dbl>, Tl <dbl>, U <dbl>
```



```
> str(hollejos)
Classes 'tbl_df', 'tbl' and 'data.frame':    413 obs. of  30 variables:
 $ Departamento: chr  "RV" "RV" "RV" "RV" ...
 $ Ti           : num  0.0001 2.032 2.046 2.319 2.187 ...
 $ V           : num  0.0001 0.086 0.069 0.069 0.05 0.0001 0.0001 0.0001 0.0001 0.0001 .
 ..
 $ Mn          : num  2.15 3.09 2.99 2.95 2.99 ...
 $ Fe          : num  3.15 36.27 36.56 37.5 38.02 ...
 $ Co          : num  0.037 0.034 0.031 0.032 0.033 0.017 0.01 0.022 0.016 0.015 ...
 $ Ni          : num  9.82 13.68 13.68 14.3 14.26 ...
 $ Cu          : num  8.25 14.57 14.66 14.39 14.64 ...
 $ Zn          : num  10.2 12.6 12.2 11.9 12.4 ...
 $ Ga          : num  0.003 0.018 0.02 0.025 0.025 0.007 0.011 0.02 0.019 0.013 ...
 $ As          : num  0.077 0.178 0.232 0.193 0.142 0.173 0.083 0.083 0.061 0.104 ...
 $ Rb          : num  1.19 1.83 1.75 1.79 1.81 ...
 $ Y           : num  0.003 0.009 0.009 0.011 0.012 0.0001 0.009 0.008 0.008 0.005 ...
 $ Zr          : num  0.057 0.191 0.176 0.171 0.18 0.142 0.796 0.806 0.781 0.798 ...
 $ Nb          : num  0.185 0.253 0.128 0.108 0.1 0.214 0.27 0.058 0.031 0.021 ...
 $ Mo          : num  8.06 5.91 4.96 4.93 5.19 ...
 $ Ag          : num  0.161 0.175 0.182 0.167 0.176 0.16 0.0001 0.0001 0.0001 0.0001 ...
 $ Te          : num  0.0001 7.354 3.618 0.892 0.0001 ...
 $ Cs          : num  0.015 0.001 0.0001 0.0001 0.0001 0.007 0.002 0.001 0.001 0.001 ...
 $ La          : num  0.074 0.045 0.046 0.045 0.044 0.052 0.053 0.053 0.058 0.055 ...
 $ Ce          : num  0.035 0.056 0.058 0.05 0.053 0.028 0.07 0.067 0.069 0.073 ...
 $ Pr          : num  0.009 0.007 0.005 0.007 0.006 0.002 0.004 0.003 0.003 0.003 ...
 $ Nd          : num  0.0001 0.012 0.005 0.02 0.02 0.0001 0.015 0.009 0.009 0.007 ...
 $ Sm          : num  1e-03 6e-03 5e-03 6e-03 6e-03 1e-04 1e-03 2e-03 1e-03 3e-03 ...
 $ Eu          : num  4e-03 4e-03 4e-03 2e-03 4e-03 1e-04 1e-03 1e-03 1e-03 1e-03 ...
 $ Gd          : num  6e-03 5e-03 5e-03 5e-03 5e-03 1e-04 1e-03 1e-03 2e-03 1e-03 ...
 $ Tm          : num  8e-03 4e-03 2e-03 2e-03 2e-03 1e-03 1e-03 1e-04 1e-04 1e-04 ...
 $ Lu          : num  1e-02 2e-03 2e-03 4e-03 4e-03 3e-03 1e-04 1e-04 1e-04 1e-04 ...
 $ Tl          : num  0.161 0.0001 0.0001 0.0001 0.0001 0.125 0.055 0.076 0.059 0.078 ..
 .
 $ U           : num  0.01 0.034 0.037 0.038 0.037 0.008 0.019 0.02 0.015 0.016 ...
```

➤ `table(hollejos$Departamento)`

```
GY JN MP RV SM
22 27 200 90 74
```

### 3 Desarrollo del modelo de RLM

- La **regresión logística multinomial** generaliza el método de regresión logística para problemas multiclase, es decir, con más de dos posibles resultados discretos, es decir la variable es categórica.
- Se trata de un modelo que se utiliza para predecir las **probabilidades** de los diferentes resultados posibles de una distribución categórica como variable dependiente, dado un conjunto de variables independientes.



## #Cargo la librería que tiene modelo de regresión multinomial

```
library(nnet)
```

## #Separo los datos en training(70%) y testing(30%)

```
index <- sample(1:413, 290, replace=FALSE)
```

```
dtraining <- hollejos[index,]
```

```
dtesting <- hollejos[-index,]
```

```
> index
[1] 16 412 101 192 378 146 405 215 240 246 324 11 368 251 30 376 310 18 408 326
[21] 164 39 340 369 383 264 244 327 194 103 104 291 389 397 298 135 65 178 95 36
[41] 335 150 353 337 211 60 64 189 69 136 72 289 40 206 234 318 351 230 170 391
[61] 88 283 379 243 29 245 176 53 93 354 371 25 331 190 357 231 400 138 144 212
[81] 216 200 263 179 100 385 281 1 275 195 97 349 303 109 413 205 198 70 15 6
[101] 273 163 32 250 286 68 118 280 258 23 62 20 183 402 266 128 202 140 33 272
[121] 228 333 158 201 365 312 374 347 129 161 262 122 356 119 42 292 61 252 28 85
[141] 21 363 276 184 22 157 382 87 287 151 14 185 82 203 214 372 196 284 261 154
[161] 304 78 343 159 307 394 268 197 46 401 125 162 186 96 227 83 116 336 247 94
[181] 112 288 114 168 199 147 130 188 92 9 221 141 269 253 350 89 145 123 317 108
[201] 219 217 47 19 294 278 187 315 338 248 50 236 2 222 139 131 362 321 235 7
[221] 152 63 300 143 224 84 308 75 191 55 54 302 44 193 142 111 133 373 404 279
[241] 17 107 98 232 375 181 156 277 332 13 370 295 223 110 322 26 80 274 282 360
[261] 48 213 237 41 239 127 361 115 319 67 297 49 358 249 171 210 233 406 254 76
[281] 366 267 27 117 38 226 134 386 388 8
```

## #Hago el modelo con mis observaciones de training

```
modelo <- multinom(Departamento~., data=dtraining)
```

```
> summary(modelo)
```

```
call:
```

```
multinom(formula = Departamento ~ ., data = dtraining)
```

```
Coefficients:
```

```
(Intercept)      Ti      V      Mn      Fe      Co      Ni      Cu
JN  2.191784 -64.82001  42.48325  48.786444 -2.364052  14.58973  0.2623107  18.28687
MP  20.707889 -58.96531  44.90463  4.457424 -2.522236  17.09780  0.1881855  19.78460
RV  24.603777 -58.68780  21.92547  4.066652 -2.446614 -18.56959  0.1784459  19.67887
SM  19.800057 -58.36798 -125.05979  6.943456 -2.619488 -24.43383  0.2112712  19.93921

      Zn      Ga      As      Rb      Y      Zr      Nb
JN -13.089963 -23.337536 -278.95805 -39.30913 -5.018079 -95.71122  44.18461
MP  -3.244801  46.475202  10.60374  17.35074  31.882016 -99.47017 -71.52054
RV  -3.073019 -26.290908  11.29867  15.86682 -16.096450 -104.61066 -80.18207
SM  -3.030093  2.755939  10.04520  13.61595 -19.249134 -102.86053 -73.30666

      Mo      Ag      Te      Cs      La      Ce      Pr
JN  5.7045003  325.15747  0.3503096  1.564548 -34.80761 -79.42632  1.406587
MP  -0.5230415 -81.86610  1.5627478  314.464187 -72.79128 -128.01940 -20.165996
RV  -0.1967334 -113.21829  1.7093118 -253.232804 -76.37412 -146.64732  47.699651
SM  -0.5239666 -72.28093  1.7251467 -72.787688 -171.23421 -84.75680 -33.377108

      Nd      Sm      Eu      Gd      Tm      Lu      Tl
JN -2.796331  0.2649217 -2.007563  2.733860 -0.2967938 -2.323490  72.19087
MP  98.119761 -20.0750291 -36.846458 -12.935413 -16.2692912 -31.527613 -42.17778
RV  -93.697009  47.9949667  53.531571  12.854907  20.7040070  42.779607 -21.42159
SM  -6.590116 -27.2011263 -15.862238 -3.092897 -4.6216581 -9.474706 -29.48342

      U
JN  -1.542413
MP  -211.667077
RV  113.942230
SM  84.958909
```

Std. Errors:

	(Intercept)	Ti	V	Mn	Fe	Co	Ni	
JN	1.849386	1.0952237	0.05318181	5.312815	11.740066	0.02244394	12.087672	
MP	1.147936	0.7742679	14.95968494	1.871528	5.782642	19.70382384	4.991373	
RV	1.063356	0.7885881	15.50827819	1.868889	5.782640	19.80621806	4.991336	
SM	1.480177	0.9090230	5.10122324	1.887893	5.782707	23.76154607	4.991399	
	Cu	Zn	Ga	As	Rb	Y	Zr	Nb
JN	10.540995	6.850306	0.01742983	0.3394343	1.4724184	0.01224619	2.378677	0.5019073
MP	3.711477	3.131557	16.64641457	2.9658125	0.5723959	0.87475928	1.319143	1.6645032
RV	3.711488	3.131478	18.59039208	2.9427944	0.5664464	1.27536783	1.255605	1.6935604
SM	3.711533	3.131414	3.03036595	3.9112461	0.6732950	0.69646312	1.557214	2.5344495
	Mo	Ag	Te	Cs	La	Ce	Pr	
JN	8.624070	0.6470648	0.6148586	0.01198912	0.04666365	0.02725962	0.003584672	
MP	2.859822	4.6034663	0.2362415	1.64320851	5.30816438	10.84848152	0.422160948	
RV	2.859289	5.2330951	0.2363384	1.71379008	5.13766912	10.39915134	0.637647562	
SM	2.860356	6.0422289	0.2436960	0.26770975	7.87280238	8.55727617	0.532673797	
	Nd	Sm	Eu	Gd	Tm	Lu	Tl	
JN	0.0139869	0.01006861	0.0001966243	0.0008282065	0.0001968502	0.000523073	0.5596299	
MP	3.4354752	0.35839865	0.2881355623	0.4439610570	0.1658561831	0.285530269	3.7479527	
RV	4.4371010	0.59097530	0.3197179472	0.5198125021	0.1829018607	0.282932374	2.8206238	
SM	1.2786410	0.30167225	0.1049540081	0.3097303857	0.1278912868	0.127693366	3.7379903	
	U							
JN	0.06589611							
MP	9.49714207							
RV	11.69764604							
SM	3.17046905							

Residual Deviance: 202.2117

AIC: 442.2117

**predict(modelo, newdata = dtesting)**

```
[1] JN RV MP RV RV SM GY JN SM MP MP RV GY GY MP MP MP SM GY GY GY GY GY GY MP JN  
[27] JN JN JN JN JN SM MP JN JN JN RV RV RV RV SM MP MP RV RV RV SM SM SM RV SM SM  
[53] SM RV RV RV RV MP MP RV RV RV SM SM RV RV MP MP MP MP MP MP MP MP MP MP MP  
[79] MP MP SM SM MP RV MP MP MP MP MP JN MP MP MP MP MP MP MP MP MP MP MP MP MP  
[105] MP MP MP MP MP MP MP MP MP MP MP MP MP MP RV MP RV MP RV MP  
Levels: GY JN MP RV SM
```



```
predict(modelo, dtesting, type = "prob")
```

	GY	JN	MP	RV	SM
1	6.396108e-70	1.000000e+00	2.167728e-15	3.056751e-17	5.106948e-17
2	1.478351e-24	3.342229e-23	1.225024e-01	8.772500e-01	2.475531e-04
3	5.108794e-05	4.393896e-45	5.800451e-01	4.194860e-01	4.177901e-04
4	4.967786e-12	8.090216e-49	4.507556e-01	5.481440e-01	1.100468e-03
5	1.057719e-51	7.289500e-43	2.523314e-02	9.711196e-01	3.647287e-03
6	1.411684e-52	7.312139e-14	1.318556e-01	5.049561e-02	8.176488e-01
7	1.000000e+00	1.415818e-09	2.425376e-10	4.665828e-17	4.455604e-26
8	4.668912e-85	9.999999e-01	5.862840e-11	4.218390e-08	1.064893e-07
.					
.					
123	8.931952e-39	2.097858e-43	8.449076e-01	1.501431e-01	4.949298e-03

## #Matriz de confusión

```
mc <- table(predict(modelo, newdata = dtesting), dtesting$Departamento)
```

	GY	JN	MP	RV	SM
GY	8	0	0	0	1
JN	0	8	1	2	1
MP	0	1	50	3	8
RV	0	0	4	16	5
SM	0	1	2	3	9

## #tasa de buena clasificación

```
TBC <- (8+8+50+16+9)/123
```

```
sum(diag(mc))/sum(mc)
```

```
#[1] 0.7398374
```

#en porcentaje es de **73.98%**

## #tasa de mala clasificación

```
TMC <- 1-TBC
```

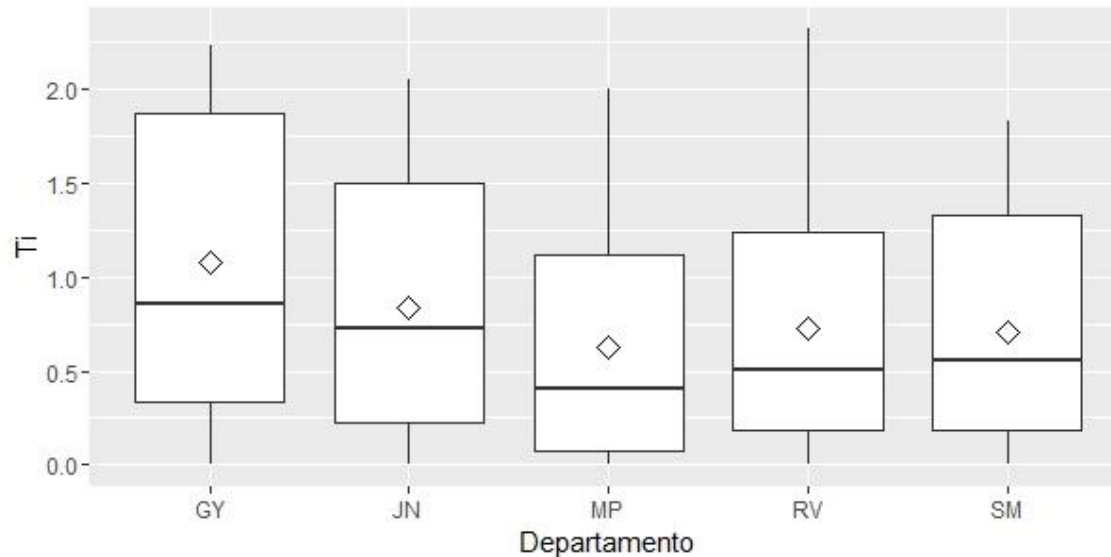
```
1-sum(diag(mc))/sum(mc)
```

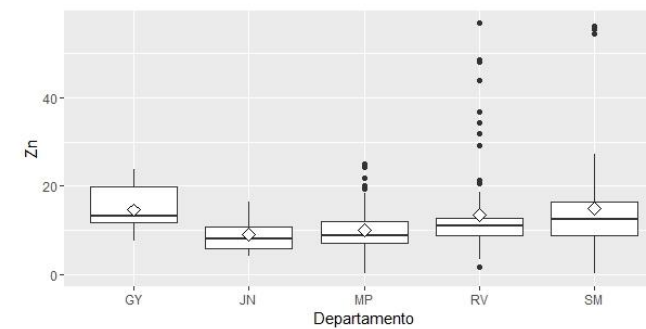
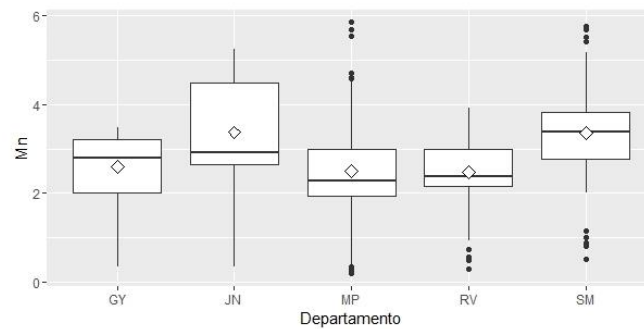
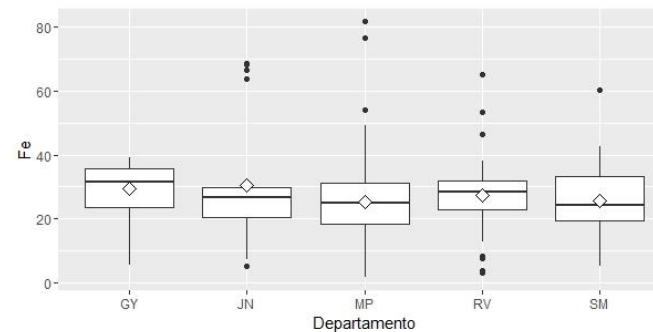
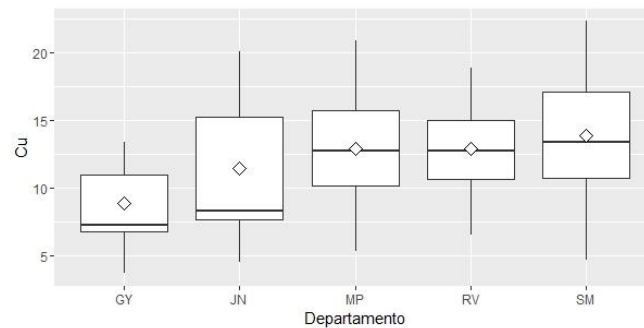
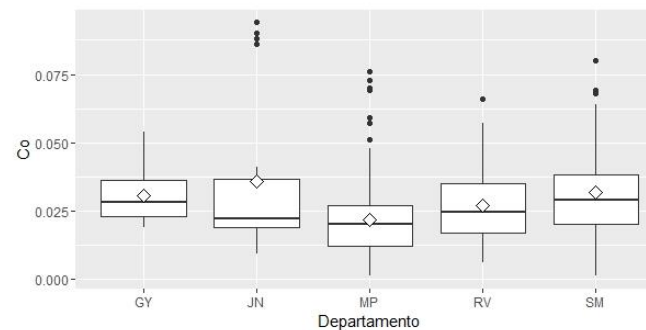
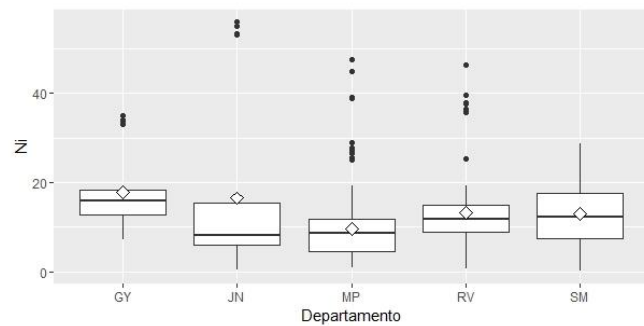
```
#[1] 0.2601626
```

#en porcentaje es de **26.02%**

## 5 Estadística descriptiva

```
g <- ggplot(hollejos, aes(x=Departamento, y= Ti)) + geom_boxplot() + stat_summary(fun.y="mean",
geom="point", shape=23, size=3, fill="white")
```







## 4. Finalizando...



- Por medio de la aplicación de la técnica RLM combinados con análisis multielemental por ICP-MS, es posible una clasificación intrarregional de hollejos de uva con una tasa de predicción de aproximadamente el 74%.
- Los metales analizados son buenos indicadores de origen de los hollejos y se pueden utilizar como criterio para garantizar la autenticidad y origen geográfico de un vino.

**GRACIAS POR TU  
ATENCIÓN!!!!**

