

Pokemon

Curso de Estadística Descriptiva

6/3/2021

Pokemon (Py -> R)

Limpieza de datos en Python

```
import pandas as pd
pokemon = pd.read_csv("../data/Pokemon.csv")
print(pokemon.head())
```

```
##           Name Type 1  Type 2  ...  Speed  Generation  Legendary
## 0      Bulbasaur  Grass  Poison  ...    45           1         False
## 1        Ivysaur  Grass  Poison  ...    60           1         False
## 2       Venusaur  Grass  Poison  ...    80           1         False
## 3  VenusaurMega  Venusaur  Poison  ...    80           1         False
## 4     Charmander   Fire    NaN    ...    65           1         False
##
## [5 rows x 12 columns]
```

```
print(pokemon.shape)
```

```
# Para solo ver los pokemon de la primera generación
```

```
## (800, 12)
```

```
pokemon = pokemon[pokemon["Generation"]==1]
```

```
# Lo filtramos aún más quedandonos con solo los del tipo 1 y 2, y la velocidad
pokemon = pokemon[["Type 1", "Type 2", "Speed"]]
print(pokemon.shape)
```

```
# Por si acaso hay variables na para que las elimine
```

```
## (166, 3)
```

```
pokemon = pokemon.dropna()
print(pokemon.shape)
```

```
## (78, 3)
```

Transmisión de los datos de Python a R

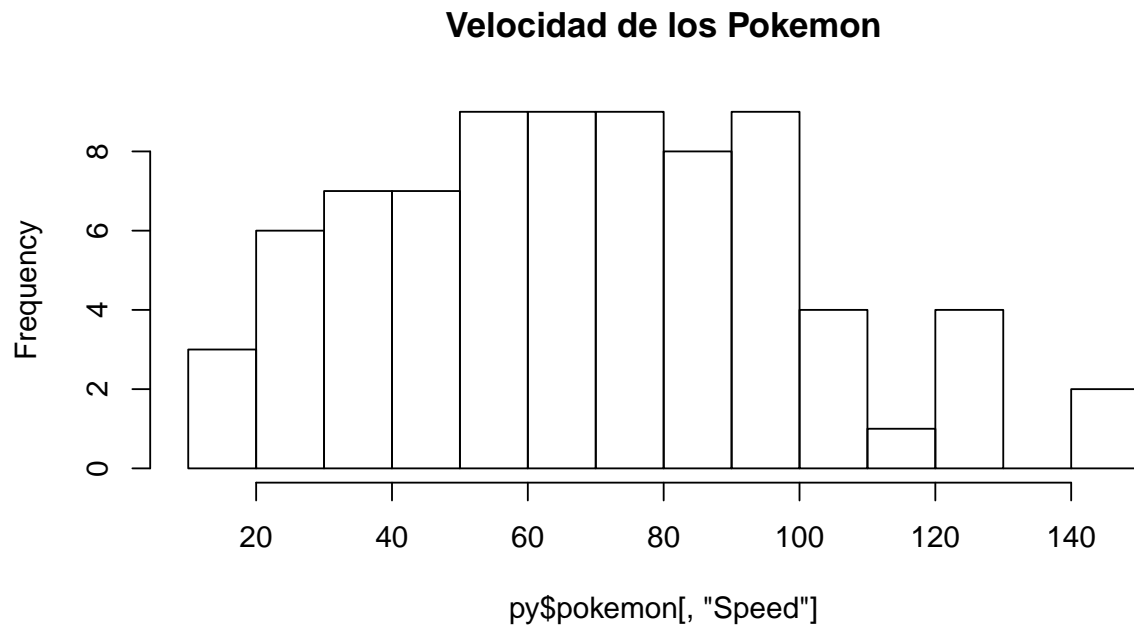
Con **fig.width** le doy el ancho a la figura

Con **fig.height** le doy la altura a la figura

Un histograma de la variable **pokemon** creada en **python**, todas las filas para la columna llamada “**speed**”.

Numero de divisiones 10

```
hist(py$pokemon[, "Speed"], breaks = 10, main = "Velocidad de los Pokemon")
```



Pokemon (R -> Py)

Carga de datos en R

```
pokemon2 <- read.csv("../data/Pokemon.csv", header = TRUE)
head(pokemon2)
```

```
##           Name Type.1 Type.2 Total HP Attack Defense Sp_Atk Sp_Def
## 1      Bulbasaur  Grass Poison  318 45    49    49    65    65
## 2        Ivysaur  Grass Poison  405 60    62    63    80    80
## 3        Venusaur  Grass Poison  525 80    82    83   100   100
## 4 VenusaurMega Venusaur  Grass Poison  625 80   100   123   122   120
## 5      Charmander   Fire         309 39    52    43    60    50
## 6     Charmeleon   Fire         405 58    64    58    80    65
## Speed Generation Legendary
## 1      45           1      False
```

```
## 2    60      1    False
## 3    80      1    False
## 4    80      1    False
## 5    65      1    False
## 6    80      1    False
```

```
library(tidyverse) # Esto es avanzado de como trabajar
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.0    v purrr   0.3.4
## v tibble  3.0.1    v dplyr   0.8.5
## v tidyr   1.0.3    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
# Lo que tenia anteriormente (pokemon2) lo filtro
# %>% significa que lo filtro
pokemon2 <- pokemon2 %>%
  filter(Generation == 1) %>% # Primero filtro y cojo solo los de la 1ª generacion
  select(Type.1, Type.2, Speed) %>% # A continuación selecciono solo las columnas que me interesan
  na.omit() # Finalmente, elimino los na
summary(pokemon2)
```

```
##      Type.1      Type.2      Speed
## Water  :31      :88    Min.   : 15.00
## Normal :24    Flying :23    1st Qu.: 50.00
## Bug    :14    Poison :22    Median : 70.00
## Fire   :14    Psychic: 7     Mean   : 72.58
## Poison :14    Ground : 6     3rd Qu.: 92.25
## Grass  :13    Water  : 4     Max.   :150.00
## (Other):56    (Other):16
```

Transmisión de datos de R a Python

```
print(r.pokemon2.head())
```

```
##   Type.1 Type.2 Speed
## 0  Grass Poison   45
## 1  Grass Poison   60
## 2  Grass Poison   80
## 3  Grass Poison   80
## 4   Fire         65
```