

DataMining en Ciencias. Trabajo Práctico 1:

Procesamiento de los datos

Freiman Sebastián sebafreiman@gmail.com, Louzau Estefanía eplouz@gmail.com, Michalla Bárbara barbaramichalla@gmail.com, Ríos Leandro Martín rios.leandromartin@gmail.com

13 de septiembre de 2016

Contents

Introducción	1
Preparación de los datos	1
Tratamiento de Outliers	5
Otros gráficos y técnicas de diagnóstico complementarias	13
Tratamiento de Datos Faltantes	20
Correlación entre variables	21

Introducción

El objetivo de esta primera parte del trabajo práctico es conocer los datos con que van a trabajar, realizar un análisis exploratorio y detectar y solucionar varios problemas que presentan los datos, para poder, en sucesivos trabajos, operarlos adecuadamente.

El trabajo realizado se organiza en las siguientes secciones.

Preparación de los datos

A continuación se describen las técnicas y herramientas utilizadas para preparar los datos que se pretende analizar.

La herramienta principal en la que se apoya este trabajo es R, utilizándose a través del RStudio y varios de los paquetes que provee CRAN y el propio software.

Configuración del entorno

Para comenzar, se setea el directorio de trabajo, comunmente denominado *Working Directory* de donde se obtendrán los datasets y donde se almacenarán los resultados parciales obtenidos.

El dataset describe un subconjunto de 3462 galaxias captadas en una región del cielo a través de unas 64 variables. El mismo se puede obtener desde la web de astrostatics

A efectos prácticos, se descarga el mismo al directorio de trabajo.

Análisis de los datos

Se realiza una previsualización de los datos y se revisa con qué formato R interpretó a los mismos.

```
## 'data.frame': 3462 obs. of 65 variables:
## $ Nr : int 6 9 16 21 26 29 45 49 50 51 ...
## $ Rmag : num 25 25 24.2 25.2 25.5 ...
## $ e.Rmag : num 0.097 0.181 0.054 0.128 0.112 0.056 0.257 0.217 0.098 0.097 ...
## $ ApDRmag : num 0.935 -0.135 0.821 0.639 -1.588 ...
## $ mumax : num 24.2 25.3 23.5 24.9 24.9 ...
## $ Mcz : num 0.832 0.927 1.202 0.912 0.848 ...
## $ e.Mcz : num 0.036 0.122 0.037 0.177 0.067 0.183 0.174 0.147 0.052 0.057 ...
## $ MCzml : num 1.4 0.864 1.217 0.776 1.33 ...
## $ chi2red : num 0.64 0.41 0.92 0.39 1.45 0.52 1.31 1.84 1.03 0.55 ...
## $ UjMAG : num -17.7 -18.3 -19.8 -17.8 -17.7 ...
## $ e.UjMAG : num 0.14 0.22 0.14 0.17 0.42 0.16 0.3 0.44 0.15 0.16 ...
## $ BjMAG : num -17.5 17.9 -19.9 -17.4 -18.4 ...
## $ e.BjMAG : num 0.25 0.55 0.14 0.31 0.83 1.37 1.94 1.81 0.15 0.19 ...
## $ VjMAG : num -17.8 -18.2 -20.4 -17.7 -19.4 ...
## $ e.VjMAG : num 0.25 0.55 0.14 0.31 0.83 1.37 1.94 1.81 0.32 0.14 ...
## $ usMAG : num -17.8 -18.4 -19.9 -18 -17.8 ...
## $ e.usMAG : num 0.14 0.22 0.14 0.17 0.42 0.16 0.3 0.44 0.15 0.16 ...
## $ gsMAG : num -17.6 -18 -20.1 -17.5 -18.7 ...
## $ e.gsMAG : num 0.25 0.55 0.14 0.31 0.83 1.37 1.94 1.81 0.32 0.14 ...
## $ rsMAG : num -18 -18.4 -20.7 -17.9 -19.9 ...
## $ e.rsMAG : num 0.25 0.55 0.14 0.31 0.83 1.37 1.94 1.81 0.32 0.14 ...
## $ UbMAG : num -17.8 -18.4 -19.8 -17.9 -17.8 ...
## $ e.UbMAG : num 0.14 0.22 0.14 0.17 0.42 0.16 0.3 0.44 0.15 0.16 ...
## $ BbMAG : num -17.5 -17.9 -19.9 -17.4 -18.4 ...
## $ e.BbMAG : num 0.25 0.55 0.14 0.31 0.83 1.37 1.94 1.81 0.15 0.19 ...
## $ VnMAG : num -17.8 -18.2 -20.4 -17.7 -19.4 ...
## $ e.VbMAG : num 0.25 0.55 0.14 0.31 0.83 1.37 1.94 1.81 0.32 0.14 ...
## $ S280MAG : num -18.2 -18 -19.8 -18.1 -13.9 ...
## $ e.S280MA : num 0.17 0.54 0.12 0.28 45.11 ...
## $ W420FE : num 0.00066 0.000324 0.013 0.0119 0.00135 0.00324 0.000898 0.00436 0.0144 0.02 ...
## $ e.W420FE : chr "3.85E-03" "3.19E-03" "4.11E-03" "2.70E-03" ...
## $ W462FE : num 0.0127 0.00514 0.0197 0.0159 0.00509 0.00332 0.00406 0.00116 0.0128 0.0212 ...
## $ e.W462FE : num 0.00372 0.00323 0.00432 0.00314 0.00268 0.00275 0.00265 0.00365 0.00492 0.00275 ...
## $ W485FD : num 0.0189 0.00273 0.0255 0.00156 0.00185 0.00401 0.00486 0.000102 0.00437 0.015 ...
## $ e.W485FD : num 0.00448 0.00485 0.00428 0.00493 0.00401 0.00497 0.00363 0.00389 0.00483 0.00375 ...
## $ W518FE : num 0.0182 0.000785 0.0159 0.00261 0.00996 0.00166 0.00178 0.00622 0.0165 0.0098 ...
## $ e.W518FE : num 0.00355 0.00485 0.00464 0.00476 0.00432 0.00342 0.00357 0.00553 0.00461 0.00351 ...
## $ W571FS : num 0.0147 0.00991 0.0229 0.00176 0.00344 0.00446 0.00537 0.00216 0.00745 0.00941 ...
## $ e.W571FS : num 0.00301 0.00284 0.00455 0.0031 0.00448 0.00311 0.00301 0.00357 0.00459 0.00297 ...
## $ W604FE : num 0.0166 0.00905 0.0234 0.00916 0.00632 0.00451 0.00262 0.00807 0.0107 0.0135 ...
## $ e.W604FE : num 0.00409 0.00445 0.00374 0.00332 0.00366 0.00429 0.00368 0.00296 0.00433 0.00382 ...
## $ W646FD : num 0.0188 0.00298 0.0231 0.00633 -0.000184 -0.000551 0.0132 0.00628 -0.004 0.0139 ...
## $ e.W646FD : num 0.00563 0.00892 0.00667 0.00596 0.0124 0.00966 0.00644 0.0147 0.00795 0.0112 ...
## $ W696FE : num 0.0246 0.00983 0.0272 0.0123 0.00554 0.00283 0.00776 0.014 0.0175 0.0168 ...
## $ e.W696FE : num 0.00351 0.00343 0.00405 0.00248 0.00293 0.00272 0.00308 0.0116 0.00284 0.00266 ...
## $ W753FE : num 0.0245 0.0142 0.0354 0.00225 0.0162 0.0174 0.0119 0.0154 0.0193 0.00767 ...
## $ e.W753FE : num 0.00524 0.00527 0.00456 0.00692 0.00497 0.0044 0.00443 0.00608 0.00468 0.00577 ...
## $ W815FS : num 0.0216 0.0147 0.0453 0.0169 0.00676 0.00829 0.00561 0.00687 0.0207 0.0128 ...
## $ e.W815FS : num 0.00266 0.00308 0.0036 0.00276 0.00314 0.00371 0.00275 0.00357 0.00285 0.00255 ...
```

```
## $ W856FD : num 0.0244 0.0114 0.0781 0.00875 0.0102 0.0039 0.00684 0.0115 0.0205 0.00587 ...
## $ e.W856FD: num 0.00546 0.00627 0.00658 0.00672 0.0061 0.00696 0.00557 0.0102 0.00524 0.00617 ...
## $ W914FD : num 0.0377 0.0103 0.0711 0.007 0.0133 0.00485 0.0144 0.0169 0.0276 0.013 ...
## $ e.W914FD: num 0.0061 0.00646 0.00613 0.00557 0.00682 0.00563 0.00615 0.00761 0.00663 0.00664 ...
## $ W914FE : num 0.0117 0.0263 0.0641 0.00587 0.0199 0.0264 0.0185 0.0106 0.0449 0.00219 ...
## $ e.W914FE: num 0.0101 0.0148 0.0127 0.0114 0.0103 0.0097 0.00876 0.00909 0.0139 0.0115 ...
## $ UFS : num 0.0187 0.00706 0.0126 0.0141 0.00514 0.00292 0.0123 0.00691 0.00677 0.0149 ...
## $ e.UFS : num 0.00239 0.00238 0.00184 0.00186 0.0017 0.00198 0.0021 0.00181 0.00187 0.00224 ...
## $ BFS : num 0.0163 0.0042 0.0183 0.0118 0.00102 0.00329 0.00622 0.00266 0.0076 0.017 ...
## $ e.BFS : num 0.00129 0.00115 0.00115 0.0011 0.00127 0.00104 0.00124 0.00137 0.00125 0.00109 ...
## $ VFD : num 1.73e-02 3.93e-03 1.88e-02 9.67e-03 3.85e-05 3.55e-03 5.04e-03 1.20e-04 8.59e-03 1.46e-03 ...
## $ e.VFD : num 0.00141 0.00182 0.00167 0.00204 0.0016 0.0013 0.00129 0.00158 0.00172 0.0017 ...
## $ RFS : num 0.0165 0.00723 0.0288 0.0105 0.00139 0.00474 0.00398 0.00162 0.0116 0.0122 ...
## $ e.RFS : num 0.000434 0.0005 0.000655 0.000416 0.000499 0.000489 0.000429 0.000552 0.000495 0.000495 ...
## $ IFD : num 0.0247 0.00973 0.057 0.0134 0.0059 0.00356 0.00271 0.00232 0.0164 0.0113 ...
## $ e.IFD : num 0.00483 0.0046 0.00465 0.0033 0.00444 0.00446 0.0048 0.00385 0.00444 0.00316 ...
```

Se aprecia que los datos corresponden mayormente a variables numéricas, dado que se incluyen estimadores del tamaño de la galaxia, de su corrimiento al rojo y datos y flujos y magnitudes absolutas obtenidas a diferentes longitudes de onda.

Pero una de las variables, e.W420FE fue importada como tipo caracter.

```
## chr [1:3462] "3.85E-03" "3.19E-03" "4.11E-03" "2.70E-03" ...
```

Por el tipo de nomenclatura utilizada, se puede deducir que se trata de un error y que debería ser considerada como variable numérica.

Se intuye que la causa del problema podría ser la notación científica esta especificada con la letra mayúscula “E”, dado que comparada con la variable VFD, se observa que la misma presenta la notación científica con la letra “e” minúscula.

```
## num [1:3462] 1.73e-02 3.93e-03 1.88e-02 9.67e-03 3.85e-05 3.55e-03 5.04e-03 1.20e-04 8.59e-03 1.46e-03 ...
```

Por lo tanto se procede a intentar convertirla al tipo numérico.

Se obtiene un Warning que indica que hubo valores que no pudo convertir, los cuales dejó como NA o nulos.

Se imprime un resumen para identificar cuál fue el volumen de registros que no consiguió convertir y cuáles son los valores de dichos registros.

```
##
## FALSE TRUE
## 3319 143
```

```
## [1] "3 3.100E-03" "3 3.089E-03" "2 1.296E-02" "3 2.858E-03"
## [5] "4 2.855E-03" "2 2.452E-02" "5 3.528E-03" "3 3.510E-03"
## [9] "3 1.961E-02" "5 3.780E-03" "4 2.387E-02" "4 3.050E-03"
## [13] "2 2.358E-02" "4 1.651E-02" "2 2.539E-02" "3 1.769E-02"
## [17] "4 2.895E-03" "3 3.554E-03" "3 1.264E-02" "2 2.717E-02"
## [21] "3 3.325E-03" "3 2.924E-03" "3 2.937E-03" "3 5.001E-03"
## [25] "3 4.780E-03" "4 2.840E-03" "4 3.076E-03" "3 3.039E-03"
## [29] "3 3.594E-03" "3 1.809E-02" "3 3.171E-03" "3 2.813E-03"
## [33] "3 3.745E-03" "3 3.603E-03" "5 3.105E-03" "3 2.865E-03"
```

```
## [37] "4 3.412E-03" "3 2.876E-03" "2 2.482E-02" "3 3.127E-03"
## [41] "4 3.826E-03" "3 3.026E-03" "3 5.249E-03" "2 2.309E-02"
## [45] "3 3.732E-03" "2 2.779E-02" "3 2.803E-03" "3 2.878E-03"
## [49] "3 3.848E-03" "3 3.246E-03" "3 9.861E-03" "4 3.337E-03"
## [53] "3 3.297E-03" "3 1.653E-02" "4 3.020E-03" "3 3.339E-03"
## [57] "3 2.891E-03" "2 2.414E-02" "3 3.448E-03" "3 2.714E-03"
## [61] "2 2.472E-02" "3 4.322E-03" "4 3.013E-03" "4 3.322E-03"
## [65] "4 2.977E-03" "3 4.217E-03" "2 2.886E-02" "3 3.107E-03"
## [69] "4 2.909E-03" "3 3.490E-03" "3 3.572E-03" "3 1.633E-02"
## [73] "3 5.927E-03" "4 2.939E-03" "3 4.587E-03" "3 3.545E-03"
## [77] "4 3.958E-03" "3 9.544E-03" "3 2.904E-03" "3 3.281E-03"
## [81] "2 1.948E-02" "3 3.152E-03" "4 3.269E-03" "4 4.315E-03"
## [85] "3 1.948E-02" "3 2.583E-02" "3 3.084E-03" "3 3.126E-03"
## [89] "3 3.332E-03" "4 3.653E-03" "3 3.048E-03" "3 3.560E-03"
## [93] "3 3.897E-03" "3 5.059E-03" "5 3.195E-03" "2 2.710E-02"
## [97] "2 2.659E-02" "3 2.969E-02" "3 7.649E-03" "3 3.046E-03"
## [101] "4 2.995E-03" "2 7.257E-03" "3 1.239E-02" "3 2.897E-03"
## [105] "3 1.576E-02" "3 4.371E-03" "3 1.721E-02" "2 2.477E-02"
## [109] "4 3.390E-03" "2 2.571E-02" "3 3.725E-03" "2 1.743E-02"
## [113] "3 3.201E-03" "2 3.566E-02" "3 3.371E-03" "3 2.995E-03"
## [117] "3 6.868E-03" "4 3.187E-03" "2 2.149E-02" "3 4.409E-03"
## [121] "3 3.013E-03" "3 3.342E-03" "3 3.422E-03" "4 3.700E-03"
## [125] "4 3.142E-03" "3 1.281E-02" "2 1.682E-02" "3 3.700E-03"
## [129] "4 3.468E-03" "2 2.686E-02" "4 3.782E-03" "2 2.835E-02"
## [133] "2 2.564E-02" "4 4.254E-03" "4 4.365E-03" "3 3.825E-03"
## [137] "4 2.928E-03" "2 1.586E-02" "2 2.082E-02" "4 3.282E-03"
## [141] "3 2.973E-03" "3 3.138E-03" "3 4.022E-03"
```

Por simple inspección se puede apreciar que los 143 datos que no fueron convertidos a “numeric”, poseen el siguiente patrón en su estructura.

`dígito + 2 espacios + numero en notación exponencial`

El equipo consensuó que por el rango esperado de la variable, se debe tomar como criterio válido que la porción correspondiente al llamado `numero en notación exponencial` es el correcto valor de la misma.

Se aplica la función `strsplit` para dividir el campo utilizando los espacios centrales. La función devuelve una lista con el `dígito` en la primera posición y el `número_notación_exponencial` en la segunda. Para obtener el segundo elemento de esa lista, dado que no es recomendable utilizar ciclos FOR en el lenguaje R, se utiliza la función `sapply`.

La misma proviene de la familia de funciones:

- `apply`
- `lapply`
- `sapply`

Una vez obtenido el dato para la variable, se repite el procedimiento de conversión a `numeric`, del cual ya no se obtienen errores o warnings.

Por lo tanto se procede a reemplazar los datos originales de la variable por los datos transformados, para poder operarlo correctamente en los sucesivos análisis, comprobándose que ahora los datos son del tipo `numeric` y no se observan valores NA.

```
## num [1:3462] 0.00385 0.00319 0.00411 0.0027 0.00371 0.00302 0.00288 0.00426 0.00384 0.00283 ...
```

```
##  
## FALSE  
## 3462
```

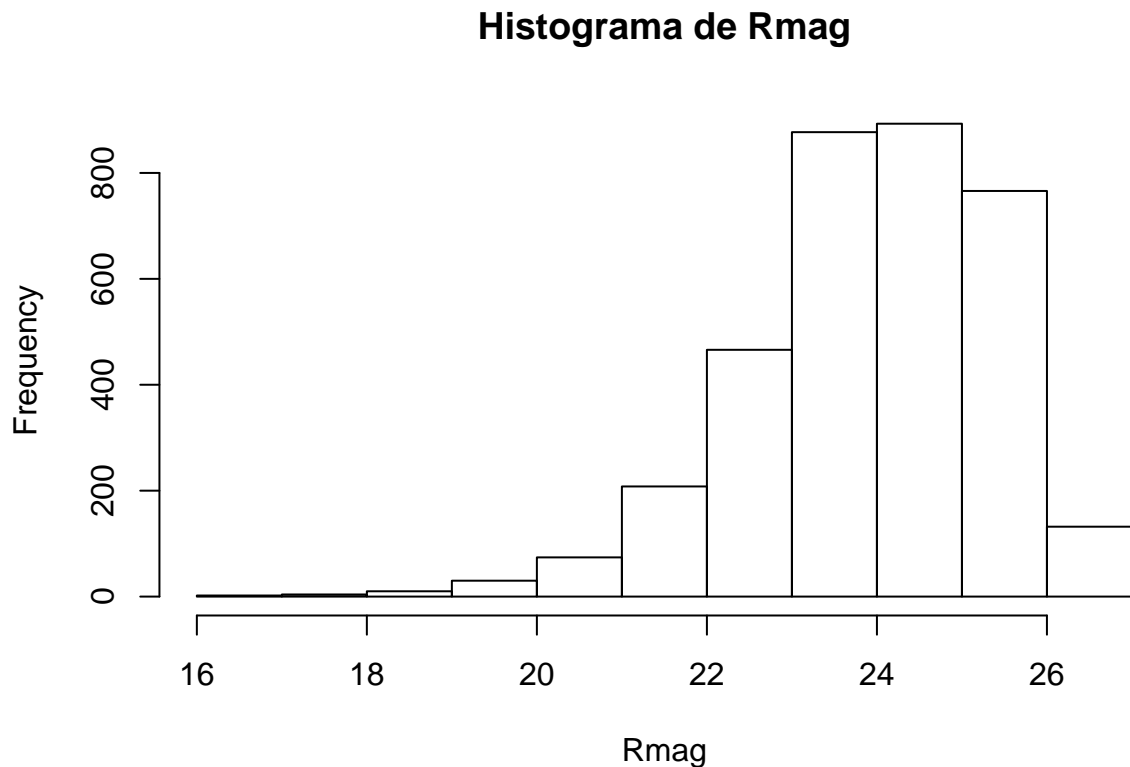
Tratamiento de Outliers

Con el objetivo de identificar la presencia de valores extremos, se realizan controles mediante la observación de tablas y gráficos exploratorios que ayuden a visualizarlos.

RMag

Se comienza analizando las variables `Rmag` y `e.Rmag`, que representan la magnitud total en la banda R y su error de medición.

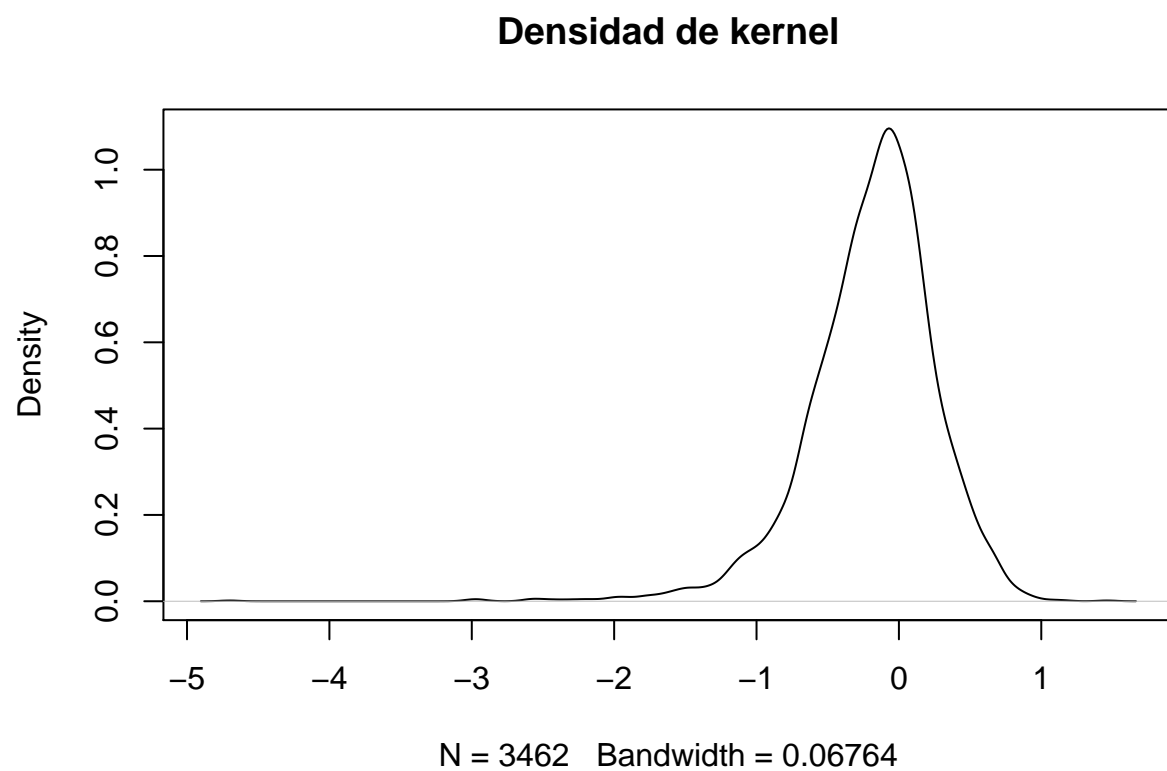
Se analiza la variable `Rmag` por medio de un histograma y un gráfico de barras de error de `e.Rmag`.



Uno de los problemas observados al estudiar los histogramas son los rangos o la cantidad de clases de la variable, los cuales son determinados de forma empírica. Otro inconveniente es que el ancho de clase debe tenerse en consideración para distribuir las observaciones. Pero debe hacerse especial hincapié en el rango desde y hasta dónde se desea visualizar.

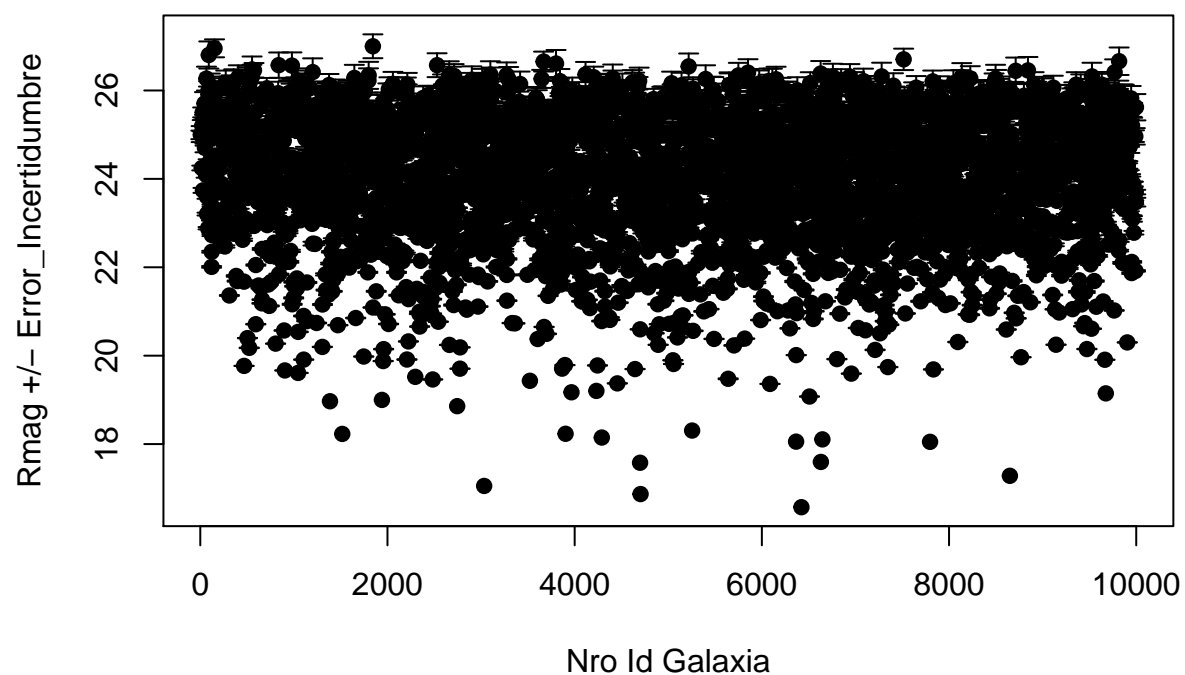
En oposición a los anteriores, suelen ser preferibles los gráficos de densidad de kernel, los cuales son una función continua y pueden someterse a una derivada. Además, son muy recomendables cuando se tienen muchos registros.

Por lo tanto, realizamos un gráfico simple de densidad de kernel.



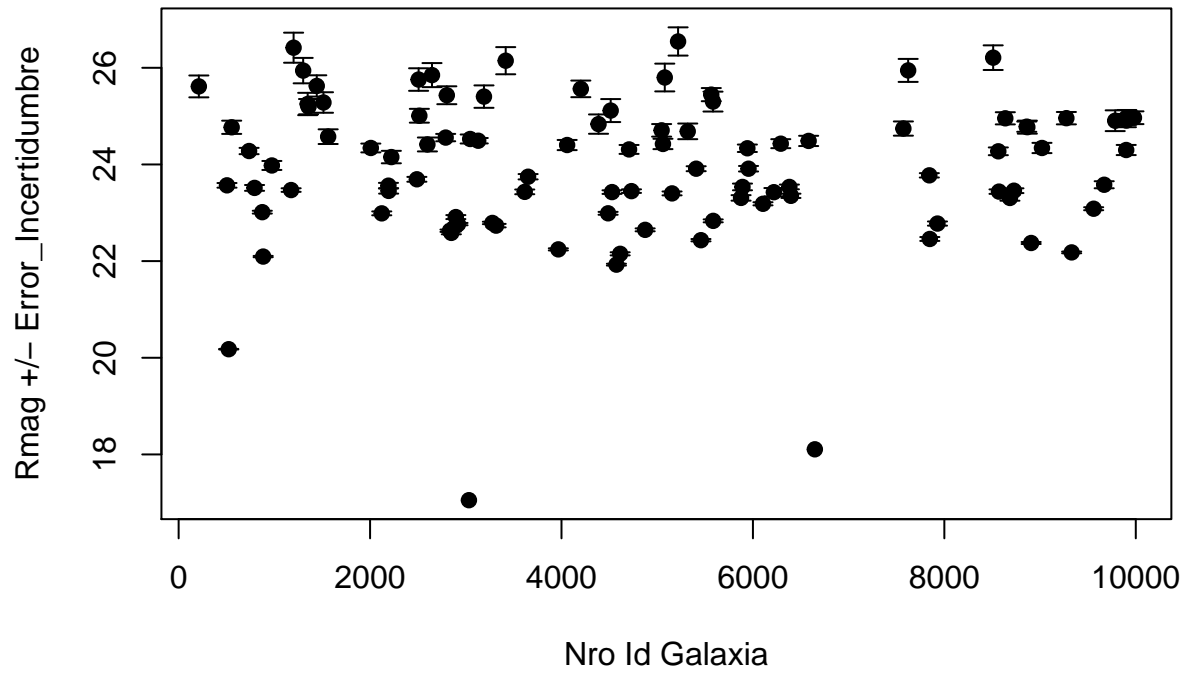
Adicionalmente, se grafica un gráfico de barras de error, a través del cual se espera tener una mejor visualización de los datos.

Grafico barras de Error_Incertidumbre



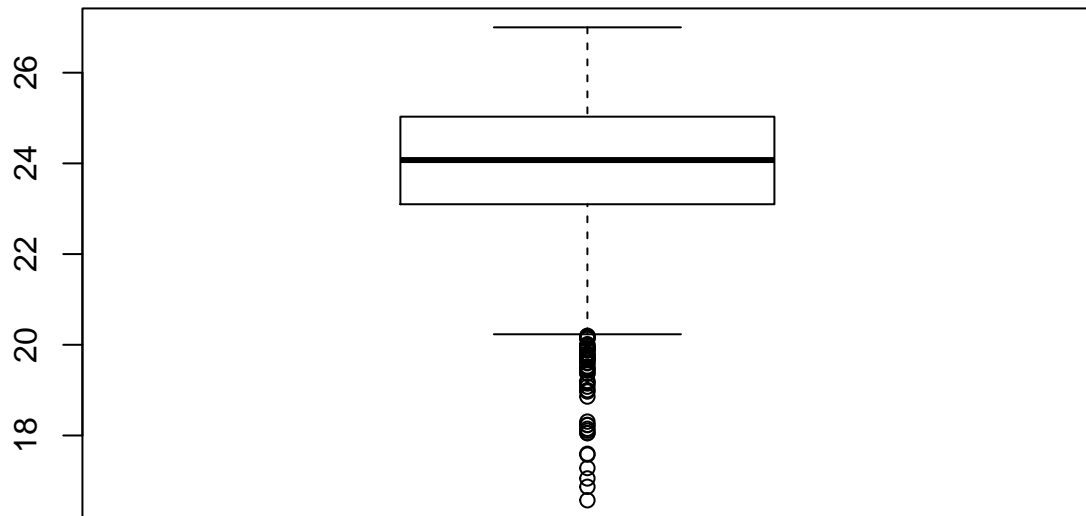
Dada la gran cantidad de registros, es difícil visualizar el comportamiento de los mismos. Se procede a realizar el mismo gráficos pero tomado una muestra al azar de 100 registros.

Grafico barras de Error_Incertidumbre



A partir de los gráficos expuestos se puede suponer la presencia de outliers.

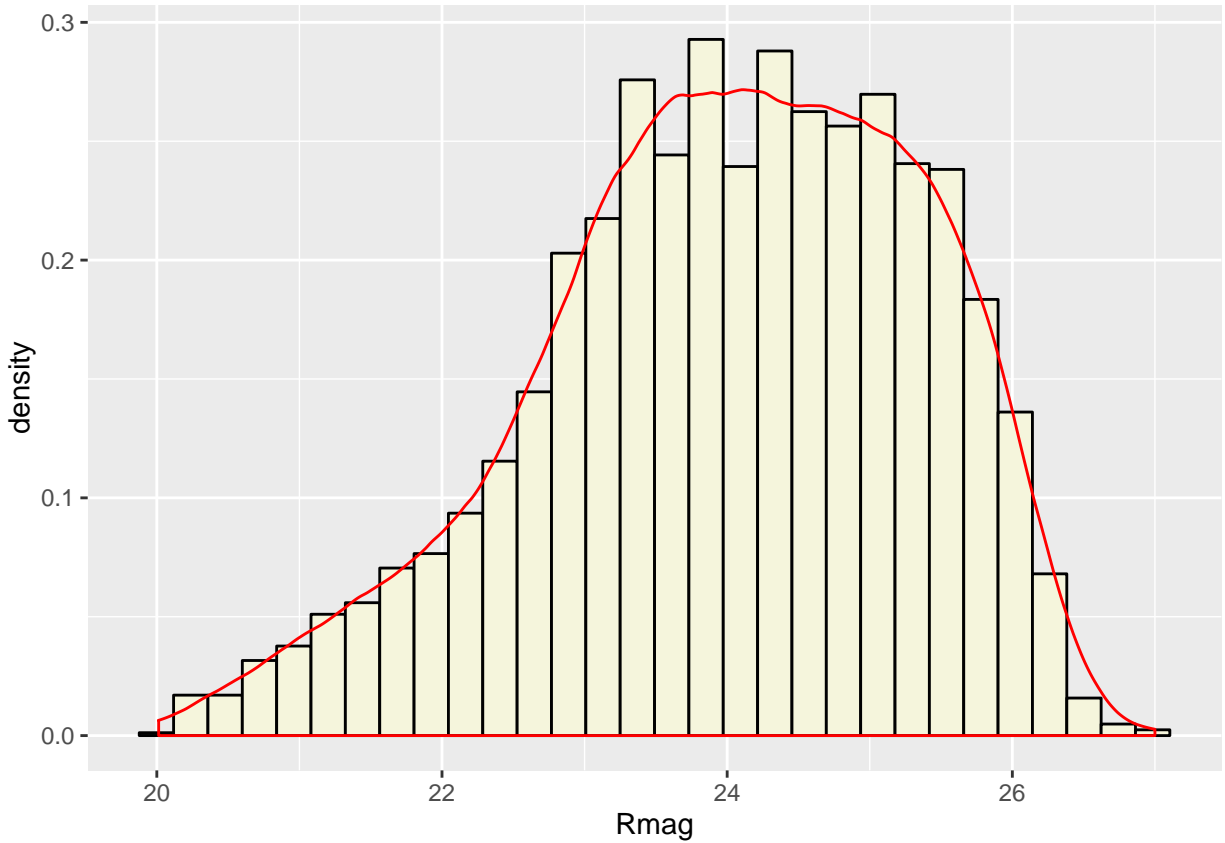
Luego, se presenta un gráfico tipo boxplot para visualizar la distribución de la variable.



El boxplot presenta de manera gráfica el valor de la media, 1er y 3er cuartil y los (supuestos) outliers. Pero debe tenerse en consideración que el diagrama de cajas fue pensado para trabajar con datos en los años 70. Hoy en día, los grandes volúmenes de datos que se manejan hacen más presente la idea de poder tener (supuestos) outliers.

Sin embargo, a partir de las observaciones realizadas con todo el análisis anterior, se determina que los valores extremos para la variable en estudio podrían ser aquellos mayores o iguales a 20.

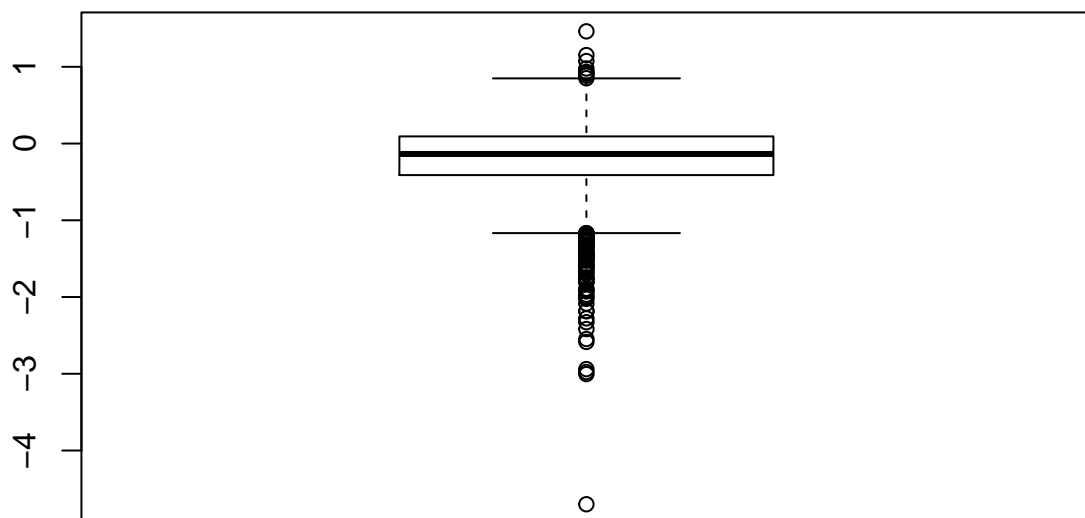
Por lo tanto, se procede a quitar las observaciones que toman valores en este rango y se vuelve a generar un gráfico de barras que muestra la distribución de la variable sin valores extremos.



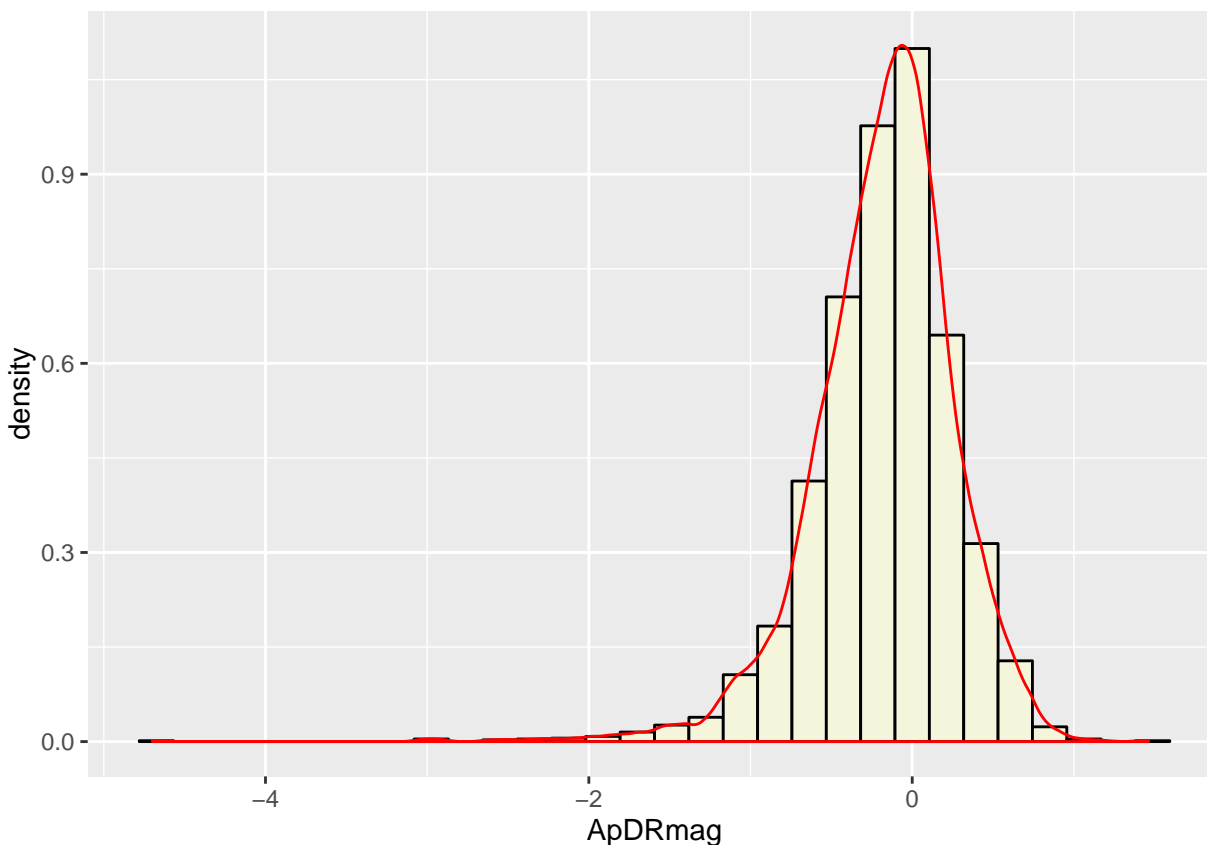
ApDRMag

Pasemos a analizar la segunda variable de interés, **ApDRmag**, que representa la diferencia entre la magnitud total y la magnitud de apertura medidas en la banda R. Es un estimador del tamaño de una galaxia. Un valor de cero indica que es una fuente puntual de luz. Los valores negativos se generan por errores de medición.

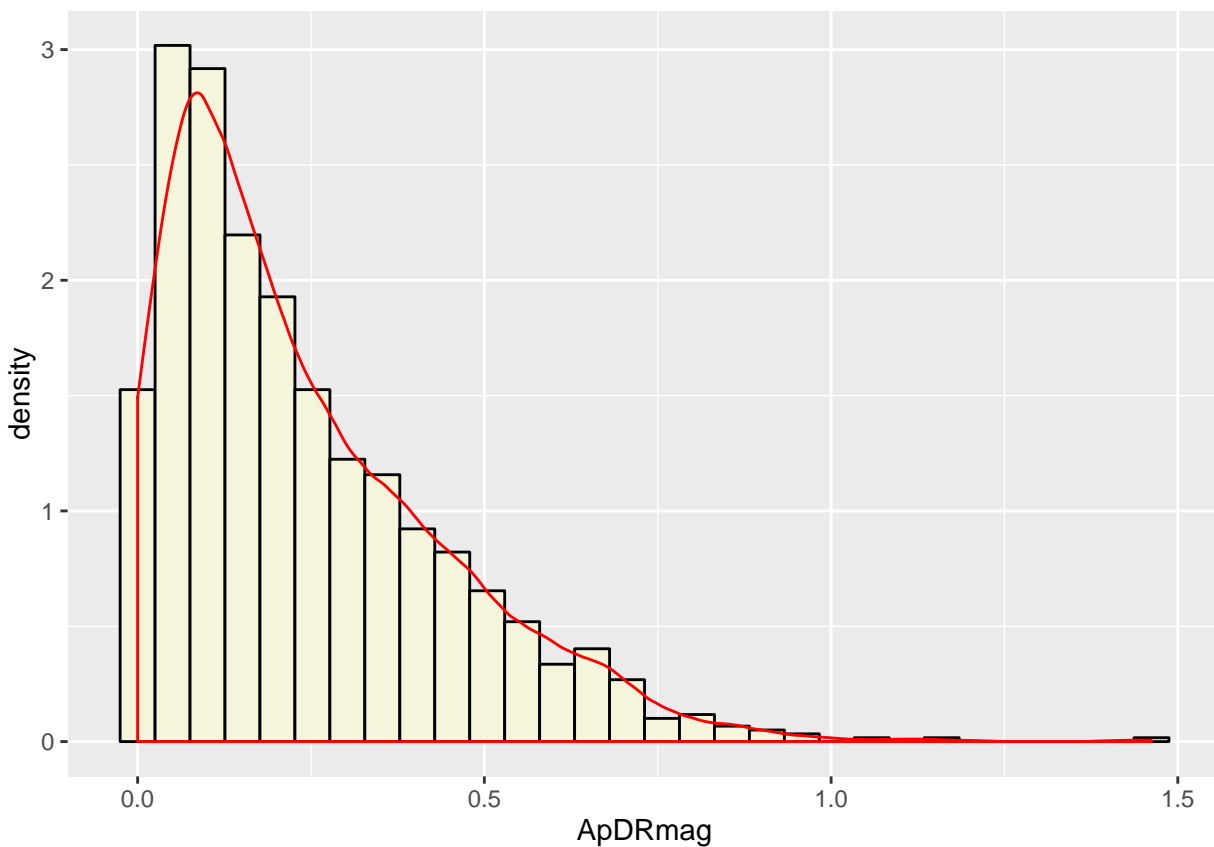
Se grafica mediante un boxplot y un histograma:



El histograma se crea con la librería `ggplot`^[1]

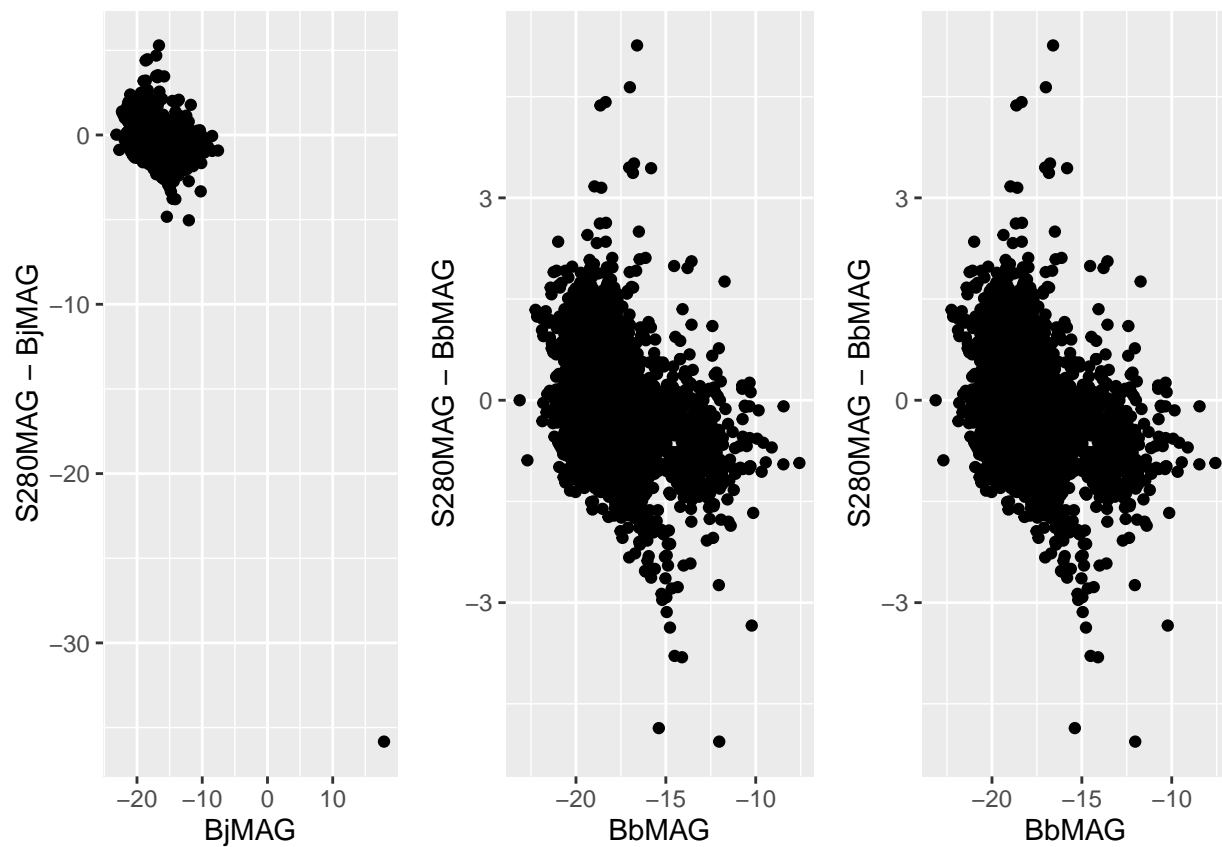


[1] Como se puede apreciar, la librería ggplot es mas vistosa para realizar gráficos. Los parámetros de “aes” están relacionados con la estética por *aesthetics*, de las variables que se quieren representar. Luego + `geom_histogram()` propone la forma en la que se representan los datos. Los signos + permiten generar los gráficos en capas; en este caso la segunda capa esta referida + `geom_density()` Cabe también destacar que la nomenclatura utilizada: `x = ApDRmag` en vez de: `x = glx$ApDRmag` es consistente con la filosofía “Tidy Data” de su creador, Hadley Wickham, que menciona en su paper publicado en “The Journal of Statistical Software, vol. 59, 2014”



Otros gráficos y técnicas de diagnóstico complementarias

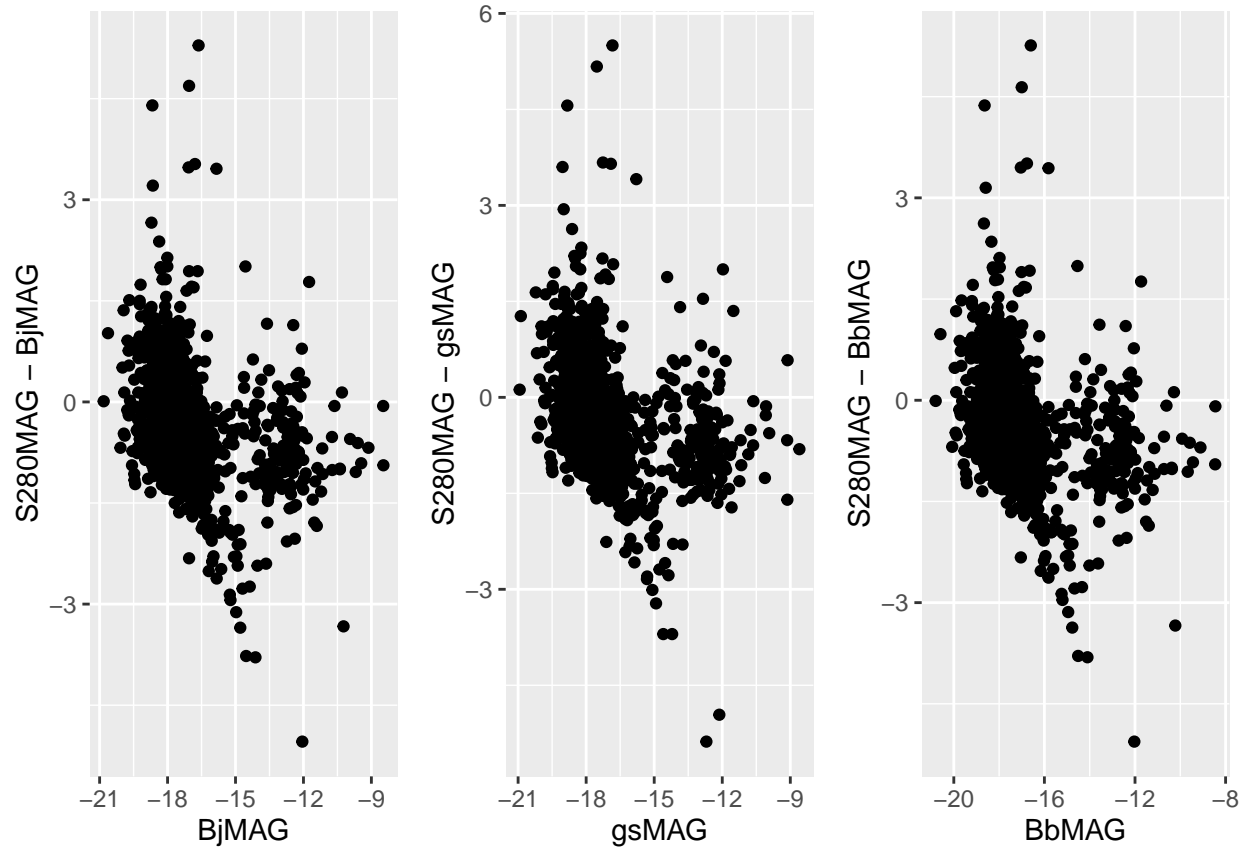
Para el análisis de los datos de las bandas espectrales, se realizarán scatter graphs donde la abcisa será la banda espectral de casa sistema fotométrico y la ordenada estará normalizada respecto de la magnitud absoluta en el ultravioleta a 280 nm **S280MAG**. La normalización se hace restando porque estas variables son logarítmicas.



Es fácil apreciar que existe un valor en BjMAG muy alejado de la nube de puntos. Procedemos a eliminarlo.

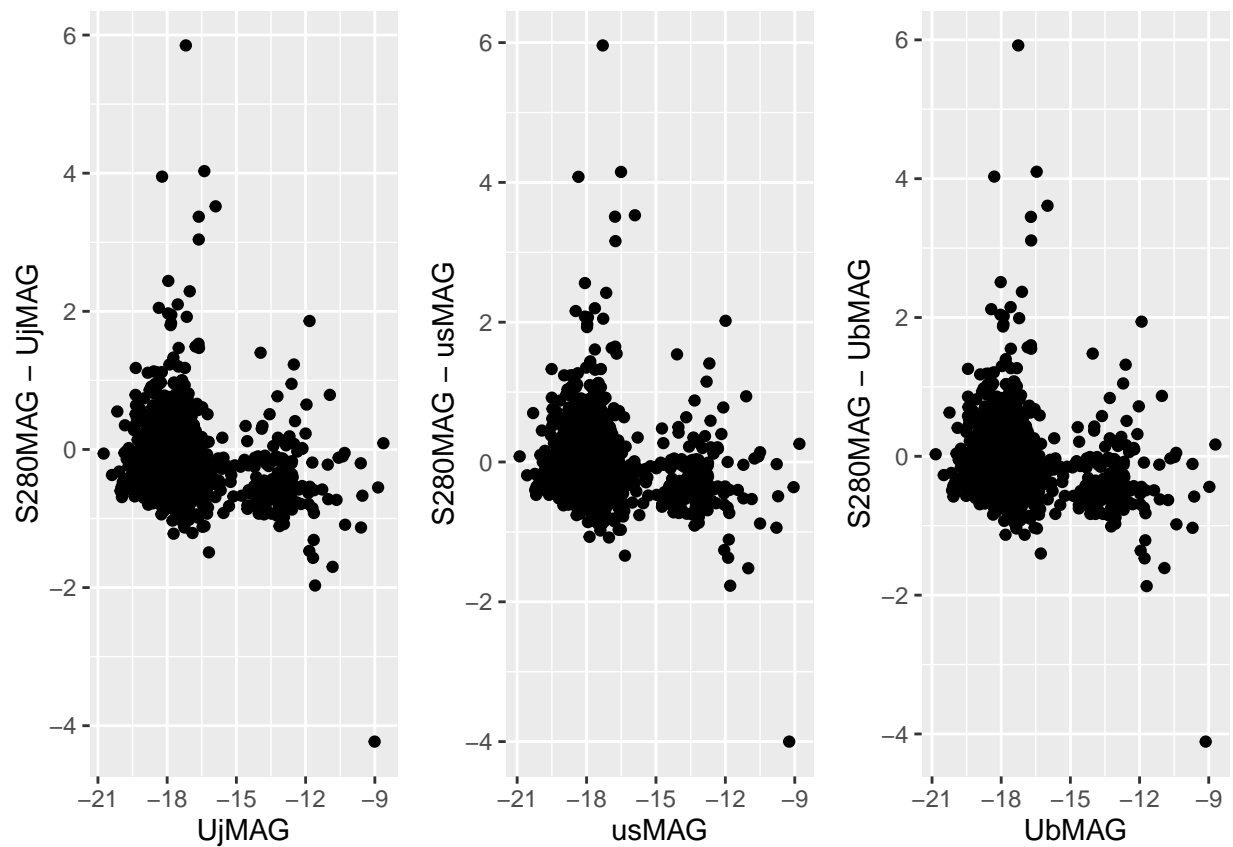
```
## [1] 2
```

```
## [1] 17.86
```



Luego de la remoción, los tres gráficos presentan la similitud esperada en la distribución.

Bajo la misma técnica se analizan las otras bandas espectrales.



Se procede a eliminar los puntos anómalos. En este caso se tomó el criterio que las bandas normalizadas se deben encontrar dentro del intervalo $[-3;+3]$

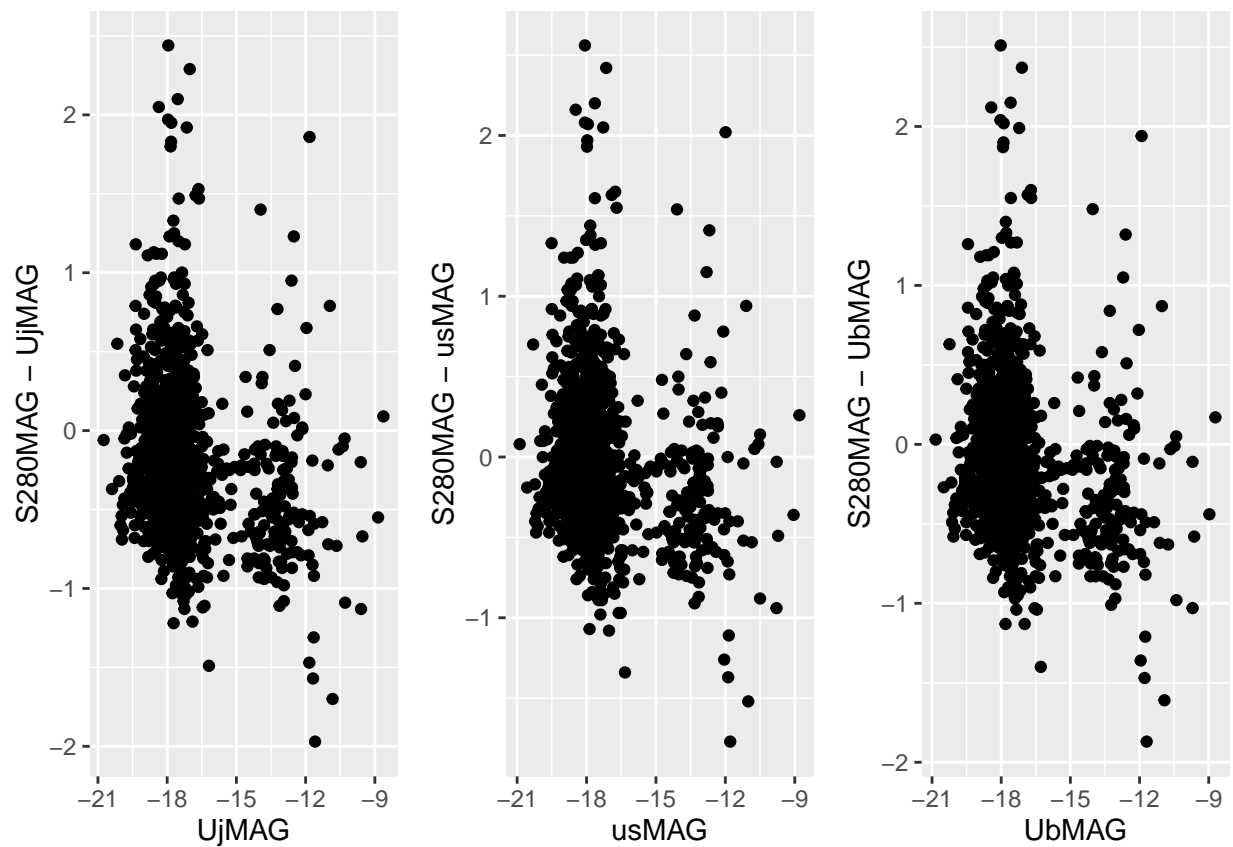
```
## [1] 977
```

```
## [1] 977
```

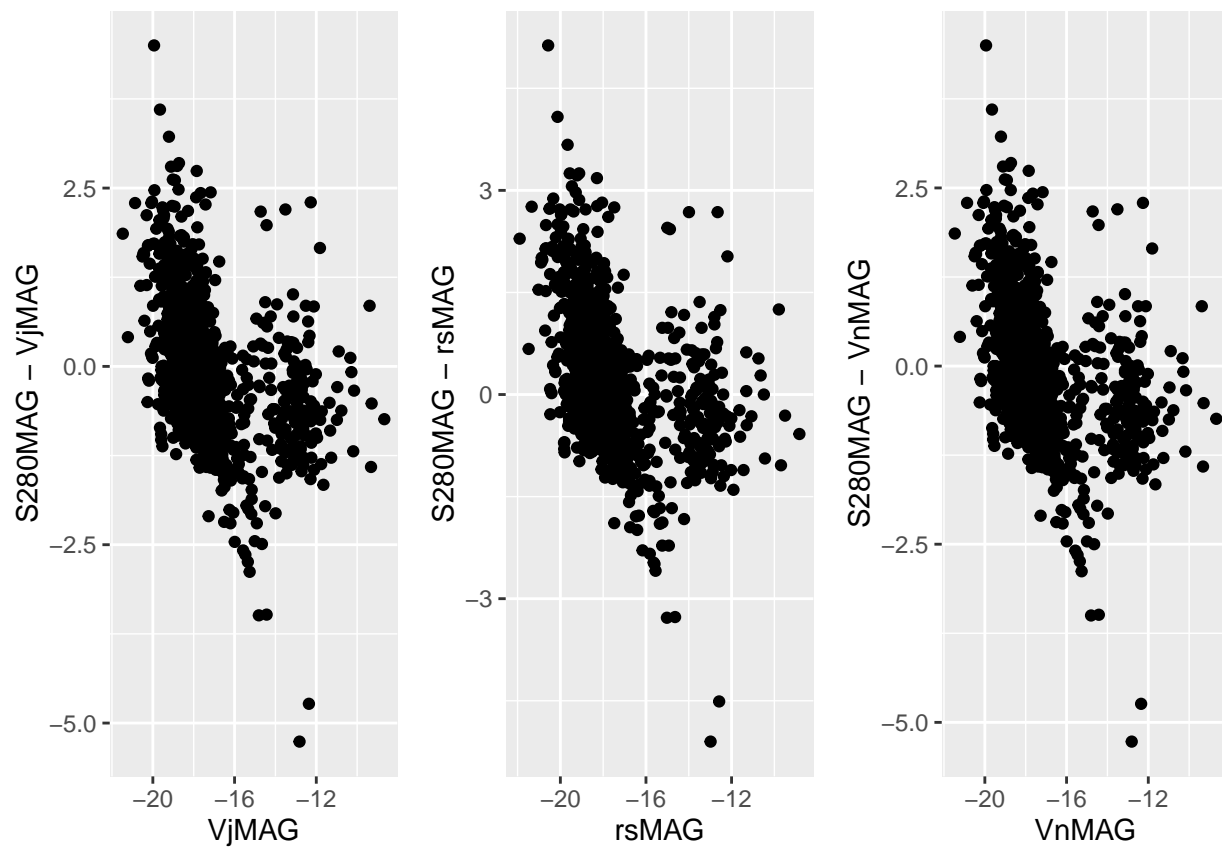
```
## [1] 1162
```

```
## [1] 1162
```

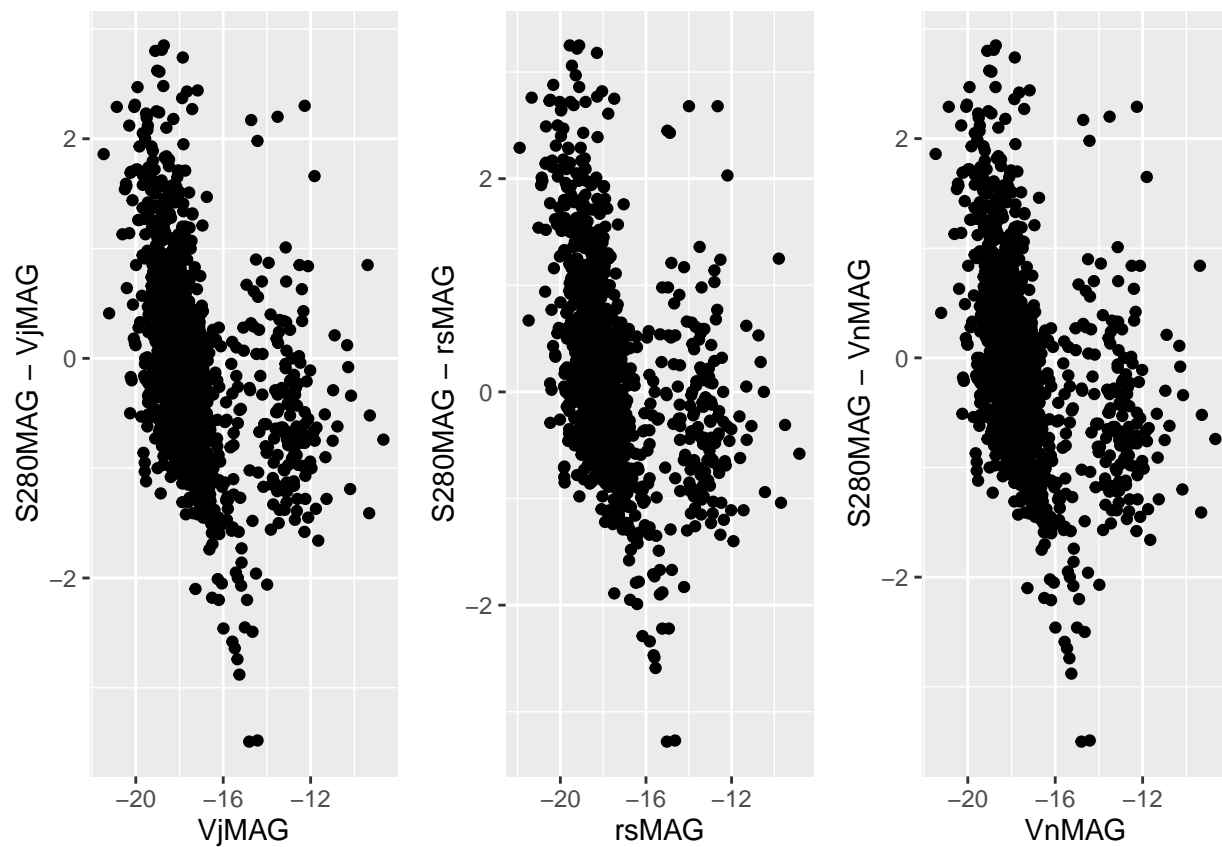
```
## [1] 89 299 414 477 866 977 1183
```

Se vuelve a generar el gráfico y por inspección visual se eliminan los puntos anómalos.



```
## [1] 57 246 352 852 303
```

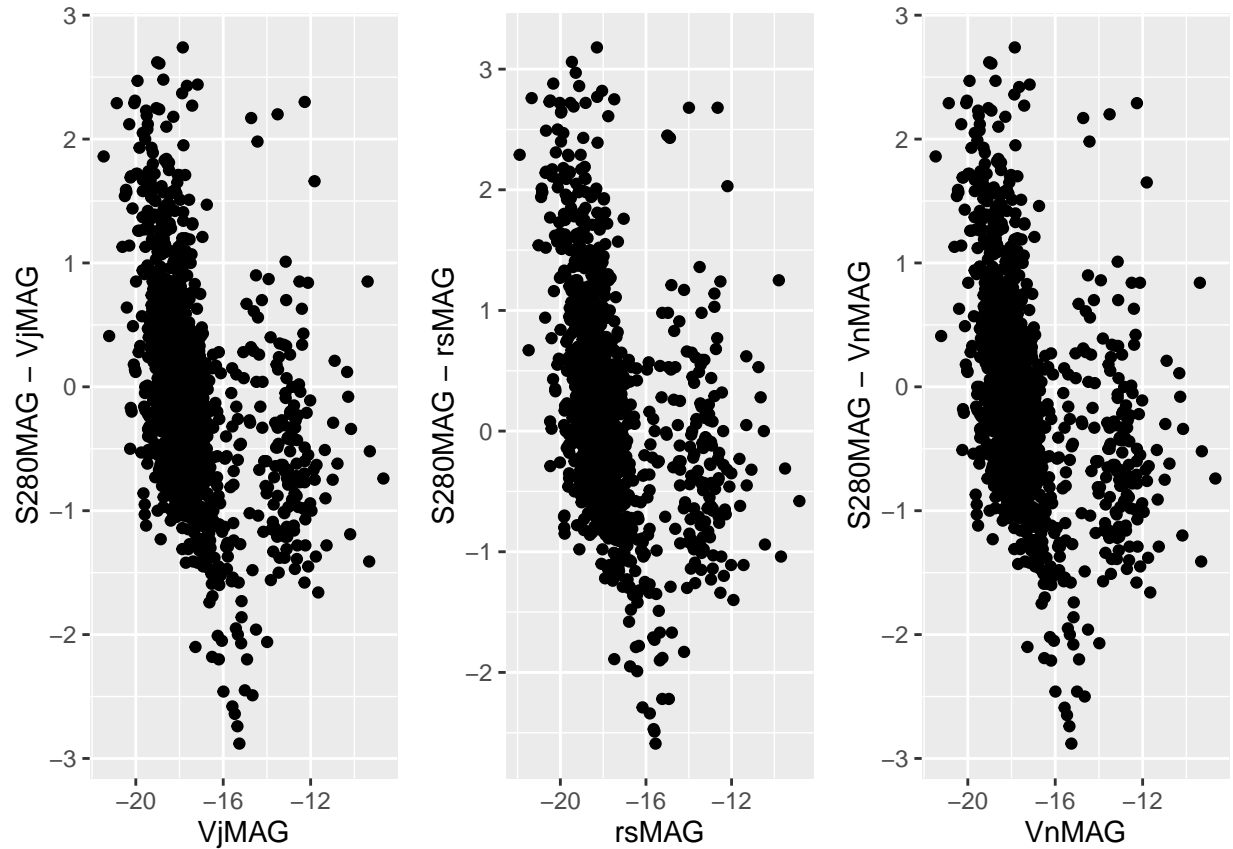


```
## [1] 133
```

```
## [1] -3.28
```

```
## [1] 133 165 498 853 989
```

Finalmente se llega a distribuciones de puntos similares para los distintos sistemas fotométricos en las distintas bandas espectrales.



Tratamiento de Datos Faltantes

Se verifica la presencia de datos faltantes en cualquiera de las variables del dataset.

El resultado indica que hay datos faltantes en 50 registros y que los mismos se distribuyen en 4 variables, como sigue:

```
##
## FALSE TRUE
## 75766 24

## [1] "VnMAG" "e.VbMAG" "S280MAG" "e.S280MA"
```

El equipo considera que la cantidad de casos con datos faltantes es despreciable, por lo que se procede a eliminar dichas observaciones para asegurar que se trabajará con datos completos y sin rellenar.

Se comprueba que luego de retirados dichos casos, el dataset obtenido ya no posee datos faltantes:

```
##
## FALSE TRUE
## 75506 24
```

Correlación entre variables

A continuación se realiza el análisis de correlación utilizando únicamente las variables que representan las bandas espectrales de las galaxias para diferentes sistemas de fotometría. Para ello, se combinan diferentes versiones del dataset para comprender su impacto en el estudio de la correlación.

En primer lugar, se muestra la matriz de correlación para los datos con valores nulos (NAs)

```
##           UjMAG      BjMAG      VjMAG      usMAG      gsMAG      rsMAG
## UjMAG      1.0000000 0.9611492 0.9501758 0.9999191 0.9568434 0.9441076
## BjMAG      0.9611492 1.0000000 0.9912516 0.9603186 0.9942098 0.9879420
## VjMAG      0.9501758 0.9912516 1.0000000 0.9487141 0.9981145 0.9993278
## usMAG      0.9999191 0.9603186 0.9487141 1.0000000 0.9558897 0.9424294
## gsMAG      0.9568434 0.9942098 0.9981145 0.9558897 1.0000000 0.9955537
## rsMAG      0.9441076 0.9879420 0.9993278 0.9424294 0.9955537 1.0000000
## UbMAG      0.9999835 0.9607499 0.9494133 0.9999535 0.9563639 0.9431966
## BbMAG      0.9609629 0.9997852 0.9906800 0.9601538 0.9939725 0.9871475
## VnMAG      NA        NA        NA        NA        NA        NA
## S280MAG     NA        NA        NA        NA        NA        NA
##           UbMAG      BbMAG VnMAG S280MAG
## UjMAG      0.9999835 0.9609629  NA      NA
## BjMAG      0.9607499 0.9997852  NA      NA
## VjMAG      0.9494133 0.9906800  NA      NA
## usMAG      0.9999535 0.9601538  NA      NA
## gsMAG      0.9563639 0.9939725  NA      NA
## rsMAG      0.9431966 0.9871475  NA      NA
## UbMAG      1.0000000 0.9606191  NA      NA
## BbMAG      0.9606191 1.0000000  NA      NA
## VnMAG      NA        NA        1      NA
## S280MAG     NA        NA        NA      1
```

Luego se recalcula la matriz de correlación pero esta vez quitando los valores nulos mediante la técnica *pairwise*, que alternará entre filas y columnas retirando aquellas que en cada variable presenten valores nulos.

```
##           UjMAG      BjMAG      VjMAG      usMAG      gsMAG      rsMAG
## UjMAG      1.0000000 0.9611492 0.9501758 0.9999191 0.9568434 0.9441076
## BjMAG      0.9611492 1.0000000 0.9912516 0.9603186 0.9942098 0.9879420
## VjMAG      0.9501758 0.9912516 1.0000000 0.9487141 0.9981145 0.9993278
## usMAG      0.9999191 0.9603186 0.9487141 1.0000000 0.9558897 0.9424294
## gsMAG      0.9568434 0.9942098 0.9981145 0.9558897 1.0000000 0.9955537
## rsMAG      0.9441076 0.9879420 0.9993278 0.9424294 0.9955537 1.0000000
## UbMAG      0.9999835 0.9607499 0.9494133 0.9999535 0.9563639 0.9431966
## BbMAG      0.9609629 0.9997852 0.9906800 0.9601538 0.9939725 0.9871475
## VnMAG      0.9499047 0.9911787 0.9999969 0.9484332 0.9980988 0.9993558
## S280MAG     0.9670750 0.9245633 0.9019804 0.9684306 0.9171050 0.8916078
##           UbMAG      BbMAG      VnMAG      S280MAG
## UjMAG      0.9999835 0.9609629 0.9499047 0.9670750
## BjMAG      0.9607499 0.9997852 0.9911787 0.9245633
## VjMAG      0.9494133 0.9906800 0.9999969 0.9019804
## usMAG      0.9999535 0.9601538 0.9484332 0.9684306
## gsMAG      0.9563639 0.9939725 0.9980988 0.9171050
## rsMAG      0.9431966 0.9871475 0.9993558 0.8916078
## UbMAG      1.0000000 0.9606191 0.9491350 0.9677775
## BbMAG      0.9606191 1.0000000 0.9908959 0.9253085
```

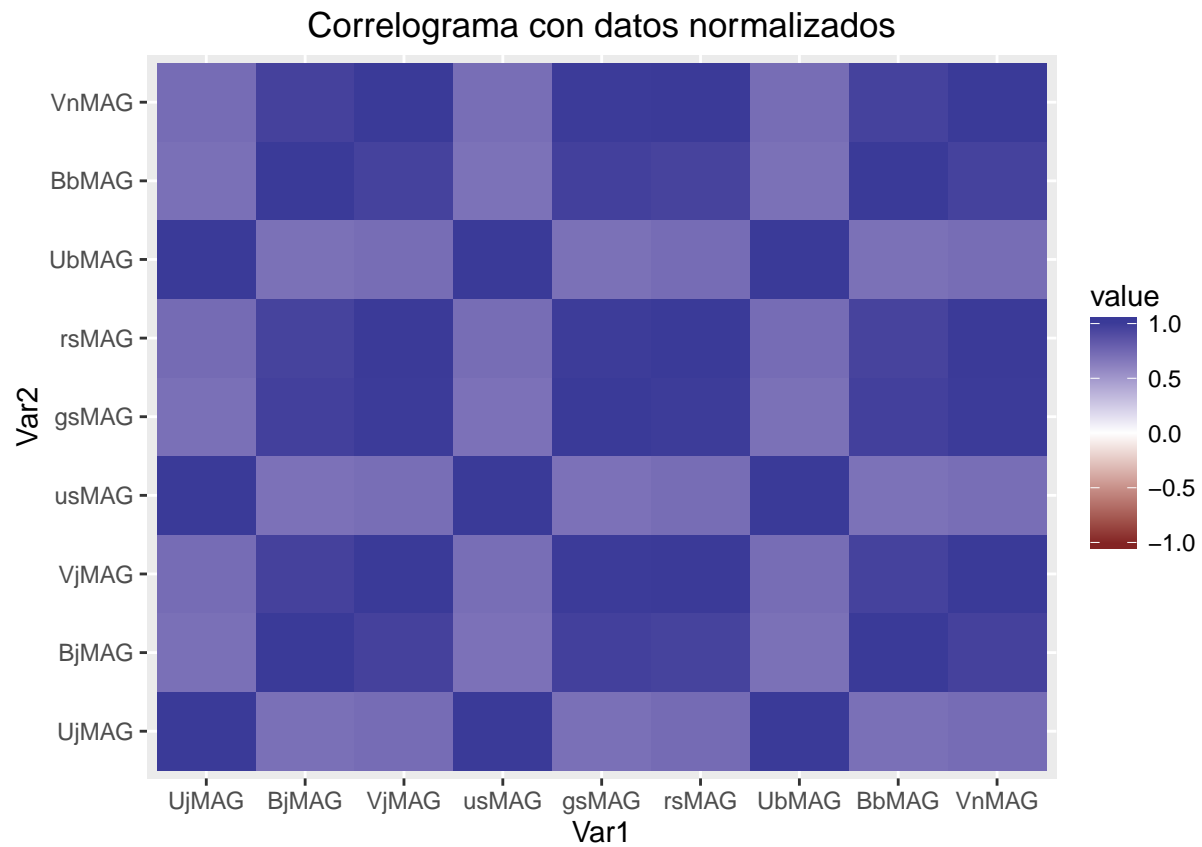
```
## VnMAG    0.9491350 0.9908959 1.0000000 0.9018503
## S280MAG 0.9677775 0.9253085 0.9018503 1.0000000
```

Finalmente, se muestra la misma matriz pero para los datos normalizados restando la variable S280Mag

```
##          UjMAG      BjMAG      VjMAG      usMAG      gsMAG      rsMAG
## UjMAG    1.0000000 0.7011085 0.7213481 0.9988352 0.7000065 0.7251835
## BjMAG    0.7011085 1.0000000 0.9571801 0.6940487 0.9634326 0.9501318
## VjMAG    0.7213481 0.9571801 1.0000000 0.7116118 0.9934702 0.9978634
## usMAG    0.9988352 0.6940487 0.7116118 1.0000000 0.6926719 0.7153144
## gsMAG    0.7000065 0.9634326 0.9934702 0.6926719 1.0000000 0.9865927
## rsMAG    0.7251835 0.9501318 0.9978634 0.7153144 0.9865927 1.0000000
## UbMAG    0.9998334 0.6976789 0.7163573 0.9992732 0.6962411 0.7198007
## BbMAG    0.6993684 0.9999595 0.9560297 0.6924564 0.9630081 0.9486464
## VnMAG    0.7213863 0.9570375 0.9999837 0.7116283 0.9933743 0.9979055
## S280MAG      NA        NA        NA        NA        NA        NA
##          UbMAG      BbMAG      VnMAG S280MAG
## UjMAG    0.9998334 0.6993684 0.7213863      NA
## BjMAG    0.6976789 0.9999595 0.9570375      NA
## VjMAG    0.7163573 0.9560297 0.9999837      NA
## usMAG    0.9992732 0.6924564 0.7116283      NA
## gsMAG    0.6962411 0.9630081 0.9933743      NA
## rsMAG    0.7198007 0.9486464 0.9979055      NA
## UbMAG    1.0000000 0.6960179 0.7163873      NA
## BbMAG    0.6960179 1.0000000 0.9558808      NA
## VnMAG    0.7163873 0.9558808 1.0000000      NA
## S280MAG      NA        NA        NA      NA
```

Luego, se presentan los gráficos tipo *heatmap* para las matrices sin normalizar y normalizadas, a modo de ofrecer una comparativa visual del efecto de la última por sobre los resultados obtenidos en la anterior.





Se puede observar que, normalizando por la variable de la magnitud absoluta, aquellas variables que presentan mayor correlación se separan del resto.