

Manual de usuario de aedR

aedR: Análisis exploratorio de datos en R

AedR es una librería que contiene algunas herramientas gráficas y de resumen estadístico para realizar análisis exploratorio de datos.

1. Instalación

Para instalar la librería **aedR**, escriba las siguientes instrucciones en la consola de R:

```
> library(devtools)
> install_github("GeomarPerales/aedR")
```

2. Ejecución en R

Llamamos a la librería “AedR”

```
library(aedR)
```

3. Ejecución en R

Importamos los datos de trabajo.

```
data("datos.Pe")
data("estaciones.Pe")
```

El archivo “datos.Pe” corresponde a las series históricas de la red de estaciones ubicada en la cuenca Huarpa. visualizamos los datos importados: datos.Pe

```
head(datos.Pe,4)
#>      Fecha San_Miguel Cochabamba Quinua Wayllapampa Huamanga Allpachaca Huanta
#> 1 1964-01-01      29.8      NA      NA      NA      NA      NA      38.5
#> 2 1964-02-01      74.2      NA      NA      NA      NA      NA      82.5
#> 3 1964-03-01     107.0      NA      NA      NA      NA      NA      60.2
#> 4 1964-04-01      44.9      NA      NA      NA      NA      NA      30.9
#>      Huanta_gore San_Pedro Acobamba Lircay Tocaz Paucarbamba Lauricocha
#> 1      NA      NA      NA      NA      NA      NA      NA
#> 2      NA      NA      NA      NA      NA      NA      NA
#> 3      NA      NA      NA      NA      NA      NA      NA
#> 4      NA      NA      NA      NA      NA      NA      NA
#>      Huancavelica Paras Chuschi Libertadores_706 Libertadores Choclococha
#> 1      66.1      NA      NA      NA      NA      NA      NA
#> 2     108.0      NA      NA      NA      NA      NA      NA
#> 3     186.0      NA      NA      NA      NA      NA      NA
#> 4      75.7      NA     86.4      NA      NA      NA
#>      Acnococha Tunel_cero
#> 1      NA      NA
#> 2      NA      NA
#> 3      NA      NA
#> 4      NA      NA
```

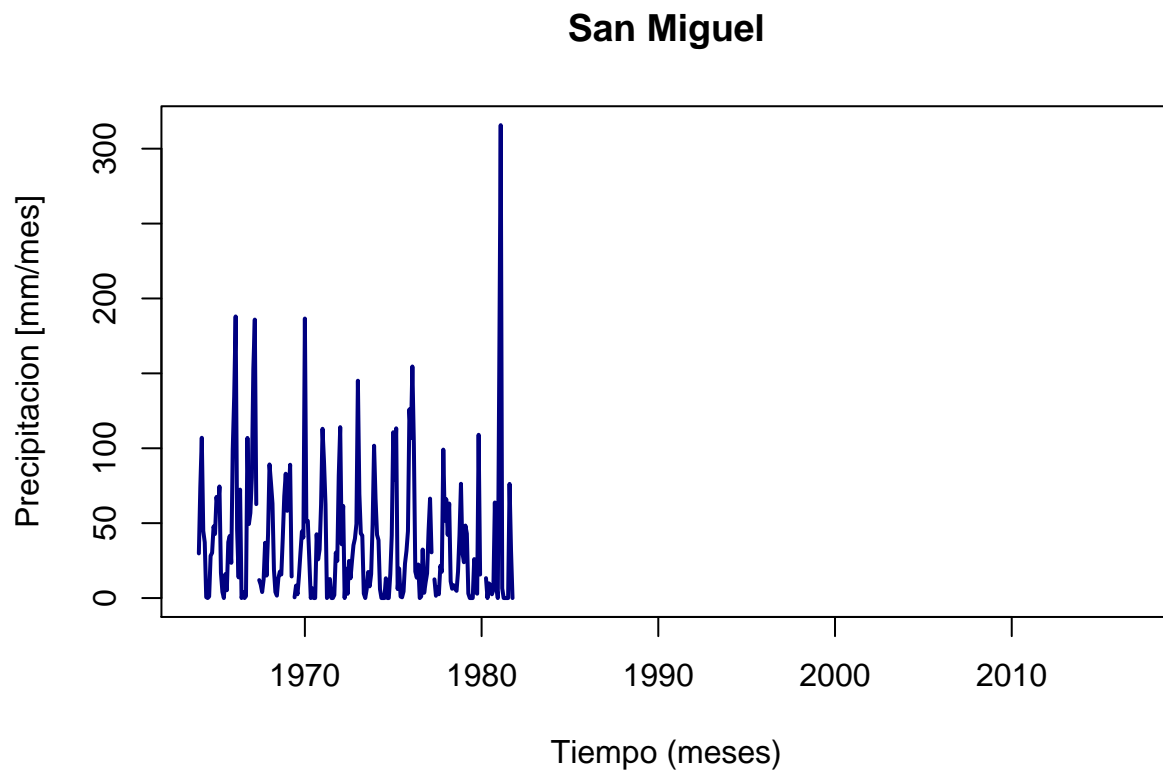
El archivo “estaciones.Pe” corresponde a la red de estaciones ubicada en la cuenca Huarpa. visualizamos los datos importados: estaciones.Pe

```
head(estaciones.Pe,4)
#>      Estacion      Lat      Lon  Alt
#> 1   San Miguel -13.01667 -73.98333 3323
#> 2 Hacienda Cochas -13.03333 -73.88333 3323
#> 3   La Quinoa -13.05528 -74.14139 3240
#> 4  Wayllapampa -13.07639 -74.21667 2470
```

4. Uso de las herramientas para AED

Genero un gráfico de la serie de tiempo de una estación para visualizar el comportamiento de la variable durante el período de estudio (1964-01 a 2016-12).

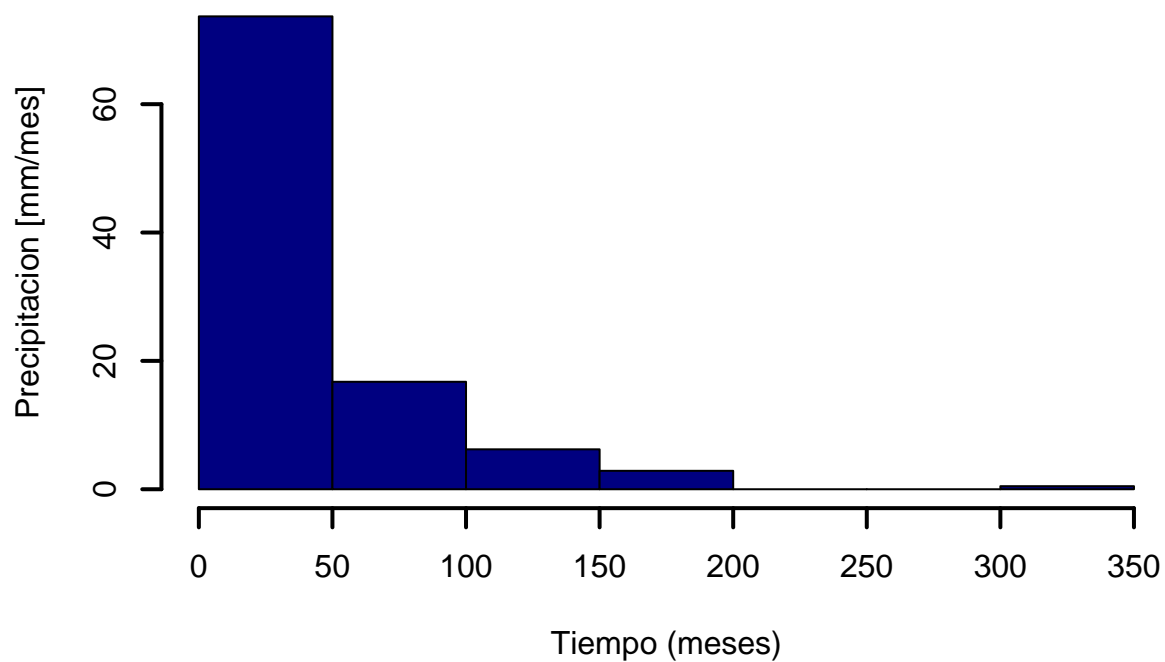
```
ts.aed(datos.Pe, estaciones.Pe)
```



Para observar la frecuencia (%) de valores de precipitación, genero un histograma donde se muestran los intervalos y sus respectivas frecuencias del conjunto de datos de las series históricas.

```
his.aed(datos.Pe, estaciones.Pe)
```

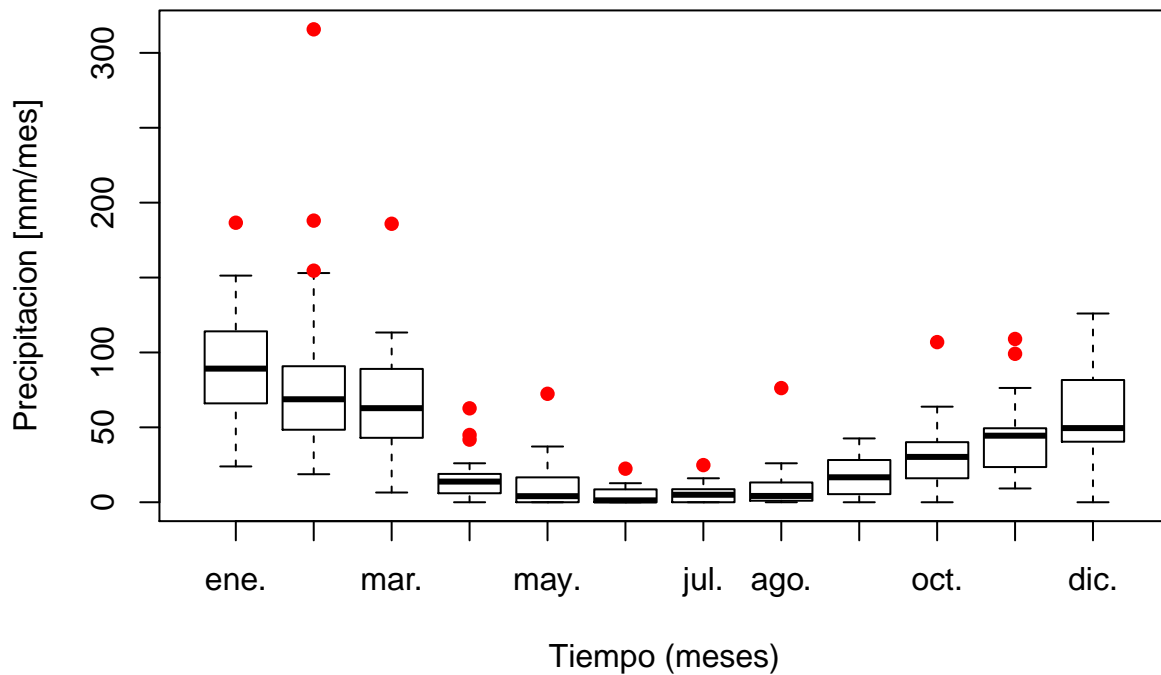
San Miguel



Los datos de la información histórica no cuentan con análisis de la calidad de la información, Genero un gráfico de diagrama de cajas (boxplot) de la estación para visualizar los datos dudosos (outliers) presentes.

```
bp.aed(datos.Pe, estaciones.Pe)
```

San Miguel



Para ver la cantidad de datos dudosos de las series históricas de precipitación, el mínimo y máximo valor de los datos dudosos, genero un resumen de los datos dudosos (outliers).

```
datos.ouliers <- infout(datos.Pe, estaciones.Pe)
#> Warning in min(bp$out): ningún argumento finito para min; retornando Inf
#> Warning in max(bp$out): ningún argumento finito para max; retornando -Inf
head(datos.ouliers)
#>      Estacion cantidad.out min.out max.out
#> 1   San Miguel          15    22.4   315.7
#> 2 Hacienda Cochas           9    59.2  1908.2
#> 3   La Quinua          16    35.1   445.6
#> 4 Wayllapampa          22    14.6   406.8
#> 5   Huamanga           12    53.5   319.2
#> 6 Allpachaca           11    20.3   438.7
```

Para analizar la información no disponible, genero un resumen general de todas las series históricas de las estaciones, observo la cantidad de datos faltantes y el porcentaje que representan respecto al total de datos.

```
info.NA <- infna(datos.Pe, estaciones.Pe)
head(info.NA)
#>      Estacion Cantidad Porcentaje
#> 1   San Miguel      427      67.14%
#> 2 Hacienda Cochas    502      78.93%
#> 3   La Quinua      162      25.47%
#> 4 Wayllapampa      228      35.85%
#> 5   Huamanga      324      50.94%
#> 6 Allpachaca      363      57.08%
```

Visualizo el máximo valor, el mínimo valor, el primer cuartil, el tercer cuartil y la fecha de ocurrencia del máximo valor.

```
rest(datos.Pe, estaciones.Pe)
#>      Estacion Max      P1Q      Fmax      P3Q      Min
#> 1      San Miguel    0  4.400 1981-02-01 37.55182 51.700
#> 2  Hacienda Cochas    0  0.000 1972-03-01 132.72313 125.000
#> 3      La Quinua    0 15.950 1967-03-01 65.89641 104.325
#> 4    Wayllapampa    0  6.175 1971-02-01 52.58625 85.725
#> 5      Huamanga    0 11.100 1986-03-01 48.89147 77.775
#> 6    Allpachaca    0 14.500 1967-03-01 68.90747 105.690
#> 7      Huanta    0  7.900 1967-03-01 44.83123 71.400
#> 8    Huanta gore    0 15.975 2001-01-01 64.62440 98.500
#> 9      San Pedro    0 11.650 1967-03-01 112.28748 102.950
#> 10    Acobamba    0 14.550 1972-01-01 59.58521 91.075
#> 11    Lircay    0 20.250 1967-02-01 66.36738 102.925
#> 12  Hacienda Tocaz    0  6.500 1974-03-01 88.98722 131.600
#> 13    Paucarbamba    0 26.000 1964-11-01 90.69870 129.100
#> 14    Lauricocha    0  6.200 1984-01-01 32.32215 47.600
#> 15    Huancavelica    0 18.750 2003-02-01 75.95226 116.650
#> 16      Paras    0  6.600 1972-02-01 70.09874 103.900
#> 17    Chuschi    0  8.300 1981-12-01 71.43622 112.100
#> 18  Libertadores 706    0 13.000 1986-01-01 74.70187 117.900
#> 19 Libertadores 156139    0 11.375 1981-01-01 75.92984 122.225
#> 20    Choclococha    0 15.475 1998-02-01 77.99791 112.175
#> 21    Acnococha    0 12.300 1984-02-01 58.50983 95.025
#> 22    Tunel cero    0 13.950 1984-02-01 72.90385 120.450
```

5. Referencias

- Pathak, M. (2014). Beginning Data Science with R.
- Zumel, N., Mount, J. (2014). Practical Data Science with R.