

1^a
Emisión

DATA SCIENCE

Módulo 05 Manipulación y visualización de datos con Python

Mtro. Ricardo Daniel Alanis Tamez



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Dirección General de Cómputo y de Tecnologías de información y Comunicación
Dirección de Docencia en TIC



Educación
Continua
1971 - 2021

Gráficos estadísticos básicos

Manipulación y visualización de datos con Python

Ricardo Alanís

Presentación

En esta lección vamos a dar un paseo por las visualizaciones de datos exploratorias partiendo de la cantidad de variables a observar por cada gráfico. Luego vamos a hacer una pequeña competencia para visualizar ciertos datos de la mejor manera.

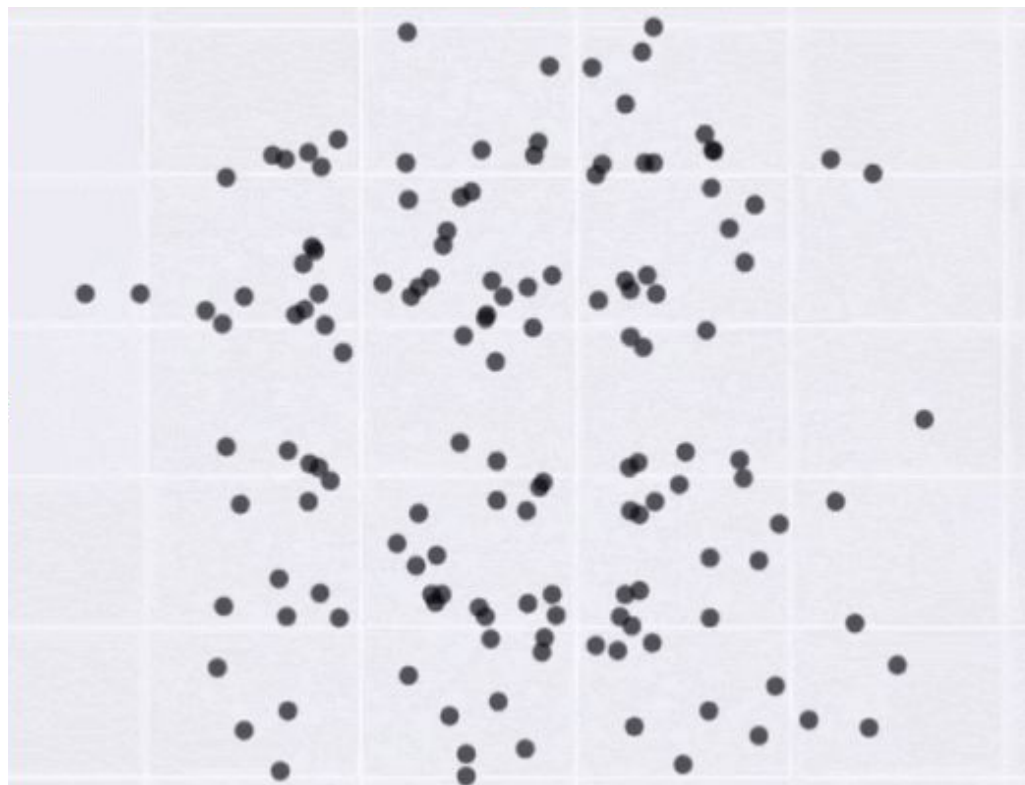
Objetivo

El participante identificará las distintas herramientas disponibles para el análisis estadístico gráfico de la información.

Agenda de hoy

- 1. Gráficos de visualización unidimensional en datos**
- 2. Gráficos de relación entre dos variables**
- 3. Gráficos para el análisis Multivariable**
- 4. Práctica de Estadísticos básicos visuales**

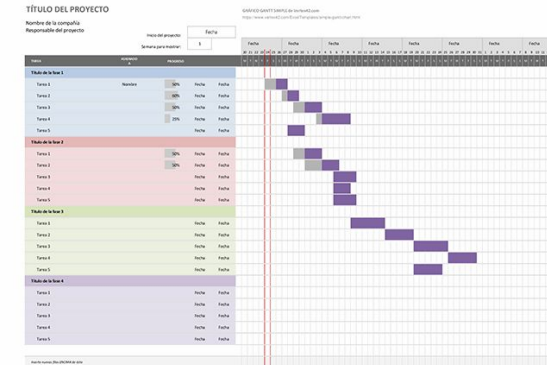
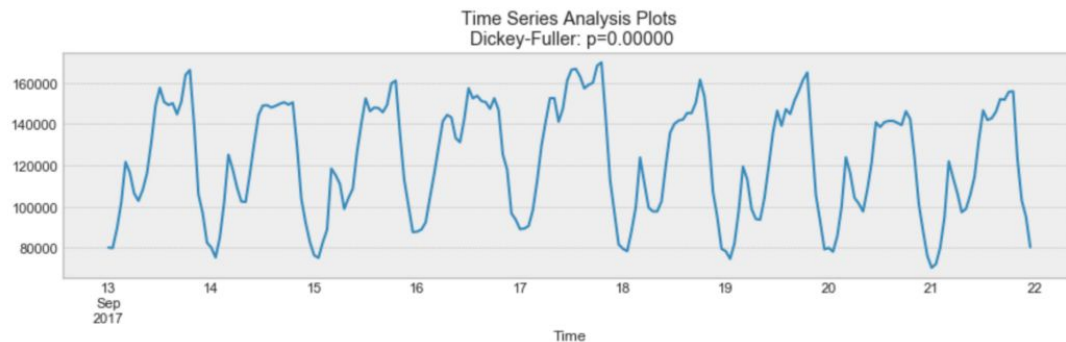
La Docena del Datasauro



X Mean: 54.26
Y Mean: 47.83
X SD: 16.76
Y SD: 26.93
Corr. : -0.06

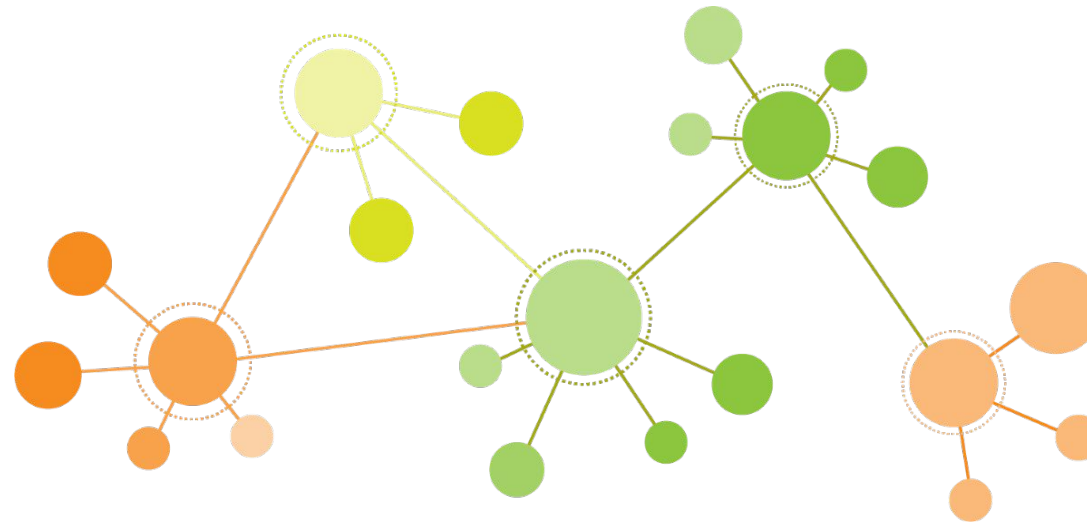
Gráficos por su objetivo

- **Visualización temporal**
- Normalmente se realiza para datos unidimensionales y muestra algún tipo de relación lineal entre puntos de datos. Dichos conjuntos de datos generalmente involucran al tiempo como una variable independiente y, por lo tanto, los datos de series de tiempo se visualizan de esta manera.
- Tipos de gráficos: gráficos de dispersión, gráficos de Gantt, líneas de tiempo, gráficos de líneas de series temporales.



Gráficos por su objetivo

- **Visualización de red**
- Como sugiere el nombre, la visualización de redes se trata de conectar varios conjuntos de datos entre sí y mostrar cómo se relacionan entre sí en una red en la que cada variable está conectada.
- Tipos de gráficos: diagramas de enlaces de nodos, gráficas de matriz, gráficas de dependencia.

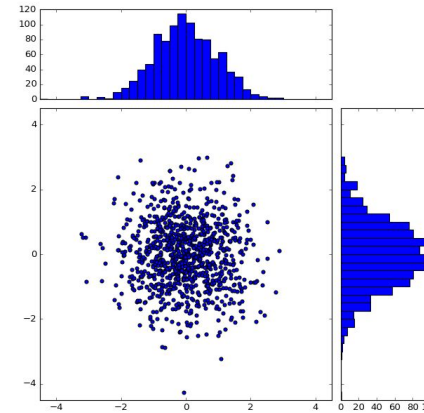


Gráficos por su objetivo

- **Visualización jerárquica**
- Se usa cuando el conjunto de datos contiene variables ordenadas conectadas entre sí. Se puede utilizar para mostrar la relación entre las variables principal y secundaria, especialmente cuando los datos se pueden agrupar en diferentes categorías.
- Tipos de parcelas: diagramas de árbol, dendrogramas, diagramas de rayos solares, diagramas de anillo.

Gráficos por su objetivo

- **Visualización multidimensional**
- Estos tipos de gráficos se utilizan cuando existen múltiples dimensiones y es posible crear un diagrama 3D - pero no se recomienda. Aunque pueden ser intrínsecamente complejos, los gráficos multidimensionales pueden albergar una gran cantidad de datos (y conocimientos).
- Tipos de gráficos: histogramas 2D / 3D, gráficos de dispersión 2D / 3D, gráficos circulares, de barras y de líneas.

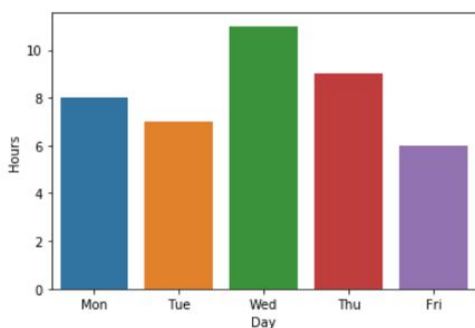


Gráficos por sus elementos gráficos

Barras

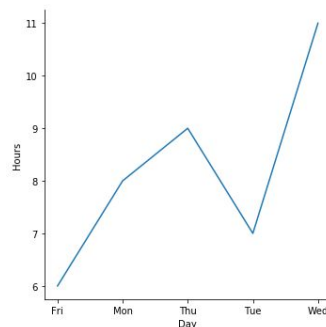
Comparar
variables
similares

<matplotlib.axes._subplots.AxesSubplot at 0x19aa9437e88>



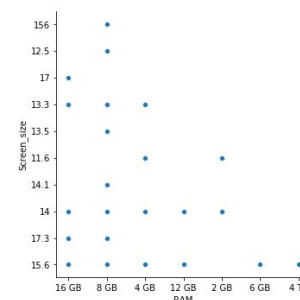
Línea

Visualizar
tendencias



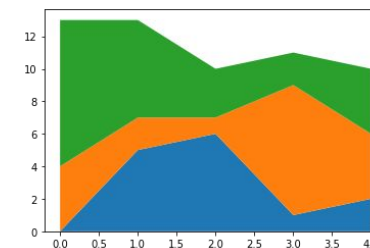
Dispersión

Relación y
patrones de
variables

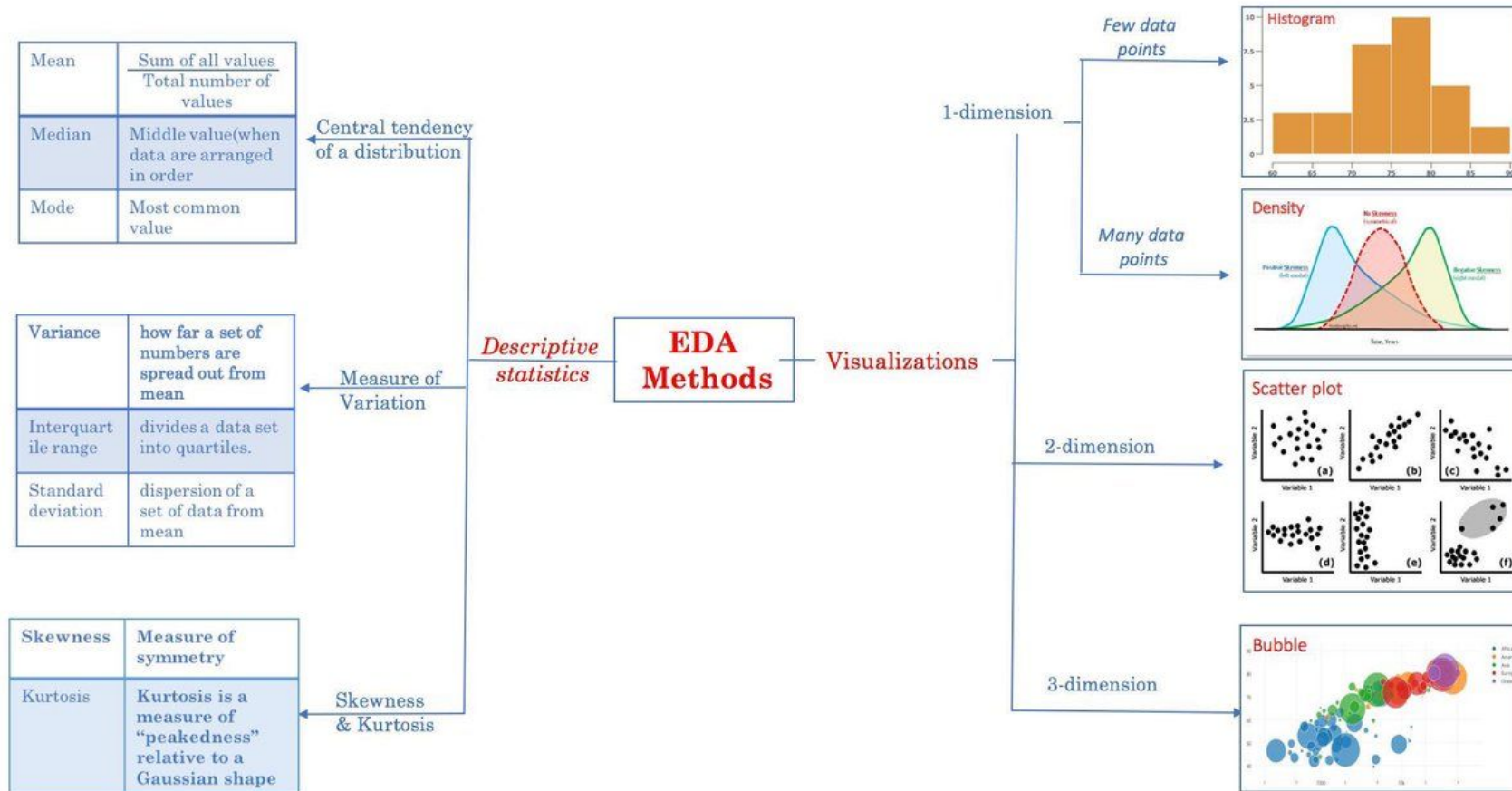


Area

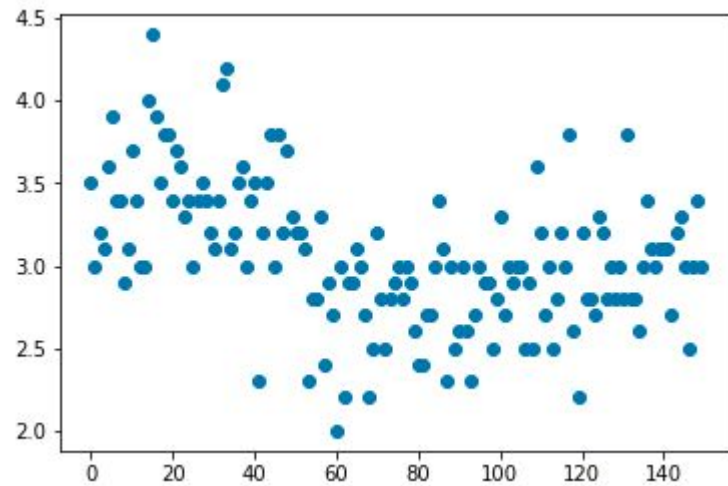
Evolución
acumulada
de variable



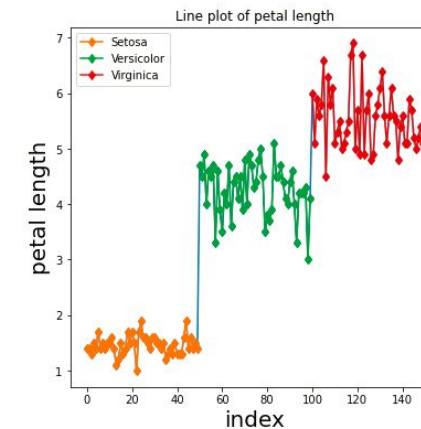
Visualización como herramienta de EDA



Gráficos unidimensionales

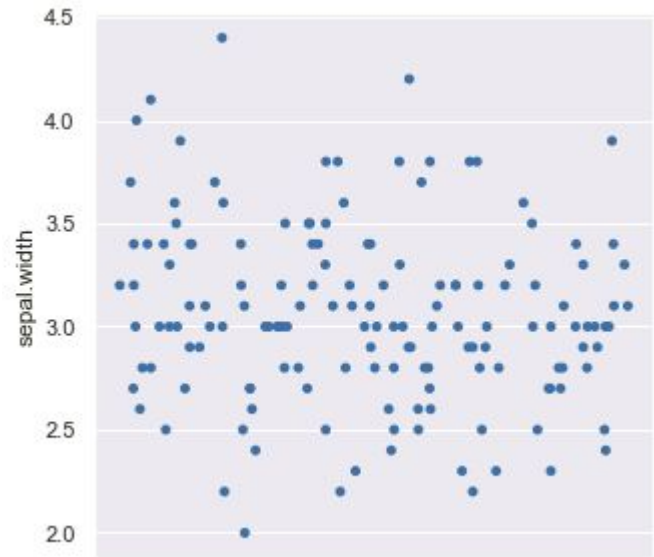


**Dispersión
univariada**

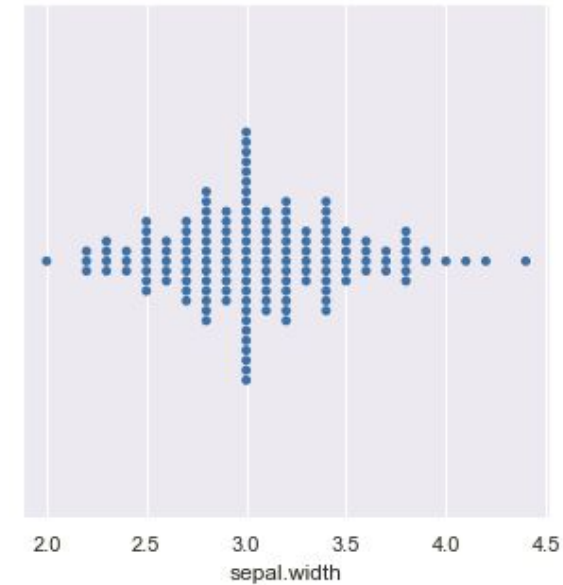


**Línea
Univariada**

Gráficos unidimensionales



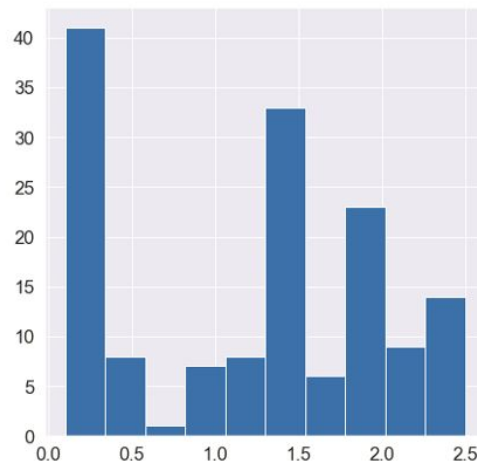
Strip plot



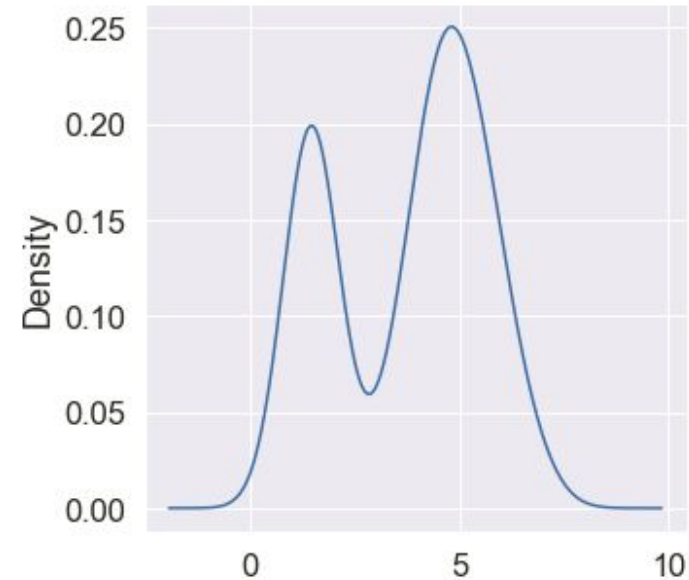
Swarm Plot

Gráficos unidimensionales: Resumen

```
In [12]: plt.hist(df['petal.width'])  
Out[12]: (array([41., 8., 1., 7., 8., 33., 6., 23., 9., 14.]),  
array([0.1, 0.34, 0.58, 0.82, 1.06, 1.3, 1.54, 1.78, 2.02, 2.26, 2.5 ]),  
<a list of 10 Patch objects>)
```



Histograma

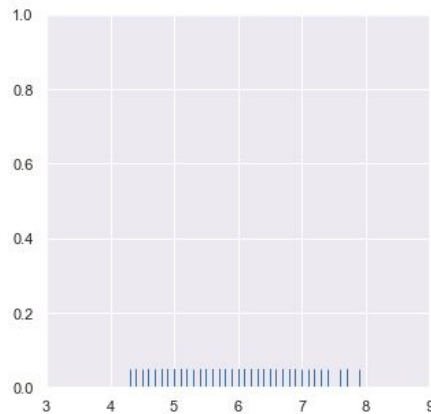


**Gráfica de
Densidad**

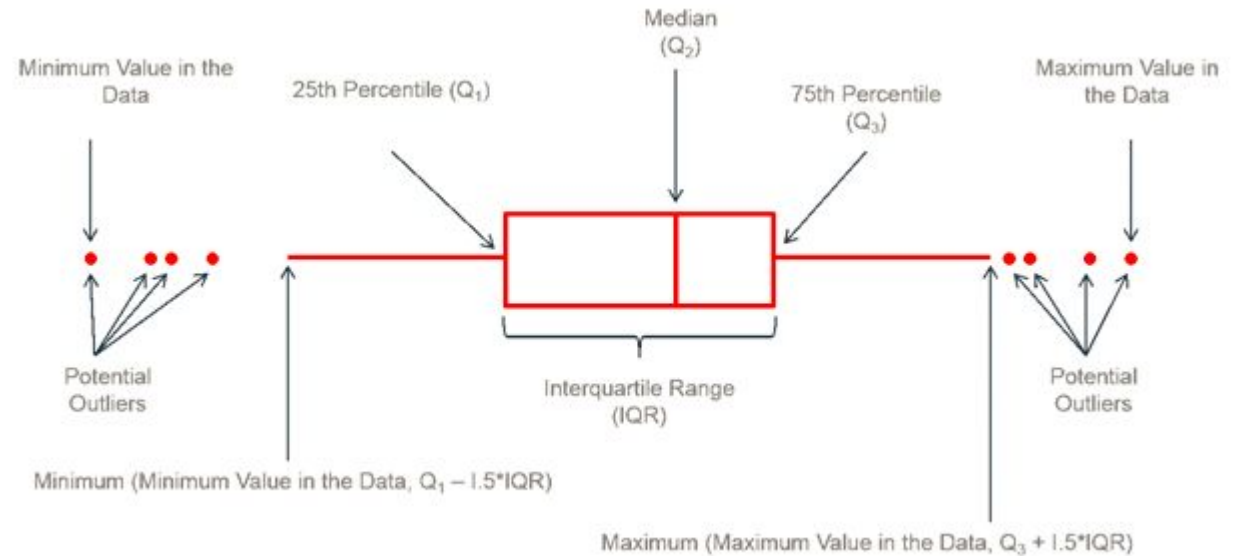
Gráficos unidimensionales: Resumen

RUG PLOT

```
fig, ax = plt.subplots()
sns.rugplot(df['sepal.length'])
ax.set_xlim(3,9)
plt.show()
```

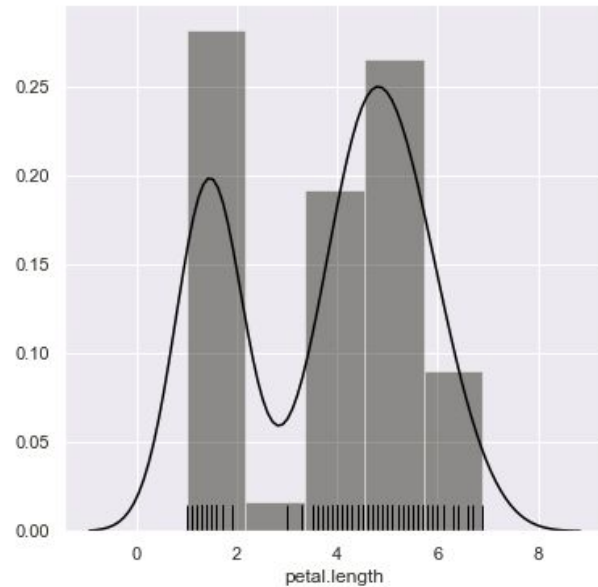


Rug plot

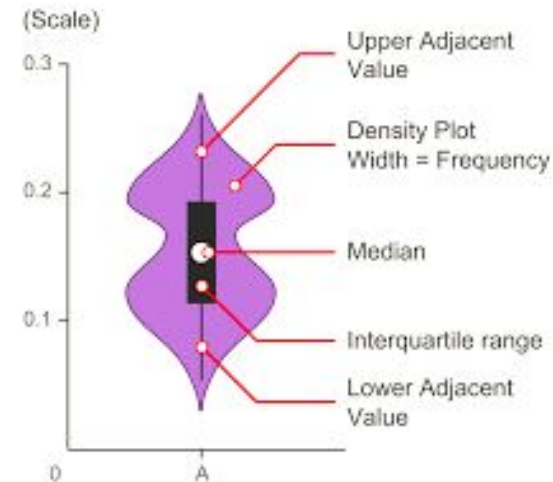


Box plot

Gráficos unidimensionales: Resumen

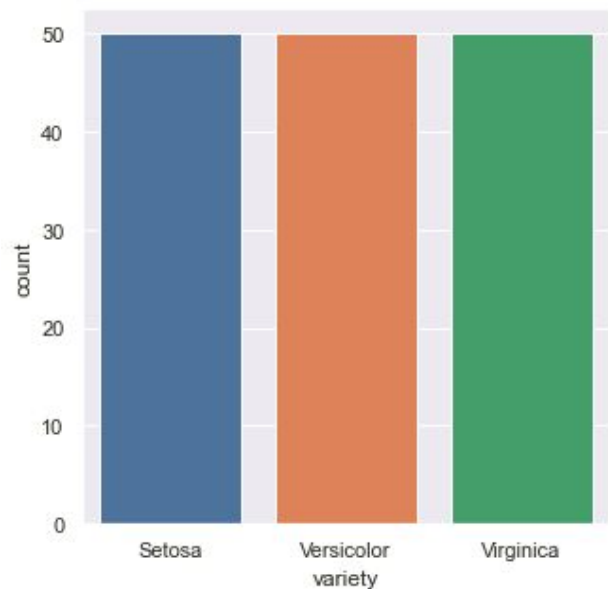


Distplot

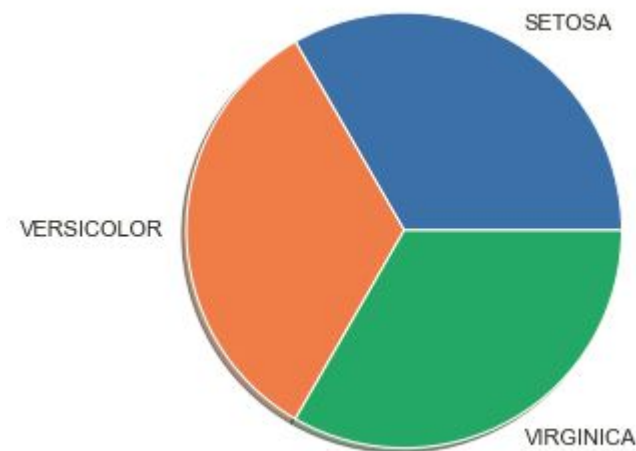


Violin plot

Gráficos unidimensionales: Categóricos

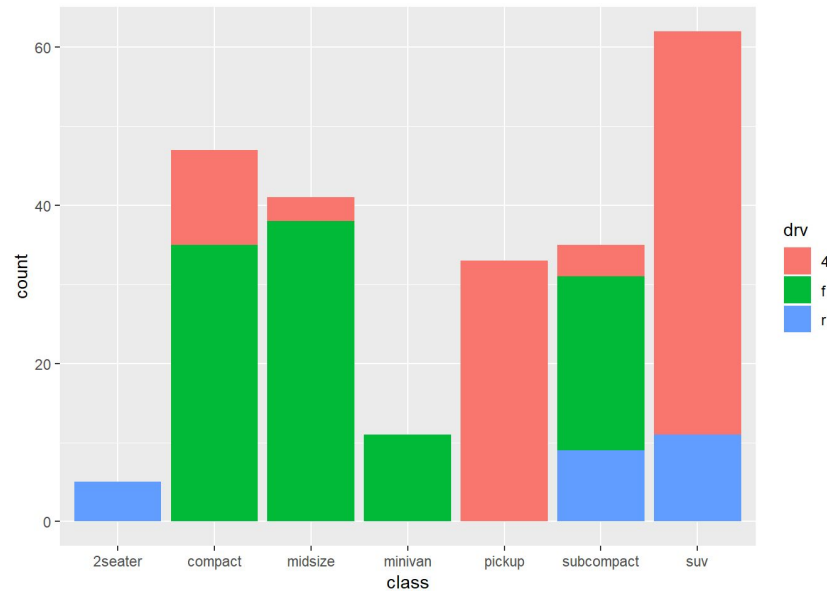


**Gráficas de
barras**

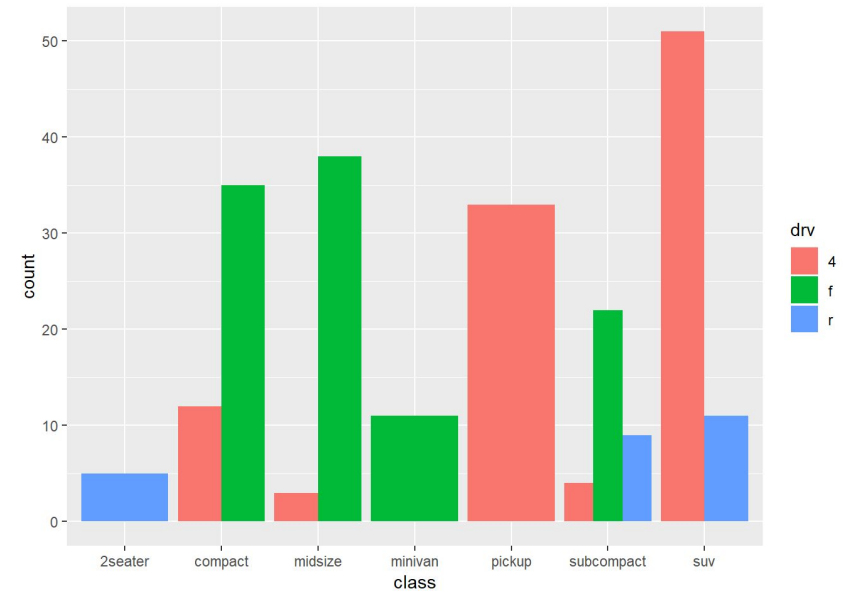


**Gráficos de
Pie**

Gráficos bivariados (Dos categóricas)

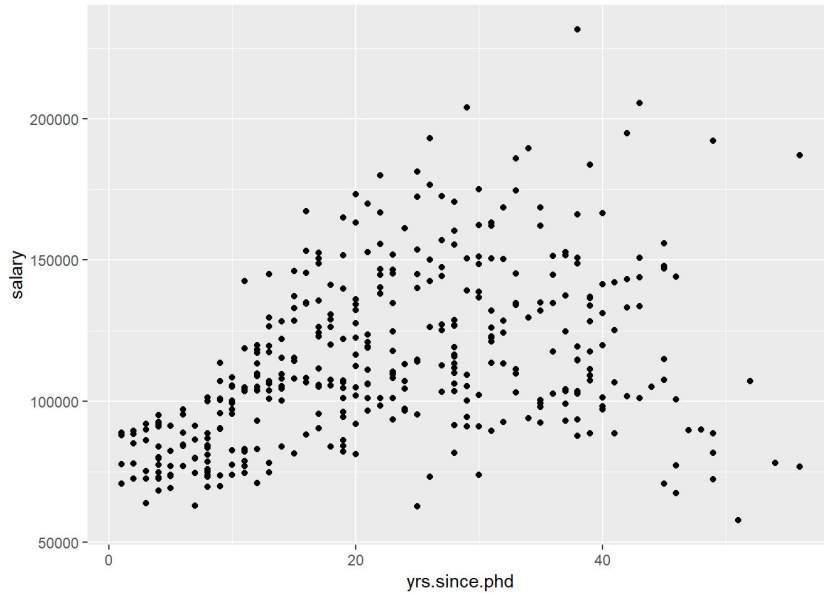


Stacked bar graph

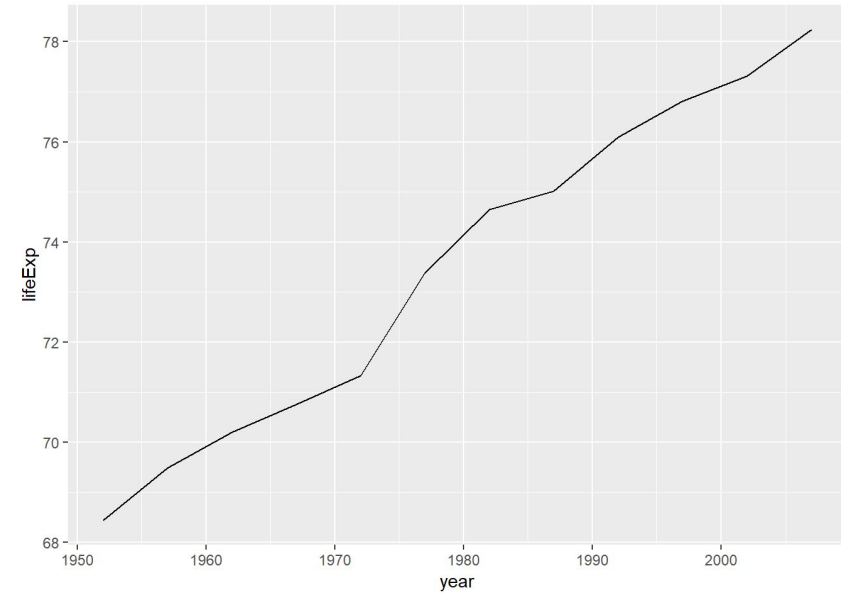


Grouped bar chart

Gráficos bivariados (2 Numéricos)

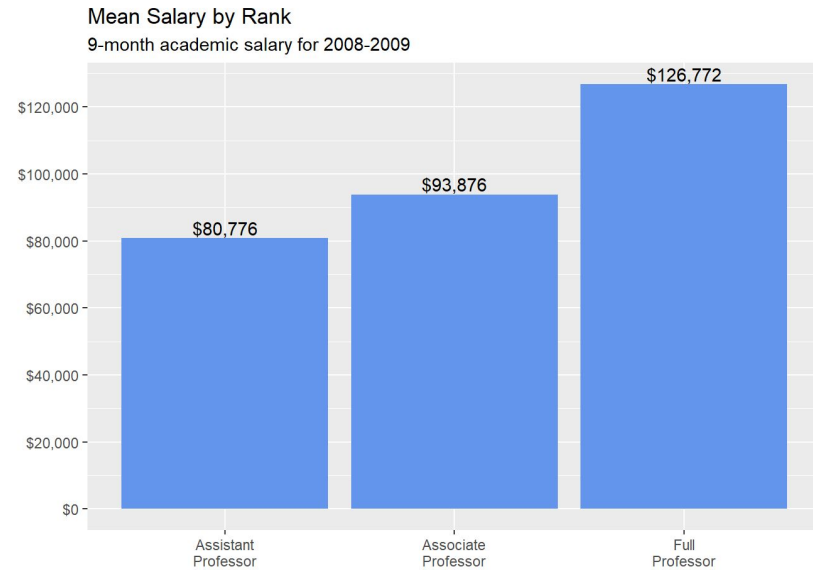


Dispersión

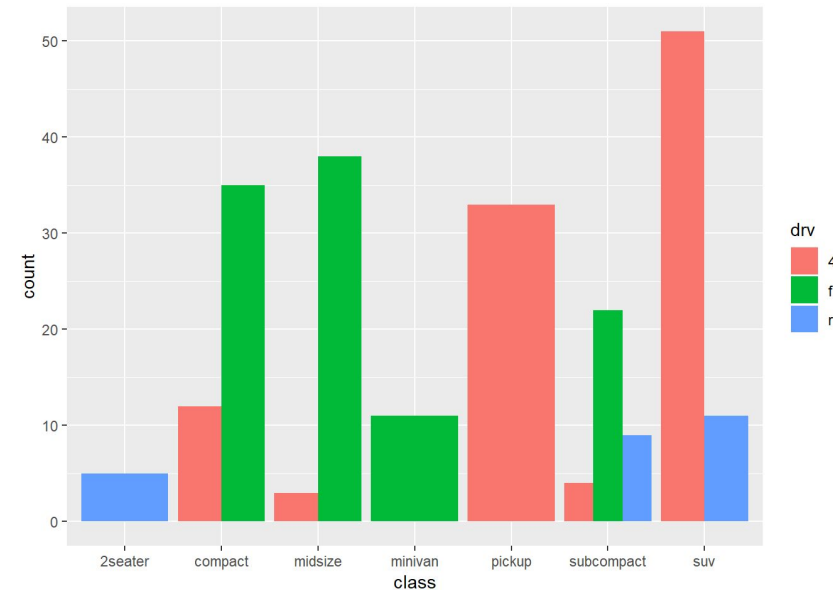


**Gráfico de
Línea**

Gráficos bivariados (Categórica vs. Numérica)

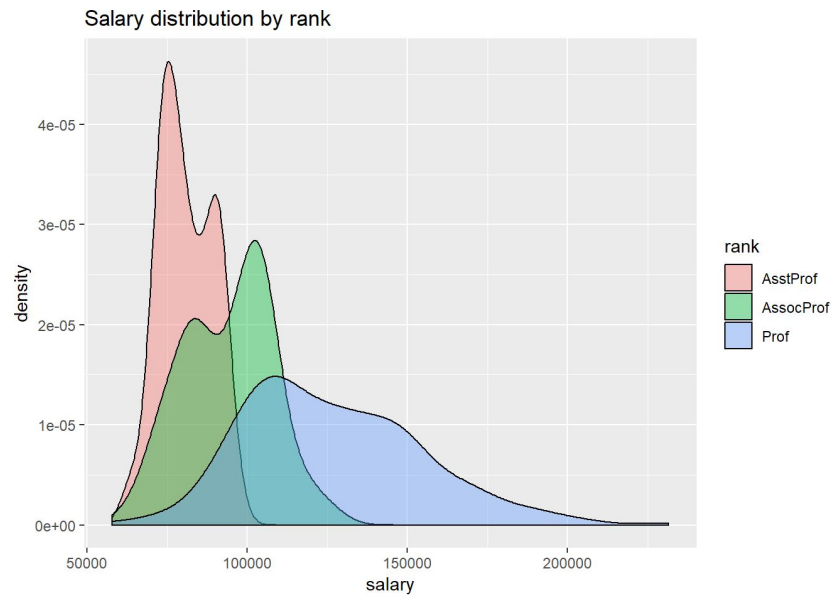


Stacked bar graph

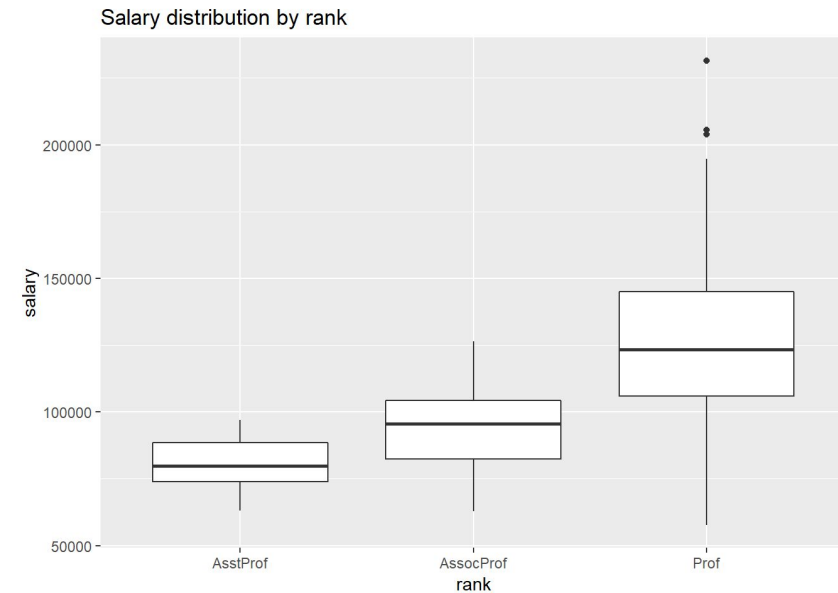


Grouped bar chart

Gráficos bivariados (Categorica vs. Numérica)

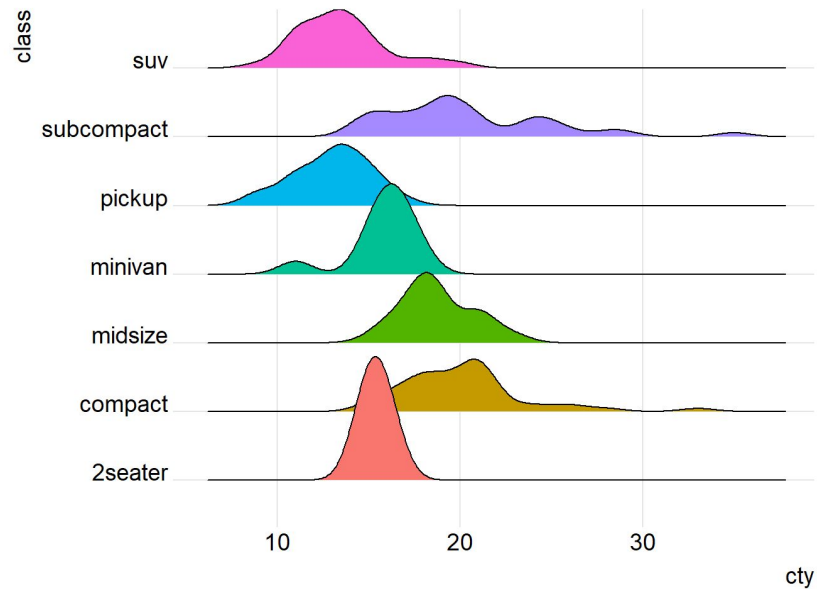


Density Plot

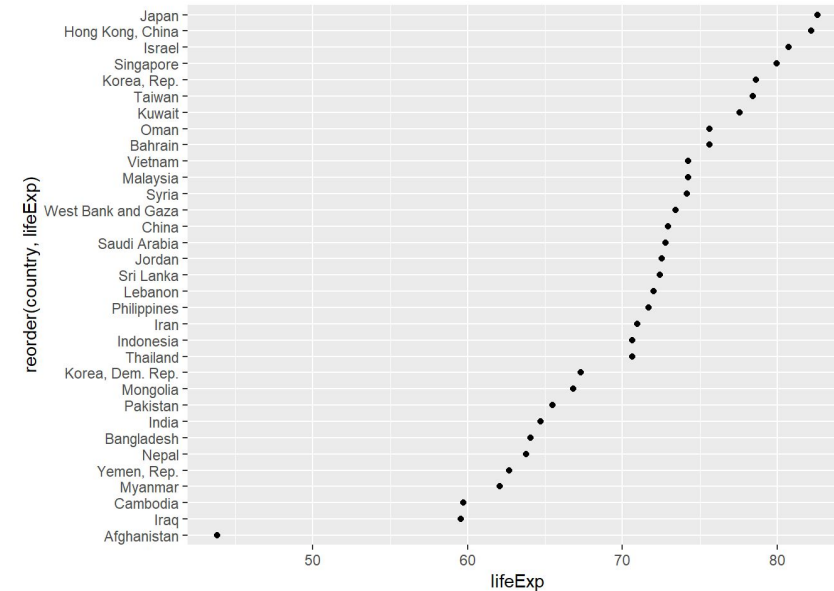


Multibox plot

Gráficos bivariados (Categorica vs. Numérica)

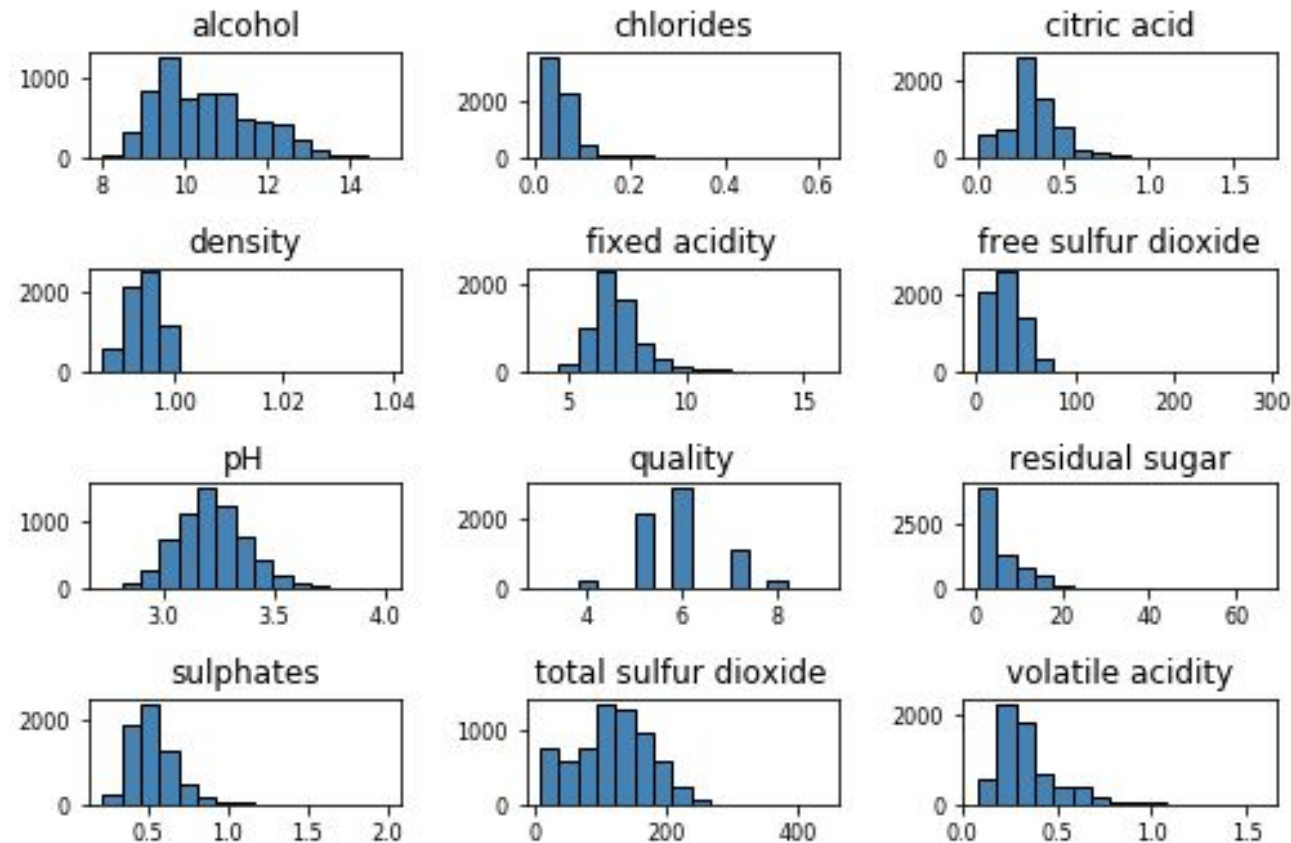


Redline plots



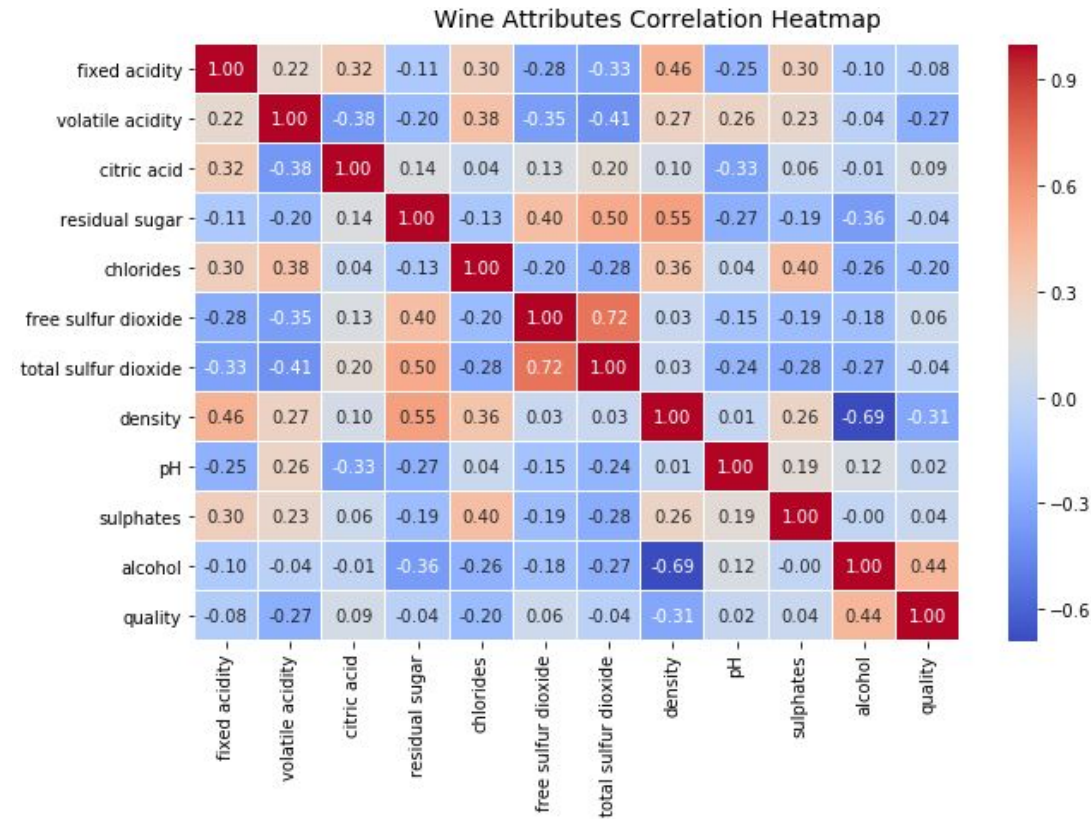
Cleveland dot charts

Multivariados



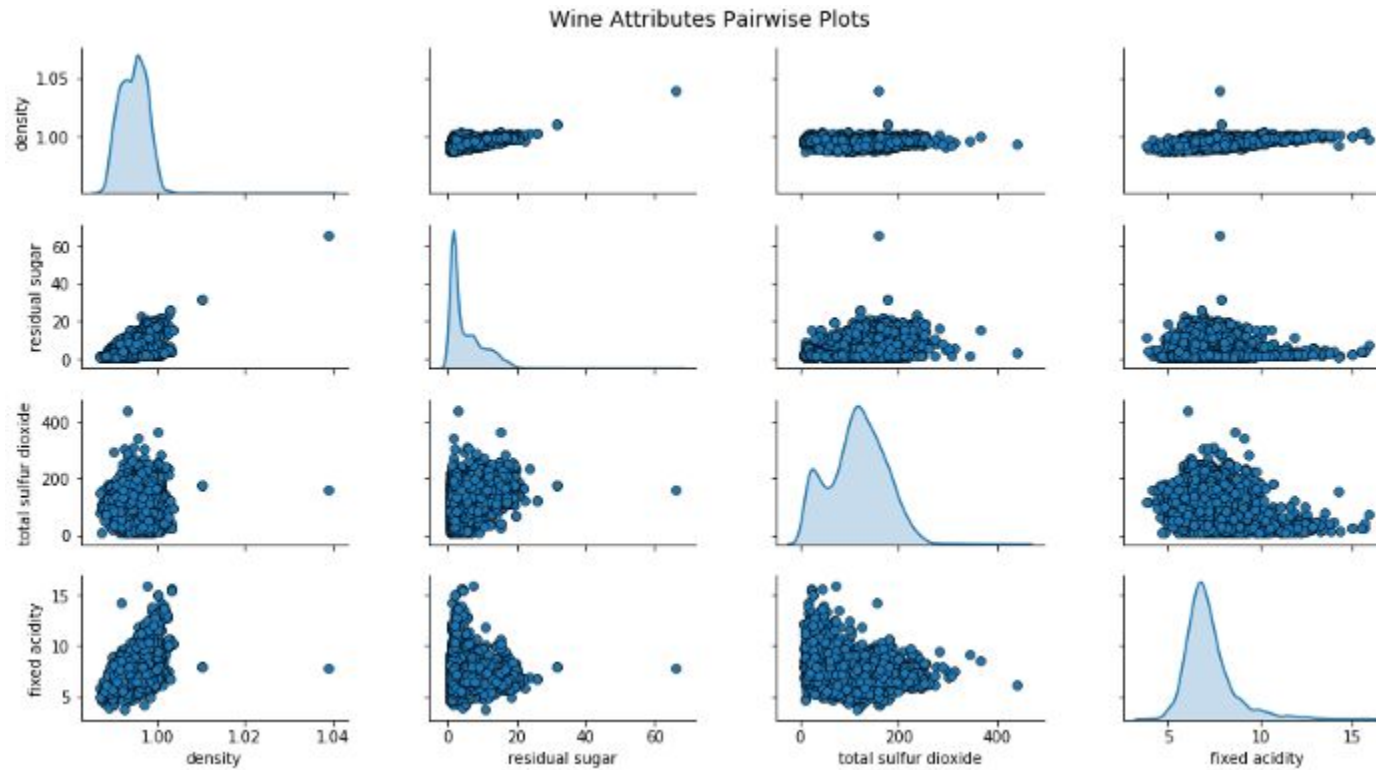
Visualización de 1d de las variables

Multivariados



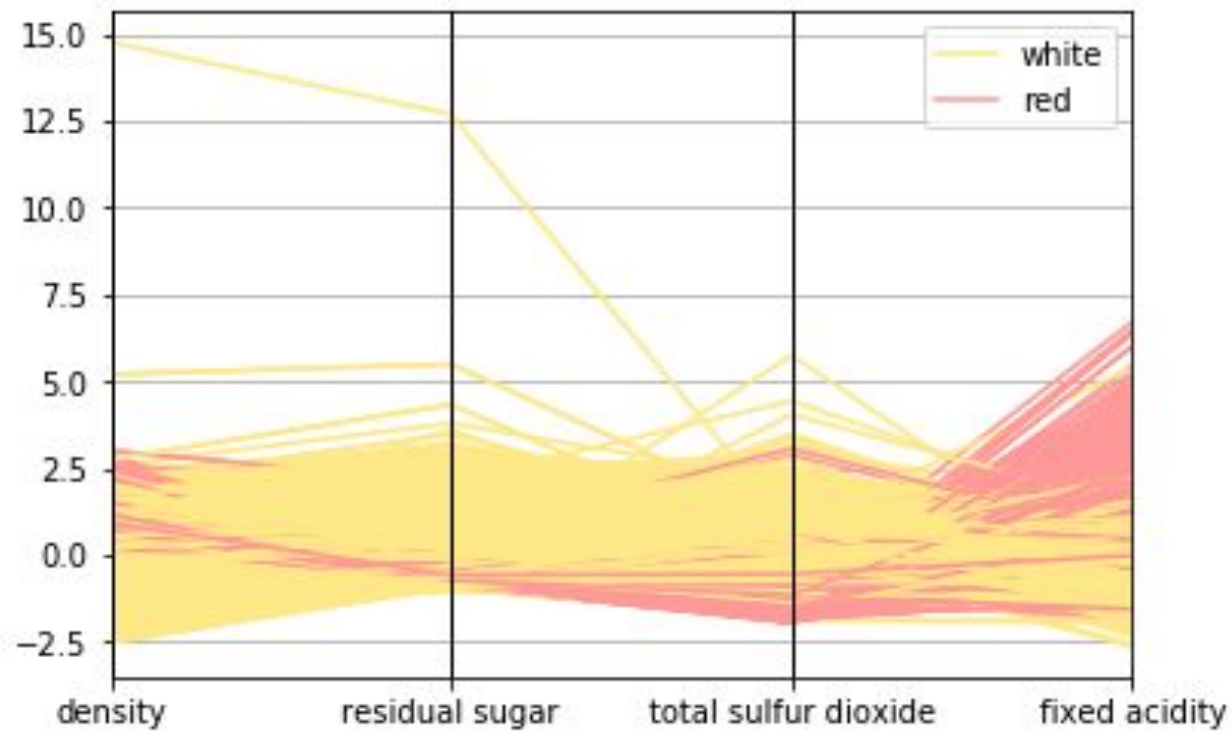
Heatmap

Multivariados



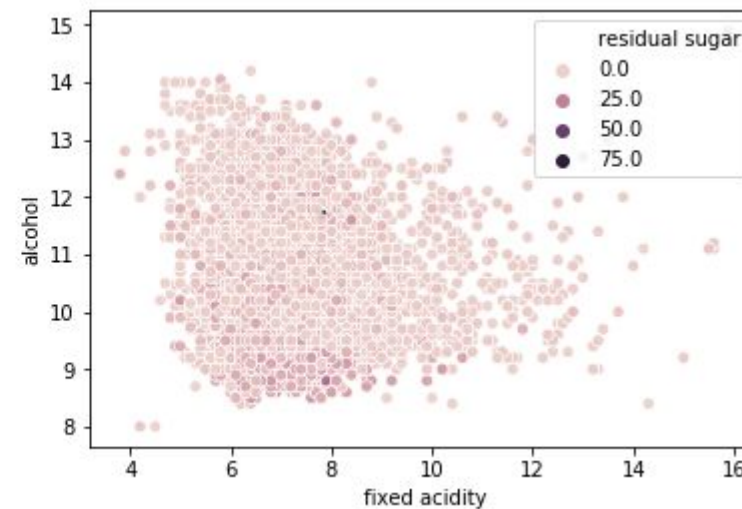
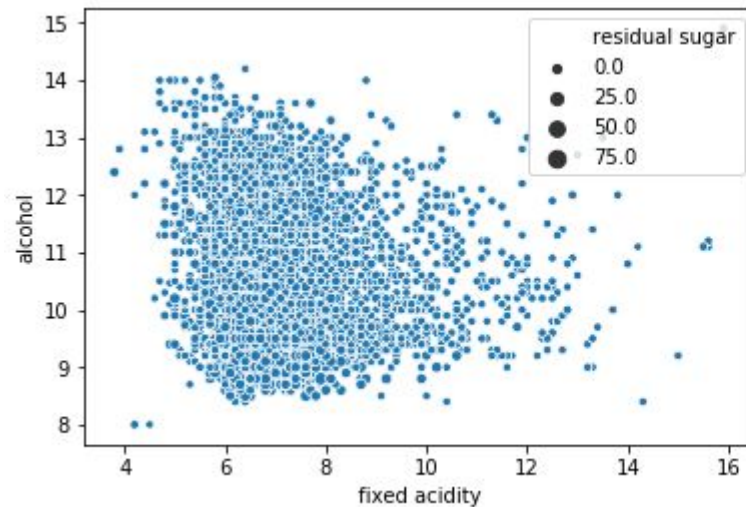
Pairwise scatter plot

Multivariados



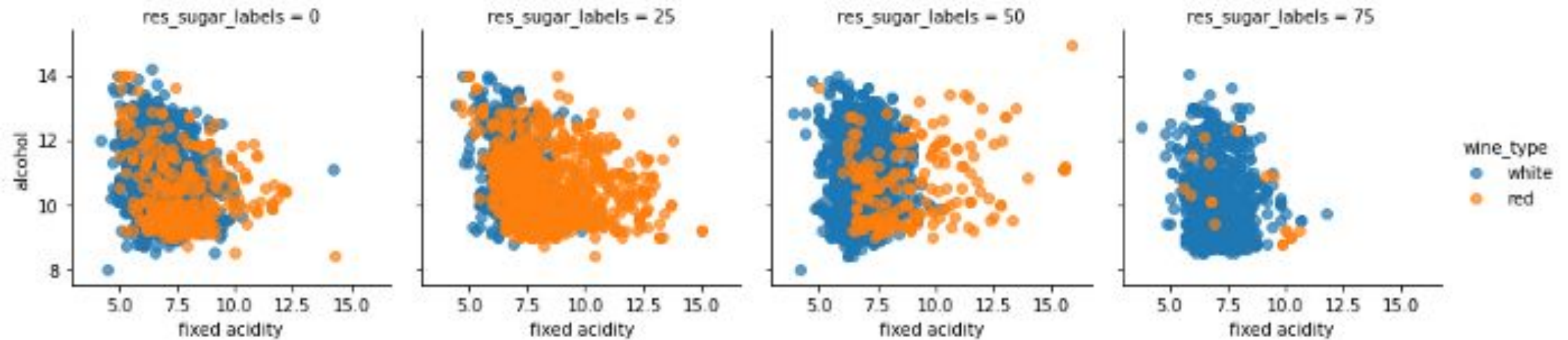
Parallel coordinates

Multivariados



**Scatter plot con variables codificadas
en transparencia**

Multivariados



Cualquier combinación de tipo de gráfico, agregar presencia y características

Veremos más la siguiente lección

Práctica 8.1: Explorando datos gráficamente

- Partamos del conjunto de datos de Equipos/Jugadores
- Carga los datos a un dataset de Pandas
- Selecciona una de las librerías que vimos al inicio de clase de visualización (Excepto vega y d3)
 - Matplotlib
 - Dataframe.plot
 - Seaborn
 - Ggplot
 - Plot.ly
 - Bokeh

(No olvidar que en jupyter tengas la magia `%matplotlib inline`)

- Averigua cómo realizar gráficos univariados, bivariados y multivariados.
- Encuentra al menos 2 insights por cada nivel de análisis, y prepara una presentación para compartirlos

30 minutos

¿Preguntas?

Referencias

- Sudheer, Sruthi., “12 Uni-variate Data Visualizations With Illustrations in Python.” Analytics Vidhya. 29 Jul, 2020,
medium.com/analytics-vidhya/12-uni-variate-data-visualizations-with-illustrations-in-python-a8fbdad4bd04
- Sarkar, Dipanjan., “Effective Visualization of Multi-Dimensional Data — A Hands-on Approach.” Start it up, 12 Dec, 2018.
medium.com/swlh/effective-visualization-of-multi-dimensional-data-a-hands-on-approach-b48f36a56ee8

Contacto

Mtro. Ricardo Daniel Alanis Tamez

ricardo@codeandomexico.org

LinkedIn: Ricardo Alanís