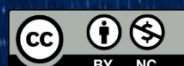


Guide

Guía de Usuario del paquete de R #rstats {MATdatatools}, Versión 0.1.0

Miguel Ángel Tarancón Morán

Catedrático de Economía Aplicada. Universidad de Castilla – La Mancha



MATdatatools Guide

Miguel-Ángel Tarancón

2025-01-15

Tabla de contenidos



1	Introducción	3
2	Funciones de MATdatatools	3
3	Descarga e Instalación	3
3.1	Desde GitHub	3
3.2	En RStudio	4
4	Detalle de cada Función	4
4.1	MATfexcel	4
4.1.1	Uso	4
4.1.2	Objetivo	4
4.1.3	Argumentos	4
4.1.4	Salida	4
4.1.5	Ejemplo	5
4.1.6	Resultados Esperados	5
4.2	MATmv	6
4.2.1	Uso	6
4.2.2	Objetivo	6
4.2.3	Argumentos	6
4.2.4	Salida	6
4.2.5	Ejemplo	6
4.2.6	Resultados Esperados	6
4.3	MATout	8
4.3.1	Uso	8
4.3.2	Objetivo	8
4.3.3	Argumentos	8
4.3.4	Salida	8
4.3.5	Ejemplo	8
4.3.6	Resultados Esperados	8

4.4	MATout_Mahalanobis	11
4.4.1	Uso	11
4.4.2	Objetivo	11
4.4.3	Argumentos	11
4.4.4	Salida	11
4.4.5	Ejemplo	11
4.4.6	Resultados Esperados	11
4.5	MATdescribe	13
4.5.1	Uso	13
4.5.2	Objetivo	13
4.5.3	Argumentos	13
4.5.4	Salida	14
4.5.5	Ejemplo	14
4.5.6	Resultados Esperados	14
4.6	MATtaf	15
4.6.1	Uso	15
4.6.2	Objetivo	15
4.6.3	Argumentos	15
4.6.4	Salida	15
4.6.5	Ejemplo	15
4.6.6	Resultados Esperados	15
5	Documentación del Paquete	16
6	¡Empieza a explorar tus datos con MATdatatools!	17

1 Introducción

El paquete **MATdatatools** ha sido diseñado para facilitar el análisis de datos y la generación de resultados gráficos y estadísticos en R. Este paquete es una herramienta ágil y accesible, especialmente útil para estudiantes y profesionales de Administración y Dirección de Empresas que buscan una manera intuitiva de explorar, analizar y visualizar sus datos sin necesidad de profundos conocimientos de programación.

3

¿Qué ofrece MATdatatools?

- Automatización de procesos estadísticos comunes.
- Gráficos de alta calidad con interpretaciones claras.
- Análisis descriptivo univariante.

2 Funciones de MATdatatools

El paquete incluye las siguientes funciones principales:

- **MATfexcel:** Importa datos desde hojas de Excel, convirtiendo fácilmente las primeras columnas en nombres de filas.
- **MATmv:** Filtra datos con casos completos, muestra gráficamente los datos faltantes y genera un resumen.
- **MATout:** Detecta y elimina valores atípicos en una variable, proporcionando gráficos y tablas explicativas.
- **MATout_Mahalanobis:** Identifica valores atípicos multidimensionales basándose en la distancia de Mahalanobis.
- **MATdescribe:** Realiza un análisis descriptivo completo, incluyendo gráficos y estadísticos.
- **MATtaf:** Genera tablas de frecuencias agrupadas en intervalos y crea histogramas relacionados.

Cada función está diseñada para ser intuitiva y generar resultados listos para interpretar. El archivo de Microsoft® Excel® de los ejemplos, `eolica_100_mv.xls`, puede descargarse [aquí](#).

3 Descarga e Instalación

3.1 Desde GitHub

1. Asegúrate de tener instalado el paquete devtools:

```
install.packages("devtools")
```

2. Descarga e instala el paquete desde GitHub:

```
devtools::install_github("teckel71/R_for_Economics/packages/MATdata  
tools")
```

3. Carga el paquete:

```
library(MATdatatools)
```

4

3.2 En RStudio

1. Abre RStudio y ve a **Tools > Install Packages**.
2. Selecciona **Install from GitHub** y escribe:

```
teckel71/R_for_Economics/packages/MATdatatools
```

3. Haz clic en **Install**.

4 Detalle de cada Función

4.1 MATfexcel

4.1.1 Uso

```
MATfexcel(file_path, sheet_name, na_values = NULL, viz = FALSE)
```

4.1.2 Objetivo

Esta función carga datos desde un archivo Excel, convirtiendo la primera columna en nombres de fila y generando un resumen.

4.1.3 Argumentos

- `file_path`: Ruta al archivo Excel.
- `sheet_name`: Nombre o índice de la hoja a importar.
- `na_values`: Vector de caracteres que representan valores NA.
- `viz`: Si es TRUE, genera una visualización automática de las variables. Es necesario que no existan NAs.

4.1.4 Salida

- Dataframe con los datos importados.
- Si `viz = TRUE`, genera un objeto con visualización de las variables (`<nombre_dataframe>_viz`). Es necesario que no existan NAs.

4.1.5 Ejemplo

```
rm(list = ls())  
library(MATdatatools)  
eolica_100 <- MATfexcel("eolica_100_mv.xlsx", "Datos",  
                        na_values = c("n.d.", "s.d."),  
                        viz = FALSE)
```

5

4.1.6 Resultados Esperados

- Un dataframe llamado eolica_100

```
##      RES      ACTIVO      FPIOS  
## Min.   : -5661.5   Min.   : 24944   Min.   : -77533  
## 1st Qu.: 669.5     1st Qu.: 34547   1st Qu.: 2305  
## Median : 2084.5    Median : 46950   Median : 11936  
## Mean   : 11529.8   Mean   : 277270   Mean   : 123743  
## 3rd Qu.: 3806.7    3rd Qu.: 85610   3rd Qu.: 28292  
## Max.   :727548.0   Max.   :13492812   Max.   :6904824  
## NA's   :1         NA's   :1  
##  
##      RENECO      RENFIN      LIQUIDEZ  
## Min.   : -2.813   Min.   : -359.773   Min.   : 0.0140  
## 1st Qu.: 1.558     1st Qu.: 2.556     1st Qu.: 0.6567  
## Median : 4.236     Median : 15.326     Median : 1.0650  
## Mean   : 5.416     Mean   : 17.243     Mean   : 2.7214  
## 3rd Qu.: 7.970     3rd Qu.: 31.307     3rd Qu.: 1.6078  
## Max.   :35.262     Max.   : 588.190     Max.   :128.4330  
## NA's   :2  
##  
##      ENDEUDA      MARGEN      SOLVENCIA  
## Min.   : 0.917     Min.   : -2248.157   Min.   : -40.74  
## 1st Qu.: 50.852     1st Qu.: 12.316     1st Qu.: 4.71  
## Median : 83.346     Median : 26.618     Median : 16.65  
## Mean   : 72.227     Mean   : 3.228       Mean   : 27.57  
## 3rd Qu.: 95.388     3rd Qu.: 39.590     3rd Qu.: 45.59  
## Max.   :140.745     Max.   : 400.899     Max.   : 99.08  
## NA's   :2         NA's   :2  
##  
##      APALANCA      MATRIZ      DIMENSION  
## Min.   : -8254.11   Length:100         Length:100  
## 1st Qu.: 16.13      Class :character    Class :character  
## Median : 161.97     Mode  :character     Mode  :character  
## Mean   : 345.03  
## 3rd Qu.: 623.13  
## Max.   :12244.35
```

4.2 MATmv

4.2.1 Uso

```
MATmv(dataframe, columnas)
```

4.2.2 Objetivo

Selecciona variables de un dataframe, identifica casos completos y genera un resumen de valores faltantes.

6

4.2.3 Argumentos

- **dataframe**: El dataframe que se desea analizar.
- **columnas**: Las columnas (variables) que se desean evaluar, separadas por comas (sin comillas). Si no se especifican las variables, se procesarán todas las variables del dataframe.

4.2.4 Salida

- Un nuevo dataframe con los casos completos de las columnas seleccionadas.
- Una lista con:
 - Tabla de valores faltantes.
 - Gráfico de valores ausentes.

4.2.5 Ejemplo

```
MATmv(eolica_100, c(RENECO, ACTIVO, RES, RENFIN, FPIOS, MARGEN))
```

4.2.6 Resultados Esperados

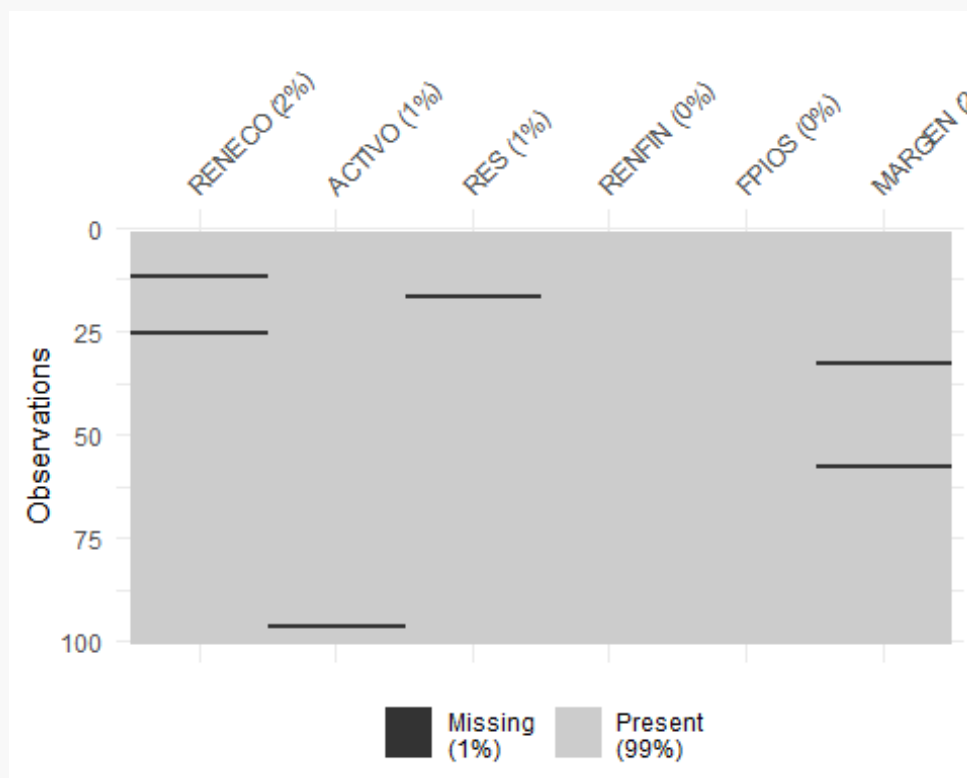
1. Un dataframe llamado `eolica_100_sm` con los casos completos en las variables seleccionadas.
2. Una lista `eolica_100_sm_info` que contiene:
 - Gráfico que visualiza los valores faltantes.
 - Tabla con los casos faltantes.

```
##          RES          ACTIVO          FPIOS
## Min.   : -5661.5  Min.    : 24944  Min.    : -77533
## 1st Qu.:  739.6   1st Qu.: 34249   1st Qu.:  2380
## Median : 2114.7   Median : 46653   Median : 11936
## Mean   : 12080.6  Mean    : 285091  Mean    : 128818
## 3rd Qu.: 3844.2   3rd Qu.: 83091   3rd Qu.: 27991
## Max.   :727548.0  Max.    :13492812 Max.    :6904824
##
```



```
##      RENECO      RENFIN      LIQUIDEZ
## Min.   :-2.813   Min.    :-359.773   Min.    :  0.0140
## 1st Qu.: 1.558   1st Qu.:   2.212   1st Qu.:  0.6675
## Median : 4.236   Median :  15.924   Median :  1.0795
## Mean   : 5.478   Mean     : 17.906   Mean     :  2.8401
## 3rd Qu.: 8.107   3rd Qu.:  34.167   3rd Qu.:  1.6350
## Max.   :35.262   Max.     : 588.190   Max.     :128.4330
##
##      ENDEUDA      MARGEN      SOLVENCIA
## Min.    :  0.917   Min.    :-2248.157   Min.    :-40.745
## 1st Qu.: 54.406   1st Qu.:  12.793   1st Qu.:   4.779
## Median : 83.346   Median :   27.638   Median : 16.653
## Mean    : 72.002   Mean     :   9.393   Mean     : 27.773
## 3rd Qu.: 95.289   3rd Qu.:  41.264   3rd Qu.: 43.812
## Max.    :140.745   Max.     : 400.899   Max.     : 99.082
## NA's    :2
##
##      APALANCA      MATRIZ      DIMENSION
## Min.    :-8254.11   Length:94          Length:94
## 1st Qu.:  25.71     Class :character   Class :character
## Median : 223.21     Mode  :character   Mode  :character
## Mean    : 378.36
## 3rd Qu.: 670.30
## Max.    :12244.35
```

eolica_100_sm_info\$grafico_vis_miss



eolica_100_sm_info\$tabla_na

Casos con datos faltantes

	RENECO	ACTIVO	RES	RENFIN	FPIOS	MARGEN
Viesgo Renovables SL.	NA	269730.00	4609.000	3.200	177707.000	11.818
Biovent Energía SA	4.551	183899.00	NA	11.952	70033.000	22.792
Sargon Energías SLU	NA	85745.00	-2216.000	26.900	-10985.000	-615.625
Parc Eolic Sant Antoni SL	1.361	69654.00	668.000	9.746	9727.000	NA
Eolica La Brujula SA	7.295	42146.98	2306.062	14.174	21694.791	NA
La Caldera Energía Burgos SL	2.643	NA	511.304	-24.857	-2752.605	14.448

8

4.3 MATout

4.3.1 Uso

MATout(data, variable)

4.3.2 Objetivo

Detectar valores atípicos unidimensionales basados en el rango intercuartílico (IQR).

4.3.3 Argumentos

- data: El dataframe que contiene los datos.
- variable: La variable numérica a analizar (sin comillas).

4.3.4 Salida

- Identifica valores atípicos y los muestra en un boxplot.
- Filtra los datos sin valores atípicos.

4.3.5 Ejemplo

MATout(data = eolica_100_sm, variable = RENECO)

4.3.6 Resultados Esperados

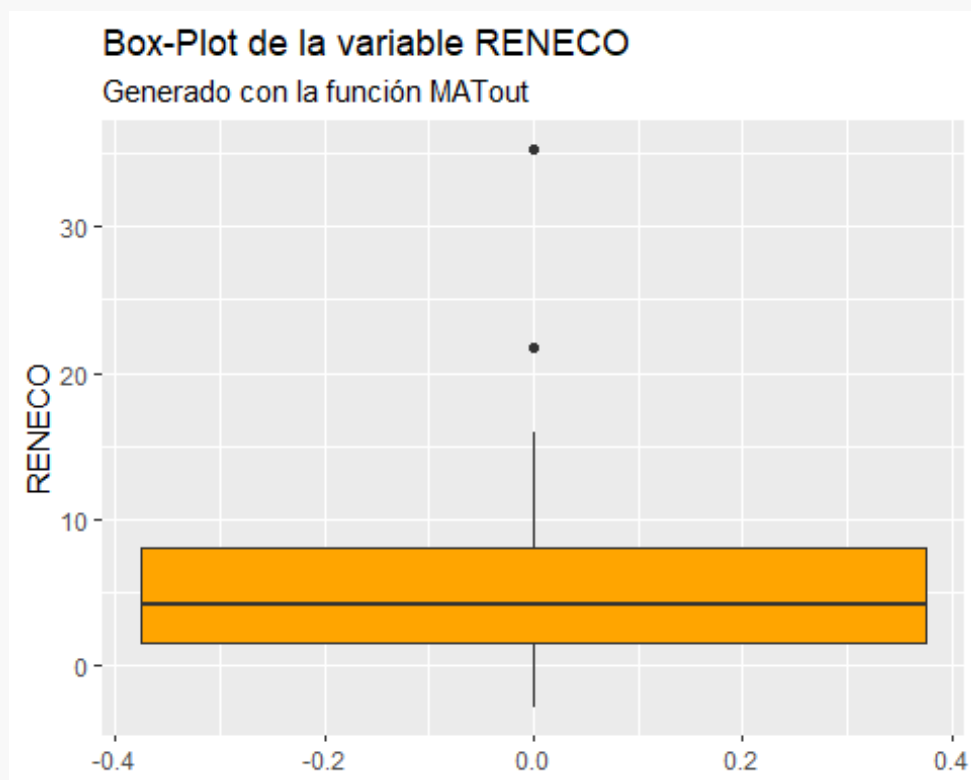
1. Un dataframe llamado eolica_100_sm_so con los datos filtrados.
2. Una lista eolica_100_sm_so_info que contiene:
 - Box-Plot de la variable analizada.
 - Tabla de valores atípicos.

```

##          RES              ACTIVO              FPIOS
## Min.    : -5661.5   Min.    :   24944   Min.    : -77533
## 1st Qu.:   718.8   1st Qu.:   34437   1st Qu.:   2305
## Median :   2016.7   Median :   46653   Median :   10870
## Mean    :  12109.0   Mean    :  290312   Mean    :  131112
## 3rd Qu.:   3666.9   3rd Qu.:   84061   3rd Qu.:   28292
## Max.    :727548.0   Max.    :13492812   Max.    :6904824
##
##          RENECO              RENFIN              LIQUIDEZ
## Min.    : -2.813   Min.    : -359.773   Min.    :   0.0140
## 1st Qu.:   1.421   1st Qu.:    1.951   1st Qu.:   0.6915
## Median :   4.144   Median :   15.460   Median :   1.1115
## Mean    :   4.977   Mean    :   17.078   Mean    :   2.8952
## 3rd Qu.:   7.904   3rd Qu.:   33.163   3rd Qu.:   1.6567
## Max.    :15.882   Max.    :  588.190   Max.    :128.4330
##
##          ENDEUDA              MARGEN              SOLVENCIA
## Min.    :   0.917   Min.    : -2248.157   Min.    : -40.74
## 1st Qu.:  56.187   1st Qu.:   12.493   1st Qu.:   4.71
## Median :  83.648   Median :   27.409   Median :  16.35
## Mean    :  72.646   Mean    :    8.631   Mean    :  27.14
## 3rd Qu.:  95.388   3rd Qu.:   39.580   3rd Qu.:  43.80
## Max.    :140.745   Max.    :  400.899   Max.    :  99.08
## NA's    :2
##
##          APALANCA              MATRIZ              DIMENSION
## Min.    : -8254.11   Length:92   Length:92
## 1st Qu.:   26.17   Class :character   Class :character
## Median :   233.75   Mode  :character   Mode  :character
## Mean    :   385.76
## 3rd Qu.:   693.49
## Max.    :12244.35

```

eolica_100_sm_so_info\$Boxplot



10

eolica_100_sm_so_info\$Outliers_Table

Casos considerados outliers

RENEKO	
Molinos Del Ebro SA	35.262
Sierra De Selva SL	21.761

4.4 MATout_Mahalanobis

4.4.1 Uso

```
MATout_Mahalanobis(data, variables)
```

11

4.4.2 Objetivo

Usa la distancia de Mahalanobis para detectar y eliminar outliers multivariados en un conjunto de variables.

4.4.3 Argumentos

- `data`: El dataframe que contiene los datos.
- `variables`: Variables a incluir en el análisis (sin comillas, separadas por comas).

4.4.4 Salida

- Un dataframe sin outliers (`<nombre_original>_so`).
- Una lista con:
 - Tabla de outliers detectados.
 - Boxplot de distancias de Mahalanobis.

4.4.5 Ejemplo

```
MATout_Mahalanobis(data = eolica_100_sm,  
                    variables = c(RENECO, ACTIVO, RES, RENFIN, FPIOS,  
                                  MARGEN))
```

4.4.6 Resultados Esperados

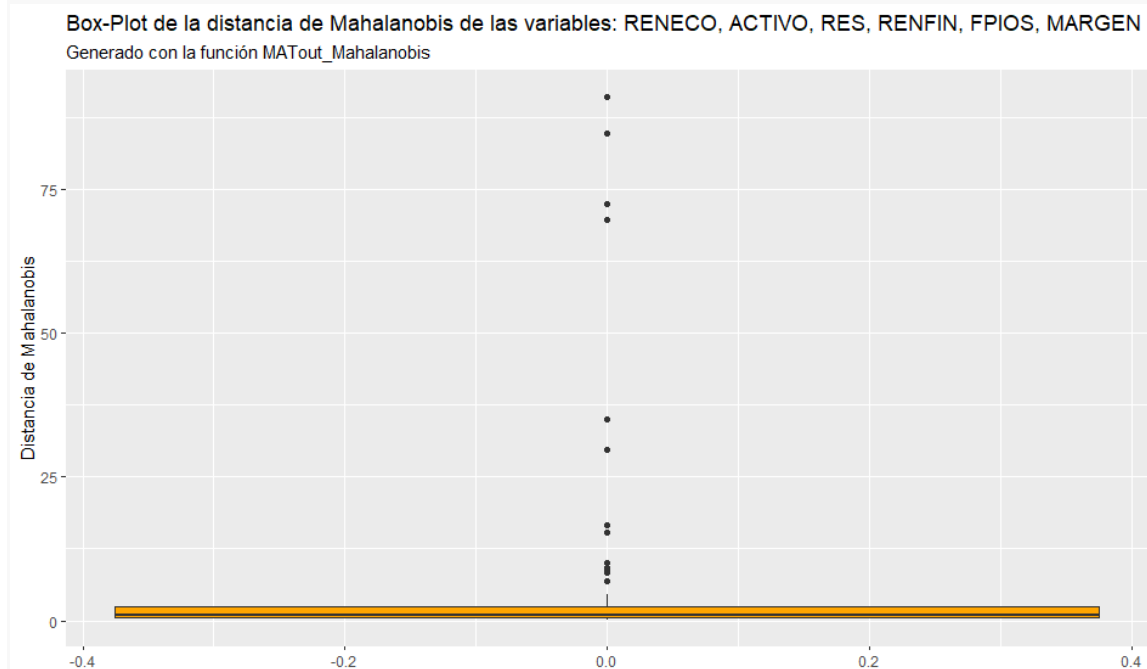
1. Un dataframe llamado `eolica_100_sm_so` con los datos filtrados.
2. Una lista `eolica_100_sm_so_info` que contiene:
 - Tabla de valores atípicos.
 - Gráficos de las distancias de Mahalanobis.

```
##      RES      ACTIVO      FPIOS
## Min.   :-5661.5   Min.    : 24944   Min.    :-77533
## 1st Qu.:  650.7   1st Qu.: 33607   1st Qu.:  2842
## Median : 1949.0   Median : 43998   Median : 10980
## Mean   : 3253.1   Mean    : 89534   Mean    : 28365
## 3rd Qu.: 3497.5   3rd Qu.: 79873   3rd Qu.: 25995
## Max.   :67033.0   Max.    :1275939   Max.    :726783
##
```



```
##      RENECO      RENFIN      LIQUIDEZ
## Min.   :-2.813   Min.    :-165.348   Min.    :  0.029
## 1st Qu.: 1.424   1st Qu.:   6.904   1st Qu.:  0.690
## Median : 4.237   Median :  16.684   Median :  1.177
## Mean   : 5.236   Mean     :  19.599   Mean     :  3.129
## 3rd Qu.: 8.560   3rd Qu.:  34.669   3rd Qu.:  1.648
## Max.   :15.882   Max.     : 207.801   Max.     :128.433
##
##      ENDEUDA      MARGEN      SOLVENCIA
## Min.    :  0.917   Min.    :-302.03   Min.    :-40.745
## 1st Qu.: 56.193   1st Qu.:  11.66   1st Qu.:   5.981
## Median : 83.098   Median :   28.52   Median :  16.901
## Mean    : 72.071   Mean     :  34.10   Mean     :  27.670
## 3rd Qu.: 94.059   3rd Qu.:  42.02   3rd Qu.:  43.796
## Max.    :140.745   Max.     :  400.90   Max.     :  99.082
## NA's    :2
##
##      APALANCA      MATRIZ      DIMENSION
## Min.    :-7770.00   Length:81   Length:81
## 1st Qu.:   31.93   Class :character   Class :character
## Median :  239.37   Mode  :character   Mode  :character
## Mean    :  435.09
## 3rd Qu.:  684.22
## Max.    : 8049.39
```

eolica_100_sm_so_info\$Boxplot



eolica_100_sm_so_info\$Outliers_Table

Casos considerados outliers

	Mahalanobis_Distance	RENECO	ACTIVO	RES	RENFIN	FPIOS	MARGEN
Holding De Negocios De GAS SL.	91.048060	5.264	13492812.00	727548.0000	10.287	6904824.0000	91.152
Global Power Generation SA.	72.365163	1.393	2002458.00	39995.0000	1.603	1740487.0000	22.403
Naturgy Renovables SLU	69.684700	1.959	1956869.00	42737.0000	12.043	318475.0000	20.442
Corporacion Acciona Eolica SL	8.471773	4.562	864606.00	29592.0000	28.990	136064.0000	20.091
Saeta Yield SA.	16.650544	0.360	796886.38	2084.4760	0.432	665319.5560	16.258
Parque Eolico Santa Catalina SL	15.261169	4.053	147742.52	3645.2780	-359.773	-1664.7550	31.780
Molinos Del Ebro SA	29.667488	35.262	62114.37	17026.2569	81.149	26991.0714	41.821
Luria De Energias SA	6.816766	4.844	49912.00	1849.0000	267.774	903.0000	20.160
Parque Eolico Sierra De Las Carbas SL	8.424602	3.676	46949.76	1252.4250	-263.639	-654.7640	19.056
Elecdey Lezuza SA	34.902738	4.657	36061.15	1258.6609	588.190	285.5387	22.346
WPD Parque Eolico Navillas SL.	84.604706	-0.416	35511.45	-110.9293	-14.302	1034.1284	-2248.157
Sierra De Selva SL	9.134843	21.761	27728.00	4525.0000	30.856	19555.0000	47.045
El Paramo Parque Eolico SL	9.970725	3.416	26634.00	671.0000	-287.974	-316.0000	16.267

13

4.5 MATdescribe

4.5.1 Uso

MATdescribe(data, variable)

4.5.2 Objetivo

Esta función realiza un análisis descriptivo y gráfico detallado de una variable numérica en un dataframe. Incluye histogramas, densidad, boxplots y pruebas de normalidad.

4.5.3 Argumentos

- **data**: El dataframe que contiene los datos.
- **variable**: La variable numérica a analizar (sin comillas).
- **bins**: Número de bins para el histograma. Si es 0, o no se especifica, se usa el método *Freedman-Diaconis* para calcular el número de intervalos de valores en el histograma.

4.5.4 Salida

Genera un objeto en el entorno global con el nombre `<variable>_describe_info`, que incluye:

- **Resumen gráfico** (histograma, densidad, boxplot, QQ-Plot).
- **Estadísticos descriptivos** (media, mediana, asimetría, curtosis, etc.).
- **Pruebas de normalidad** (Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling).

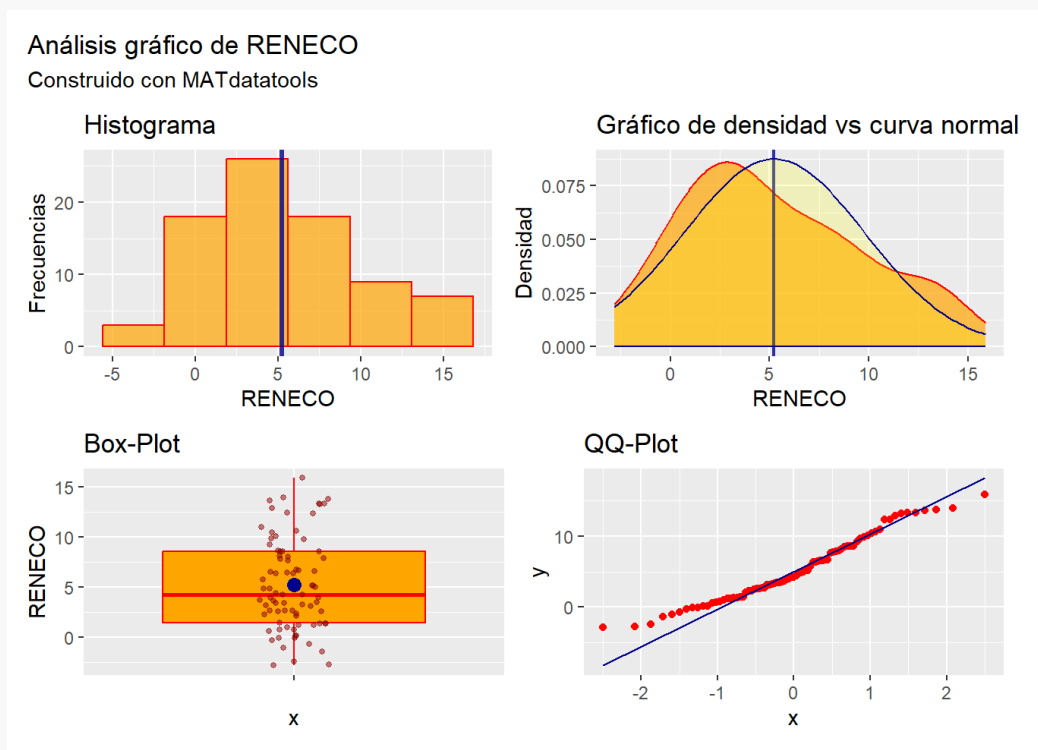
4.5.5 Ejemplo

```
MATdescribe(eolica_100_sm_so, RENECO, bins = 0)
```

4.5.6 Resultados Esperados

1. Una lista llamada `RENECO_describe_info` que contiene:
 - Gráfico resumen con varios paneles.
 - Tabla de estadísticos descriptivos.
 - Resultado de la prueba de normalidad de Shapiro-Wilk, Kolmogorov-Smirnov y Anderson-Darling (para significación de 0.05).

`RENECO_describe_info$grafico_resumen`



RENECO_describe_info\$estadisticos

Principales Estadísticos de RENECO

Media	Desviación Típica	Mediana	Valor mínimo	Valor máximo	C. Asimetría Fisher	C. Curtosis Fisher
5.236346	4.554305	4.237	-2.813	15.882	0.4115856	-0.6729009

RENECO_describe_info\$normalidad

RENECO - Pruebas de Normalidad

Prueba	p-valor	Conclusión
Shapiro-Wilk	0.026	NO-NORMALIDAD
Kolmogorov-Smirnov	0.080	NORMALIDAD
Anderson-Darling	0.024	NO-NORMALIDAD

4.6 MATtaf

4.6.1 Uso

```
MATtaf(data, variable, breaks = NULL)
```

4.6.2 Objetivo

Generar una tabla de frecuencias agrupadas en intervalos y un histograma relacionado.

4.6.3 Argumentos

- **data**: El dataframe que contiene los datos.
- **variable**: La variable numérica para analizar (sin comillas).
- **breaks**: Número de intervalos (opcional, se calcula automáticamente si no se especifica, mediante el método de Freedman-Diaconis).

4.6.4 Salida

- Una tabla de frecuencias.
- Un histograma con los mismos intervalos.

4.6.5 Ejemplo

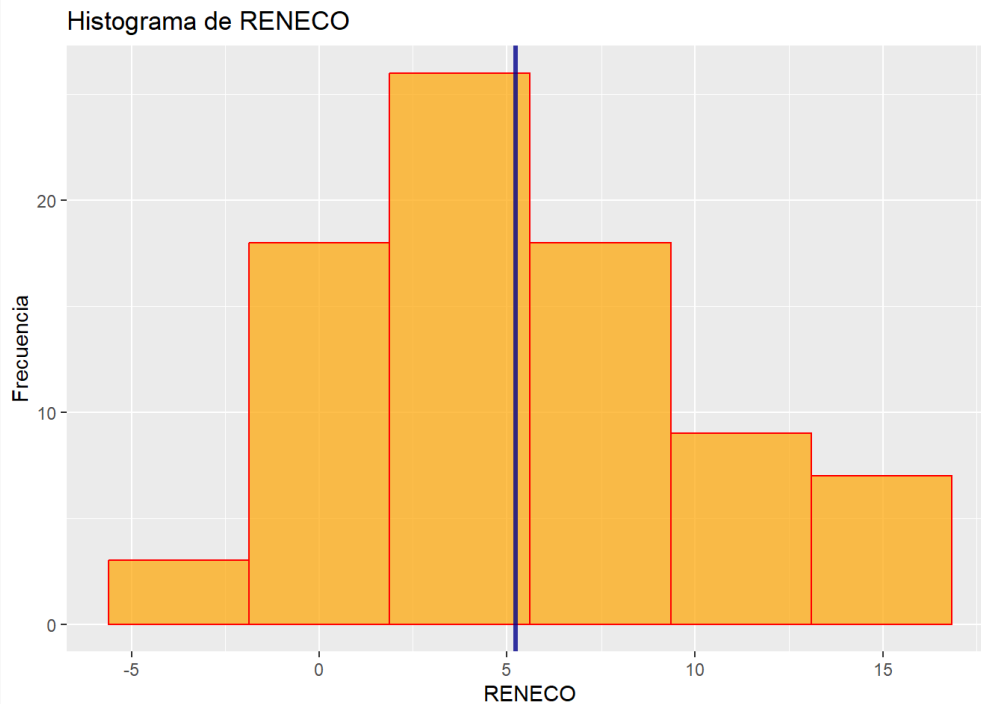
```
MATtaf(eolica_100_sm_so, RENECO)
```

4.6.6 Resultados Esperados

1. Una lista llamada `RENECO_intervalos_frecuencia` que contiene:
 - Histograma con los mismos intervalos que la tabla de frecuencias.

- Tabla de frecuencias agrupadas en intervalos.

RENECO_intervalos_frecuencia\$histograma



16

RENECO_intervalos_frecuencia\$tabla

Distribución de frecuencias agrupadas en intervalos de RENECO

Intervalo	Frecuencia absoluta n(i)	Frecuencia absoluta acum. N(i)	Frecuencia relativa f(i)	Frecuencia relativa acum. F(i)
[-2.83,0.303]	11	11	0.136	0.14
(0.303,3.42]	22	33	0.272	0.41
(3.42,6.53]	19	52	0.235	0.64
(6.53,9.65]	13	65	0.160	0.80
(9.65,12.8]	8	73	0.099	0.90
(12.8,15.9]	8	81	0.099	1.00

5 Documentación del Paquete

- **Autor:** Miguel Ángel Tarancón miguelangel.tarancon@uclm.es
- **Versión:** 0.1.0
- **Fecha:** 2025-02-14
- **Licencia:** MIT

6 ¡Empieza a explorar tus datos con MATdatatools!

Este paquete está diseñado para facilitarte el trabajo y ayudarte a generar análisis rápidos y visuales. ¡Pruébalo hoy y optimiza tus estudios de datos!

