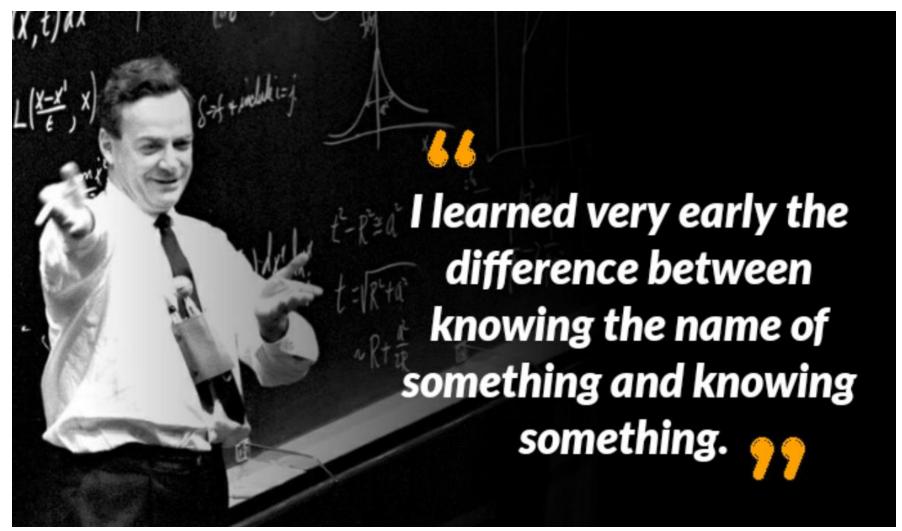
# Visualización Interactiva de la Información

## **B. J. Ferro Castro**





# Visualización Interactiva de la Información B. J. Ferro Castro

## 01 Presentación y Prerrequisitos



Doctor en Ciencias Técnicas, en el Instituto de <u>RadioIngeniería</u> y Electrónica de la Academia de Ciencias de la República Checa en el área de Procesamiento de Señales mediante microcomputadoras.

Experiencia docente de más de 30 años en el área de la Computación. En los últimos 8 años se ha dedicado mayormente a la docencia de Licenciatura y Posgrado con certificaciones en competencias de los programas de desarrollo de habilidades docentes. Formador de recursos humanos de licenciatura y posgrado acreditada con más de 15 tesistas de licenciatura, 25 tesistas de maestría y seis tesistas de doctorado, todos graduados en la actualidad.



# Algunas cosas que veremos en el curso

- ¿Por qué hacer visualizaciones?
- Evolución de la visualización
- Visualizaciones históricas
- Visualización sobre mapas
- Características de los datos
- Limpieza de datos
- Creación de nuevos datos
- Principios para diseñar buenas gráficas
- Análisis de Redes Sociales \*
- Dashboards
- Visualización con Python\*
- Sesiones de preguntas y respuestas (selección del foro)
- Sesiones de revisión de las mejores tareas
- Recursos de Inspiración



# Herramientas (prerrequisitos)

## Gráficas básicas y Dashboards

Excel, Google Sheets, Zoho, Open Office, Libre Office, etc.

## Visualización programática

- Conocimiento básico de cualquier lenguaje de programación
- Usaremos Python, se recomienda instalar jupyter (pip install jupyter)

### Análisis de Redes Sociales

- Tener una versión de Java actualizada
- Usaremos Gephi <a href="http://gephi.org">http://gephi.org</a>
- Instalar el plugin TwitterStreamerImporter
- Tener cuenta de Twitter

## Durante todo el curso

- Opcional: un micrófono que funcione
- Opcional: cámara deseable
- Deseable: Llenar el perfil del estudiante



# Estructura de los Foros

# La participación es muy importante, se tomará en cuenta en la actividad de "foro de participación"

La participación no solo es interactuar con el profesor, sino entre compañeros

## Pregúntale al Profesor

- El foro "principal" y oficial del curso
- Todas las preguntas y dudas sobre la materia
- El profesor debe contestar en 48hs o menos

### Tus Visualizaciones

- Foro para mostrar los trabajos de visualización que vayas desarrollando durante el curso, pueden ser de tareas o por iniciativa propia
- Criticar constructivamente los trabajos de los demás y recibir crítica de los propios



# Estructura de los Foros

# La participación es muy importante, se tomará en cuenta en la actividad de "foro de participación"

La participación no solo es interactuar con el profesor, sino entre compañeros

## Las mejores y peores visualizaciones

- Comparte visualizaciones que te inspiren y explica porqué
- Comparte las peores visualizaciones que te encuentres, explica porqué son malas y que harías para mejorarlas

### Otros

- En este foro "todo vale", hablaremos de cosas generales relacionadas a la visualización, y sobre otras cosas que puedan ser interesantes para el grupo
- Presentaciones personales, intereses, aplicaciones de la visualización
- Ligas a recursos
- Artículos
- Podcasts



# **Otros criterios**

Tareas con entrega tardía

## Cero tolerancia a plagios

- Por favor, citen bien sus fuentes. Una cita no es de cinco párrafos. Se descontarán puntos a una tarea/trabajo mal citado, dependiendo de la gravedad
- Cualquier plagio detectado, será calificado con cero y reportado a la Universidad para que se analice el caso y se actúe en consecuencia



## **Tareas**

#### Fecha Tarea

#### Visualización de Perfiles

- No hay una gráfica perfecta para cada ocasión, pero hay mejores gráficas que otras
- Piensen qué gráfica es compacta y permite comparar diferentes variables entre sujetos
- Piensen en el público objetivo
- Piensen en la gráfica como parte de un informe de recursos humanos
- No solo describan sus gráficas, tomen una decisión con ella

### Trabajo colaborativo de visualización de datos

- Hay mucha información como para hacer un trabajo muy interesante
- Desafortunadamente muy desactualizada
- Si pueden conseguir datos actualizados sería muy interesante
- Piensen no solo en visualizaciones estáticas
- En el material de UNIR hay muchas herramientas que podrían utilizar

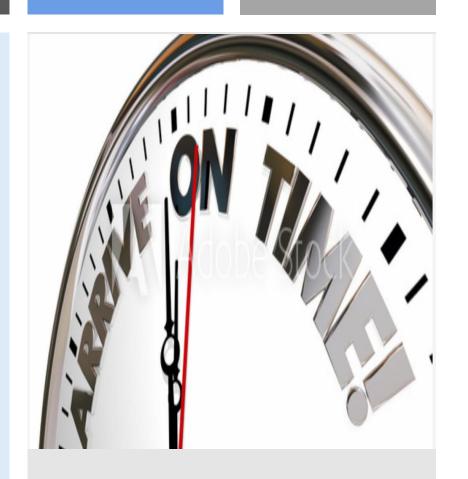
### Rediseño: Emigración

- Esta es la tarea que mayor latitud les da para ser creativos, aprovéchenla
- Este juego de datos se puede mezclar muy bien con otros datos
- Hay que tener mucho cuidado en la información y cómo se presenta



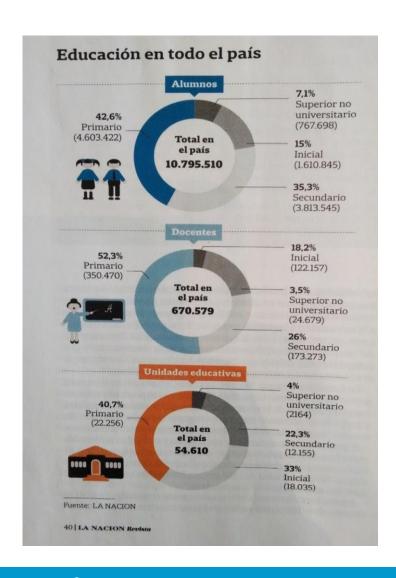
# Tareas

Caso práctico: Rediseño. La	28 de jun en 23:59
educación en Argentina	2.5 puntos posibles2.5 pts
Caso práctico: Rediseño. Time vs.	5 de jul en 23:59
Newsweek	2.5 puntos posibles2.5 pts
Laboratorio: Trabajo colaborativo de	26 de jul en 23:59
visualización de datos	2.5 puntos posibles2.5 pts
Caso práctico: Visualizar perfiles	2 de ago en 23:59
	2.5 puntos posibles2.5 pts
Trabajo: Rediseño. Emigración	9 de ago en 23:59
	2.5 puntos posibles2.5 pts
Foro: Restos del open data.	No disponible hasta 24 jul en 17:00 24 de jul en 17:00





# Tarea 1: Rediseño: Educación en Argentina



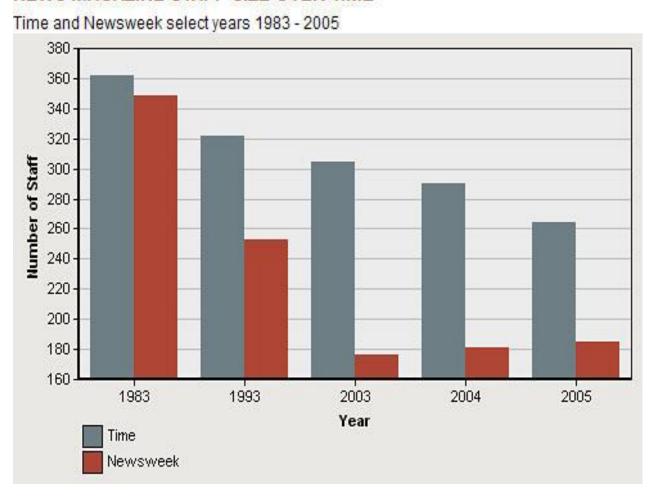
Analiza el gráfico e indica brevemente (no más de uno o dos párrafos) cuáles son los principales problemas que presenta.

Realiza un rediseño del gráfico utilizando los mismos datos. El gráfico lo puedes realizar con la aplicación que desees (Excel, Tableau...) o incluso dibujarlo a mano (y escanearlo).

Incluye una breve justificación de por qué la propuesta que haces es mejor que la original (uno o dos párrafos).

## Tarea 2: Rediseño: Time vs. Newsweek

#### NEWS MAGAZINE STAFF SIZE OVER TIME



Analiza el gráfico e indica brevemente (no más de uno o dos párrafos) cuáles son los principales problemas que presenta.

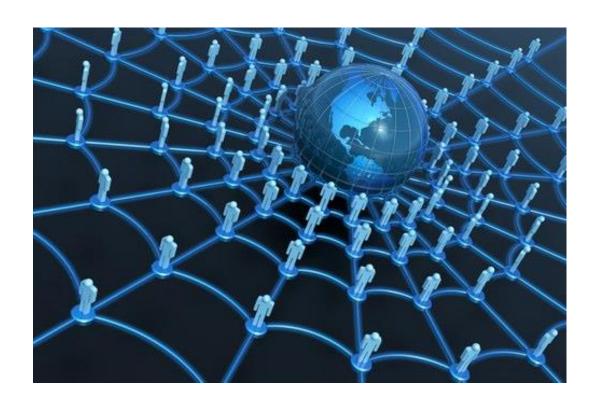
Realiza un rediseño del gráfico utilizando los mismos datos. El gráfico lo puedes realizar con la aplicación que desees (Excel, Tableau, Python) o incluso dibujarlo a mano (y escanearlo).

Incluye una breve justificación de por qué la propuesta que haces es mejor que la original (uno o dos párrafos).



# Tarea 3: Trabajo colaborativo de visualización de datos :

Revealing the People Defining Social Networks – Brian Solis

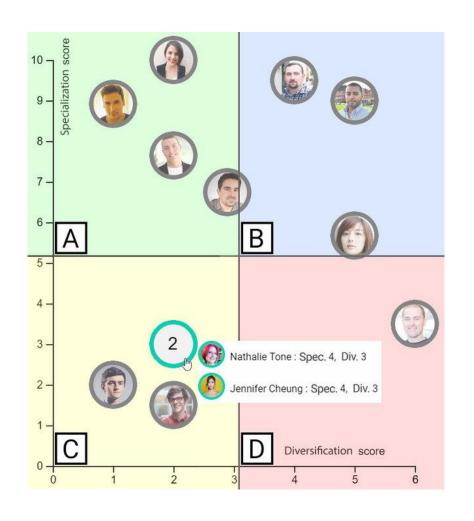


En 2009, Brian Solis publicaba algunos datos muy interesantes sobre los usuarios de las principales redes sociales (Facebook, Twitter, YouTube...), como datos sobre su género, ingresos, nivel de estudios, etc.

El objetivo del laboratorio es elaborar una visualización sobre los aspectos más interesantes, reveladores o llamativos que ofrezcan estos datos.



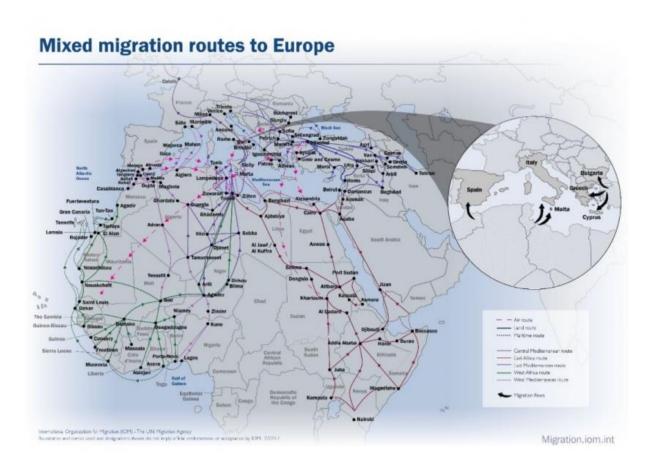
## Tarea 4: Visualización de Perfiles



Un departamento de RRHH dispone de una base de datos con el perfil de candidatos. Entre los datos de que dispone la base de datos se encuentra la valoración cuantitativa del perfil de los candidatos, a través de 6 áreas (Experience, Communication, Friendliness, Subject matter knowledge, Presentation, Education).

Debes elaborar dos visualizaciones gráficas diferentes de esos datos, cada una con un propósito o función diferenciada

# Tarea 5: Rediseño. Emigración



### http://yusef.es/ejemplosunir/emigracionCSV.csv

Analiza el gráfico de emigración española en el documento de tarea e indica brevemente (no más de uno o dos párrafos) cuáles son los principales problemas que presenta.

Realiza un rediseño del gráfico utilizando los mismos datos.

Incluye una breve justificación de por qué la propuesta que haces es mejor que la original (uno o dos párrafos).

# Infografía y visualización de datos

¿En qué coinciden, en qué se distinguen?

Semana 3
24/05/2021 -
28/05/2021

# Tema 1. Introducción a la visualización de datos (continuación)

- 1.4. Importancia de la infografía y la visualización de datos
- 1.5. Estadios de la visualización
- 1.6. Referencias bibliográficas

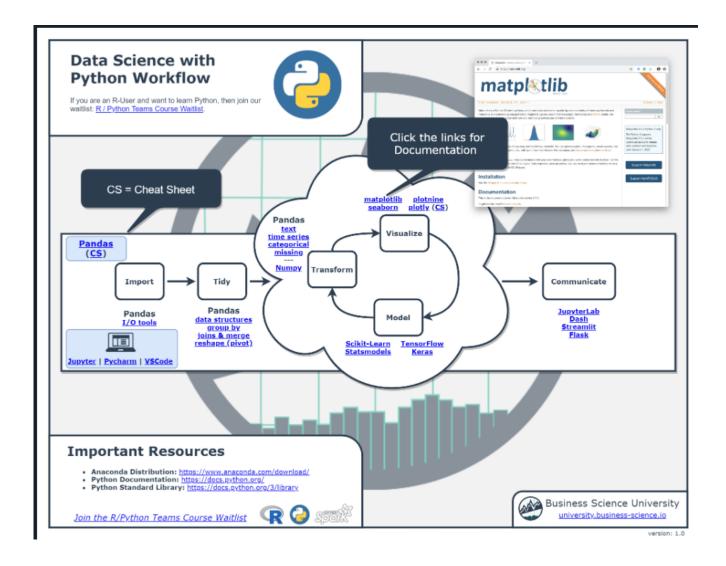
**Test** Tema 1 (0,1 puntos)

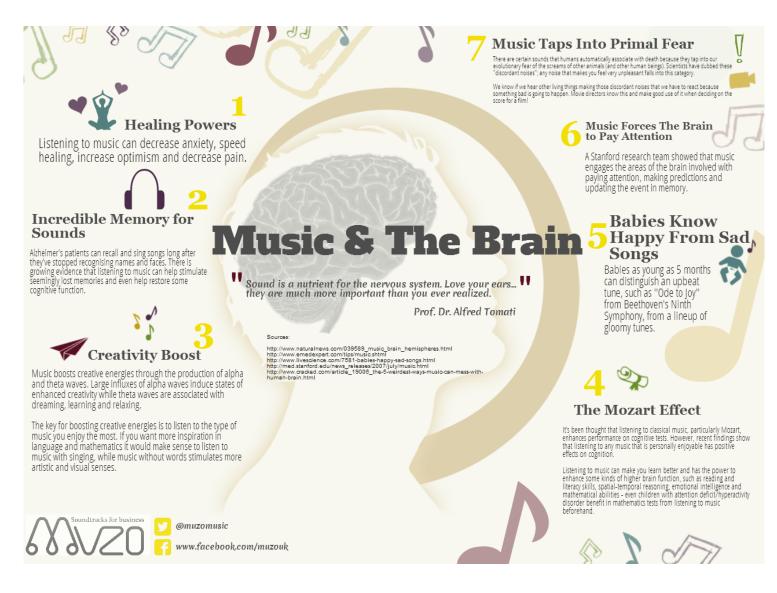


Infografía:
(a picture or diagram or a group of pictures or diagrams showing or explaining information)

Una infografía es una representación visual de una gran colección de información. Cuenta una historia más completa que una sola visualización de datos y, a menudo, ofrece subjetividad.

Si bien una infografía puede cubrir un tema con muchos datos, también puede explorar uno muy subjetivo, como las mejores rutas marítimas en el océano o un proceso de diseño de software más efectivos.

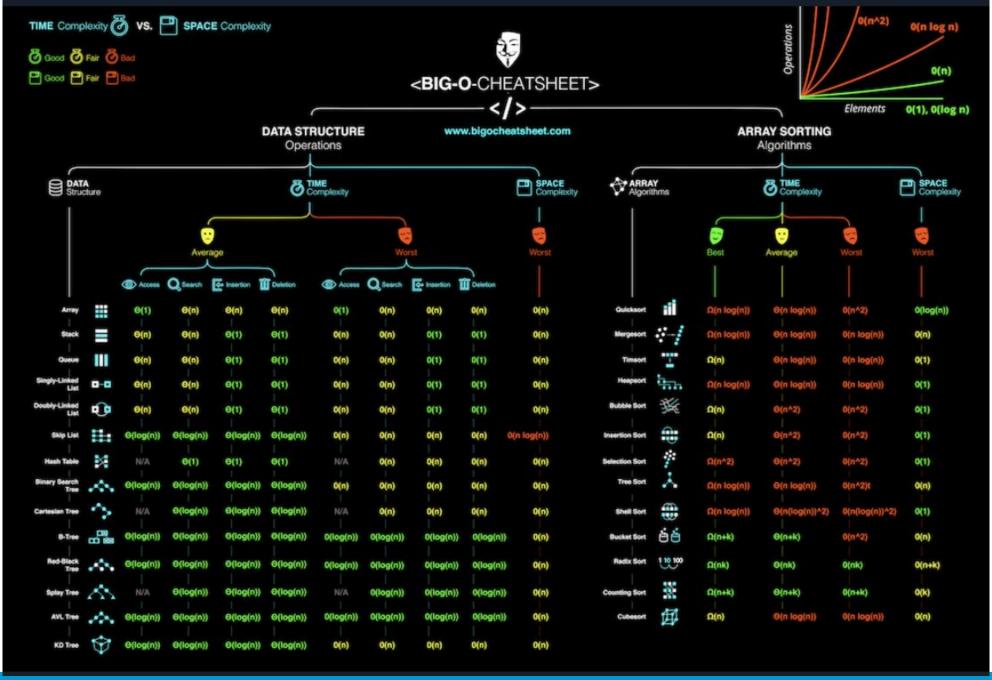




What does listening to music actually do to you? How does it boost creativity, memory and concentration? Muzo Music's graphic gathers together some of the fascinating thoughts and theories.

Classical music - in incredible infographics - Classic FM





Una visualización de datos es una representación de un conjunto específico de datos numéricos o categóricos. Presenta información objetiva y cuantificable. Las visualizaciones de datos suelen centrarse en un conjunto de datos específicos. Convierten lo que de otro modo podría ser una colección complicada de números o cantidades discretas en algo que se puede comprender de un vistazo (20 20)

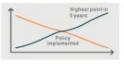
## **Core Principles of Data Visualization**

#### Audience



Always consider your audience -whether they need a short, written report, a more in-depth paper, or an online exploratory data tool.

#### Include annotation



Add explanatory text to help the reader understand how to read or use the visualization (if necessary) and also to guide them through the content.

#### Use pie charts with care



We are not very good at discerning quantities from the slices of the pie chart. Other chart types-for example, bars, stacked bars, treemaps, or slope charts-may be a better choice.

#### Avoid 3D

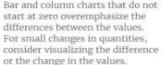


Using 3D when you don't have a third variable will usually distort the perception of the data and should thus be avoided.

#### Start bar and column charts at zero



start at zero overemphasize the differences between the values.



#### Make labels easy to read



When applicable, rotate bar and column charts to make the labels horizontal. If possible, make vertical axis labels horizontal, possibly below the title. In general, make labels clear, concise, and easy for your reader to understand.

#### Try small multiples



Breaking up a complicated chart into smaller chunks can be an effective way to visualize your data.

#### Use maps carefully



Use maps carefully, always being sure it is the geographic point you are trying to make. Column and bar charts, for example, are often better at enabling comparisons between geographic units.

#### Color and font considerations



Avoid default colors and fonts-they all look the same and don't stand out.

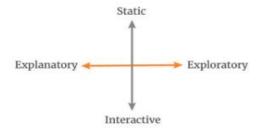


Consider color blindnessabout 10% of people (mostly men) have some form of color blindness.



Avoid the rainbow color palette-it doesn't map to our number system and there is no logical ordering.

#### Visualization Mapping: Form and Function





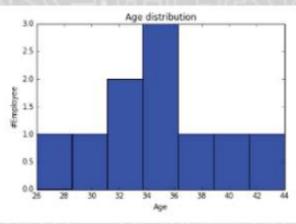




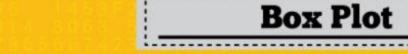
## Histogram

fig=plt.figure()
ax = fig.add\_subplot(1,1,1)
ax.hist(df['Age'],bins = 7) # Here you can
play with number of bins Labels and Tit
plt.title('Age distribution')
plt.xlabel('Age')

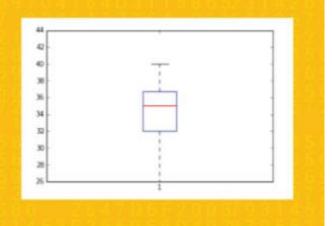
plt.ylabel('#Employee') plt.show()





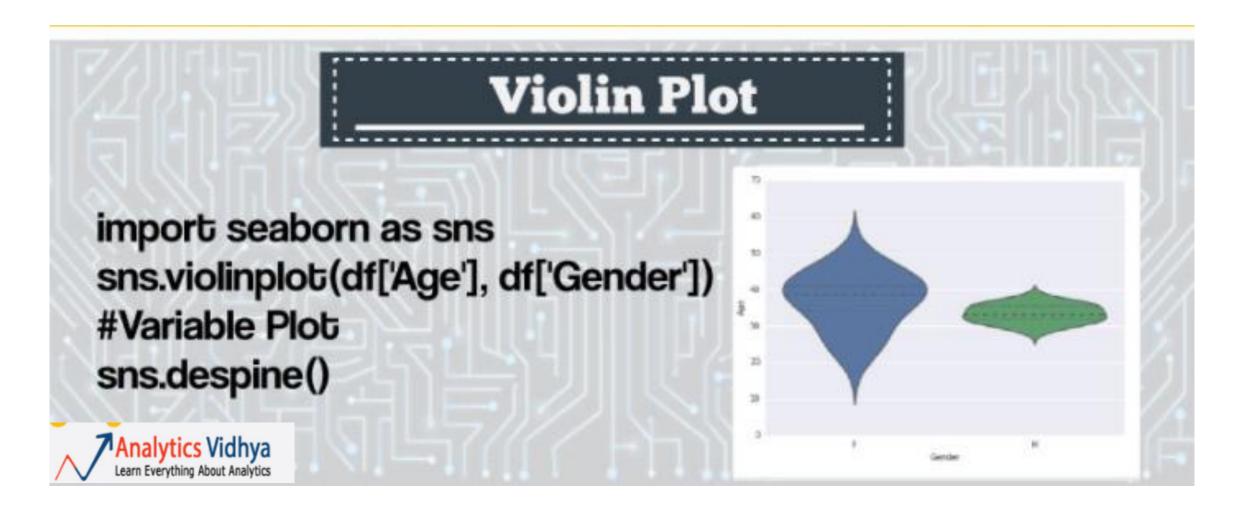


importmatplotlib.pyplot as plt import pandas as pd fig=plt.figure() ax = fig.add\_subplot(1,1,1) x.boxplot(df['Age']) plt.show()













## **Bar Chart**

var = df.groupby('Gender').Sales.sum()

#grouped sum of sales at

Gender level

fig = plt.figure()

ax1 = fig.add\_subplot(1,1,1)

ax1.set\_xlabel('Gender')

ax1.set\_ylabel('Sum of Sales')

ax1.set\_title("Gender wise Sum of Sales")

var.plot(kind='bar')







## **Line Chart**

var = df.groupby('BMI').Sales.sum()

fig = plt.figure()

ax1 = fig.add\_subplot(1,1,1)

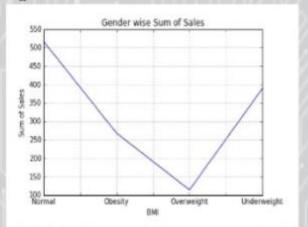
ax1.set\_xlabel('BMI')

ax1.set\_ylabel('Sum of Sales')

ax1.set\_title("BMI wise Sum

of Sales")

var.plot(kind='line')

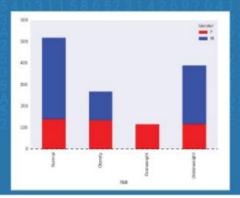






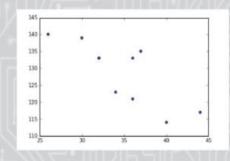
## **Stacked Column Chart**

var = df.groupby(['BMl','Gender']).Sales.sum()
var.unstack().plot(kind='bar',stacked=True, color=['red','blue'], grid=False)



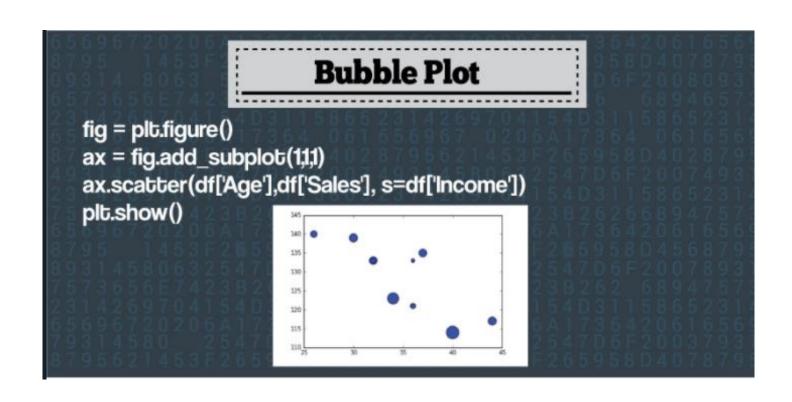
### **Scatter Plot**

fig = plt.figure()
ax = fig.add\_subplot(1,1,1)
ax.scatter(df['Age'],df['Sales'])
plt.show()











## **Pie Chart**

var=df.groupby(['Gender']).sum().stack()

temp=var.unstack()

type(temp)

x\_list = temp['Sales']

label\_list = temp.index

pyplot.axis("equal") #The pie chart

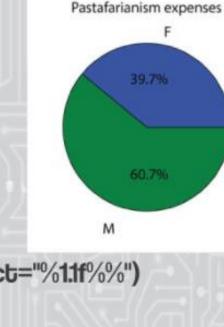
is oval by default. To make it a

circle use pyplot.axis("equal")

plt.pie(x\_list,labels=label\_list,autopct="%1.1f%%")

plt.title("Pastafarianism expenses")

plt.show()



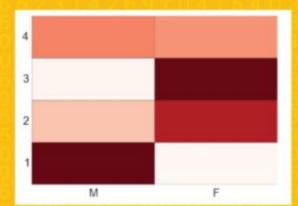




## **Heat Map**

import numpy as np data = np.random.rand(4,2) rows = list('1234') #rows categories columns = list('MF') #column categories fig,ax=plt.subplots()

ax.set\_xticks(np.arange(0,2)+0.5)



ax.set\_yticks(np.arange(0,4)+0.5)
ax.xaxis.tick\_bottom()
ax.yaxis.tick\_left()
ax.set\_xticklabels(columns,minor=False,fontsize=20)
ax.set\_yticklabels(rows,minor=False,fontsize=20)
plt.show())

ax.pcolor(data,cmap=plt.cm.Reds,edgecolors='k')



Una breve introducción



1. Investigación

2. Selección de Datos

3. Creación de un Boceto

4. Elaboración

5. Percepción y Conocimiento

1. Investigación

2. Selección de Datos

3. Creación de un Boceto

4. Elaboración

5. Percepción y Conocimiento

### Investigación

- Búsqueda de información de fuentes sólidas
- Algunas características de la información con buena calidad son:
  - Completa: Es muy común que la información esté incompleta para ciertos registros, generalmente no se puede hacer mucho al respecto, pero hay que tener estrategias
  - Confiable: conviene más que provenga de fuentes oficiales o de reputación, o la recolección por nosotros mismos
  - Precisa: diferentes problemas requieren diferentes niveles de precisión
  - Actualizada: (o que corresponda al periodo que se quiere investigar)
  - **Verificable**: Que la información se pueda demostrar y comprobar
  - Relevante: deben ser hechos relacionados al fenómeno
  - Simple: Que la interpretación de la información sea fácil



1. Investigación

2. Selección de Datos

3. Creación de un Boceto

4. Elaboración

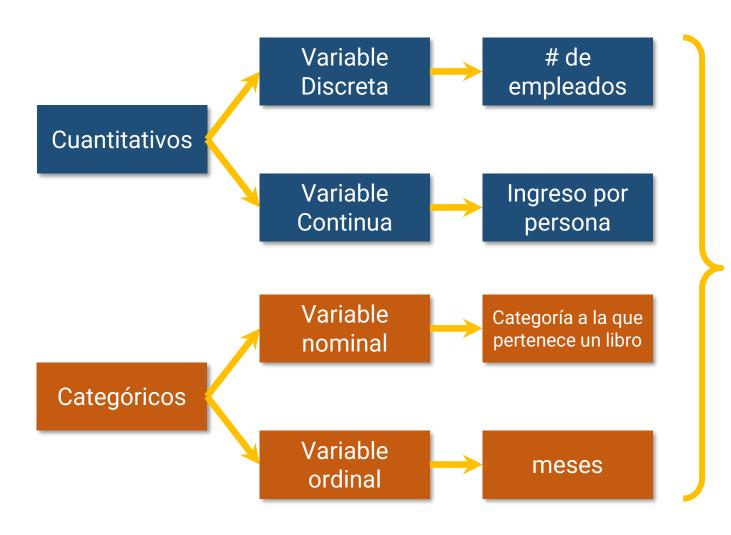
5. Percepción y Conocimiento

### Investigación

- Búsqueda de información de fuentes sólidas
- Algunas características de la información con buena calidad son:
  - Completa: Es muy común que la información esté incompleta para ciertos registros, generalmente no se puede hacer mucho al respecto, pero hay que tener estrategias
  - Confiable: conviene más que provenga de fuentes oficiales o de reputación, o la recolección por nosotros mismos
  - Precisa: diferentes problemas requieren diferentes niveles de precisión
  - Actualizada: (o que corresponda al periodo que se quiere investigar)
  - **Verificable**: Que la información se pueda demostrar y comprobar
  - Relevante: deben ser hechos relacionados al fenómeno
  - Simple: Que la interpretación de la información sea fácil



# **Tipos de Datos**



Los tipos de datos **se pueden codificar** de
forma diferente en una
visualización



1. Investigación

2. Selección de Datos

3. Creación de un Boceto

4. Elaboración

5. Percepción y Conocimiento

### Creación de un boceto

- No es necesario centrarse en análisis muy complejos
- "Tu gran idea es tan buena como tu capacidad para comunicarla"
- No tiene que ser de alta calidad, es una herramienta para darnos una idea únicamente
- No solo es necesario hacer visualizaciones técnicamente buenas, sino que también es necesario tener una buena idea de:
  - Diseño visual
  - Contar historias (storytelling) → Comunicación
  - Cautivar a la audiencia (TED Talks) → Comunicación

1. Investigación

2. Selección de Datos

3. Creación de un Boceto

4. Elaboración

5. Percepción y Conocimiento

### Elaboración

- Una vez que tenemos nuestro boceto, podemos pasar a la acción y construir la visualización que deseamos
- En este curso usaremos varias herramientas:
  - Google Maps
  - Dashboards
  - Python

### Percepción y Conocimiento

 Si transmitimos bien el mensaje y se hizo una visualización agradable, seguramente quedará en la memoria de nuestros lectores



# **Casi terminamos**

Veamos la arquitectura de un paquete de visualización: Matplotlib



Arquitectura de tres capas:

Backend

**Artist** 

Scripting



#### Arquitectura de tres capas:

#### Backend

Esta es la capa más inferior donde los gráficos se muestran en un dispositivo de salida. Puede ser cualquiera de las interfaces de usuario que admite Matplotlib.

Dos tipos:

Backends de interfaz de usuario (para su uso en pygtk, wxpython, tkinter, qt4 o macosx, etc., también denominados "backends interactivos")

Backends de copia impresa para crear archivos de imagen (PNG, SVG, PDF, PS; también denominados "backends no interactivos ").



Arquitectura de tres capas:

**Artist** 

Esta es la capa intermedia del stack. Matplotlib usa el objeto artist para dibujar varios elementos del gráfico. Cada elemento que vemos en el gráfico es un artista. Esta capa proporciona una API orientada a objetos para trazar gráficos con la máxima flexibilidad. Esta interfaz está destinada a programadores experimentados de Python, que pueden crear aplicaciones de dashboards complejas.



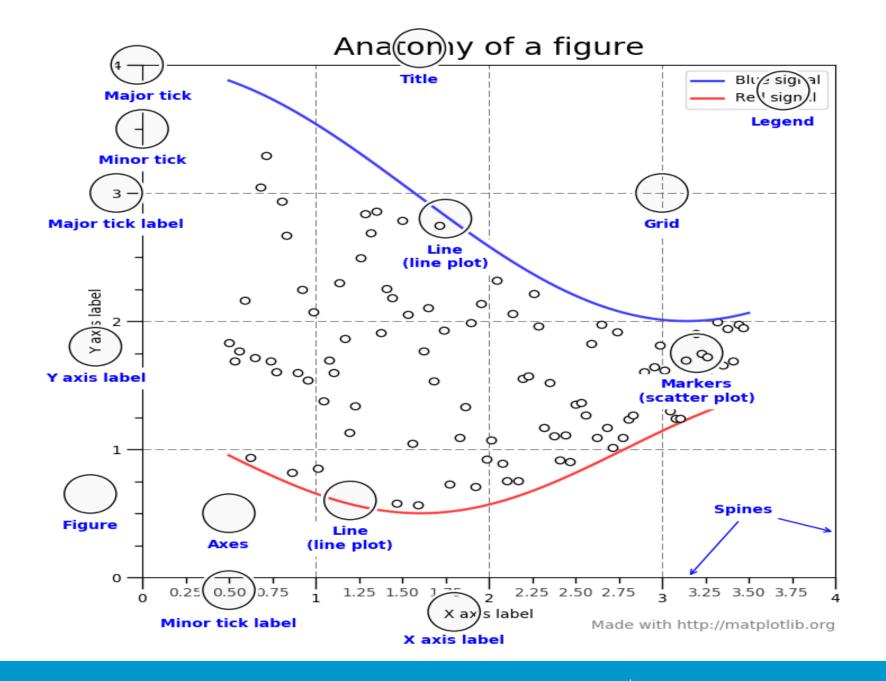
Arquitectura de tres capas:

Scripting

Esta es la capa superior de la pila. Esta capa proporciona una interfaz sencilla para crear gráficos. Está destinada a usuarios finales que no tienen mucha experiencia en programación. El API de pyplot.

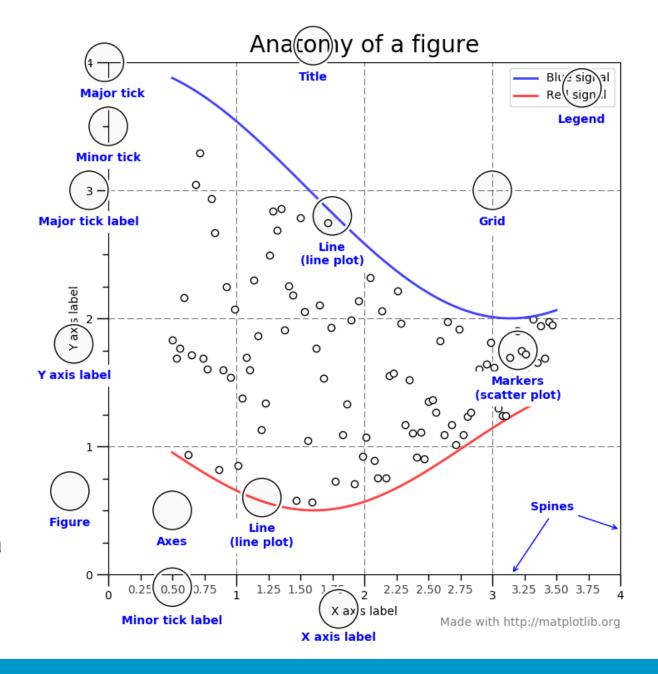


### **Figure**



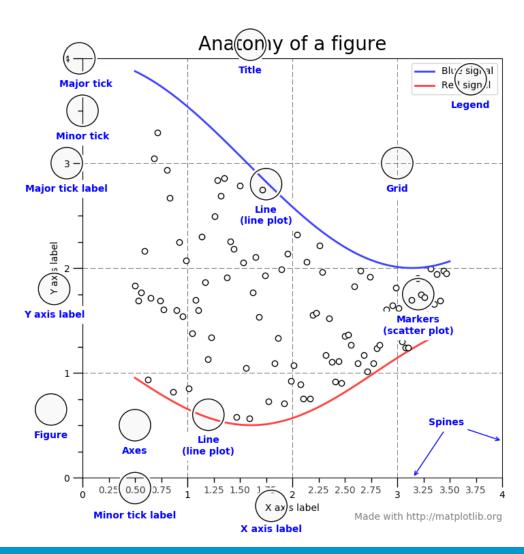
# Elementos de una Figura

El objeto matplotlib de alto nivel que contiene todos los elementos del gráfico de salida se llama *Figure*. Se pueden organizar varios gráficos de diferentes formas para formar una figura. Cada uno de los elementos de la Figura es personalizable.



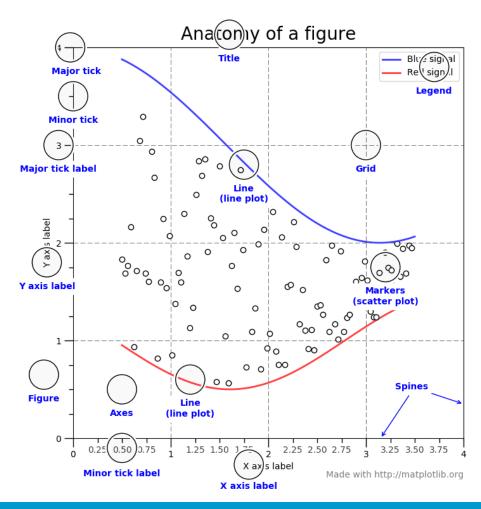


# **Figure**



La imagen completa que hemos mostrado es el objeto Figure. Contiene todos los elementos de la imagen.

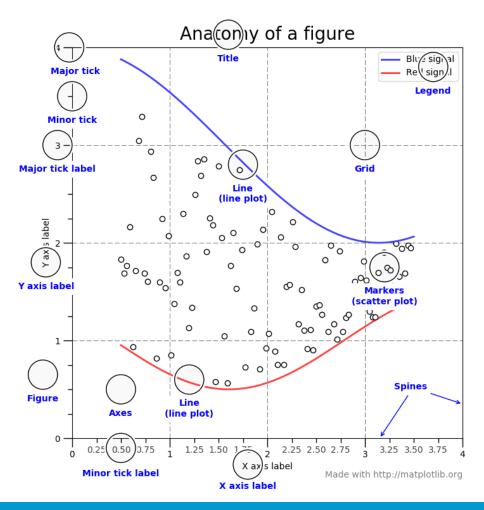
#### Axes



Axes es una subsección de Figura, donde se traza un gráfico. Axes tiene un título, una etiqueta x y una etiqueta y. Una figura puede tener muchos axes, cada uno de los cuales representa uno o más gráficos. En la figura anterior, solo hay un eje, dos gráficos de líneas en colores azul y rojo.

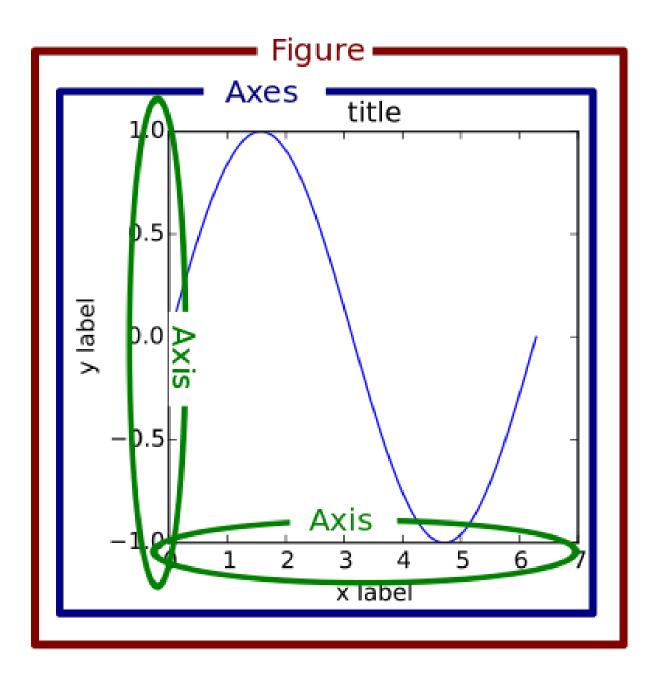


#### Axis

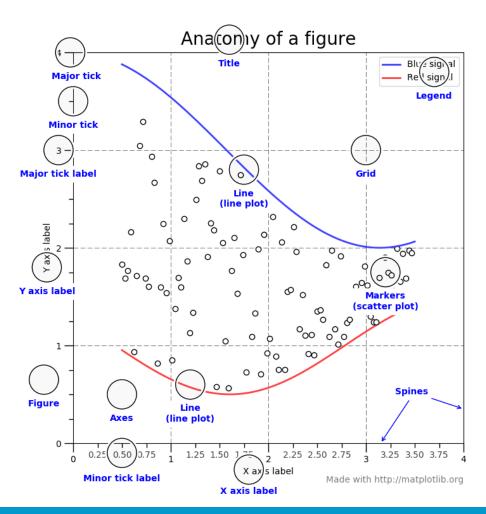


Estas son líneas numéricas que representan la escala de los gráficos que se trazan. Los gráficos 2D tienen el eje X y el eje Y, los gráficos 3D tienen el eje X, el eje Y y el eje Z. No se confunda entre Axes y Axis. Axis es un elemento de Axes.

#### Axes versus Axis

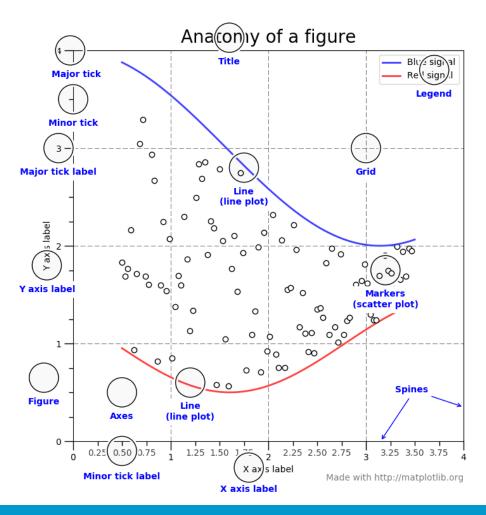


#### Labels



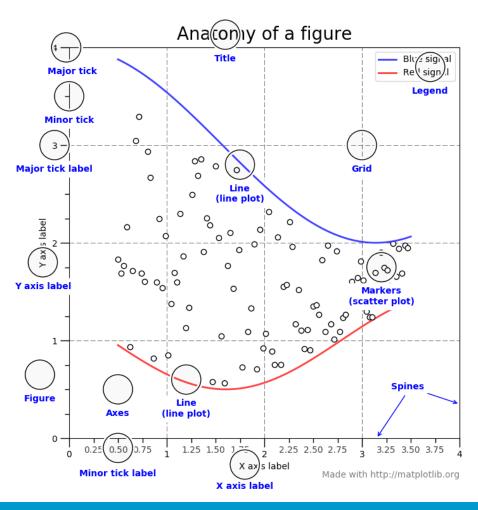
Es el nombre que se le da a varios elementos de la figura, p. Ej. Etiqueta del eje X, etiqueta del eje Y, etiqueta del gráfico (señal azul, señal roja en la figura anterior), etc.

#### Legends



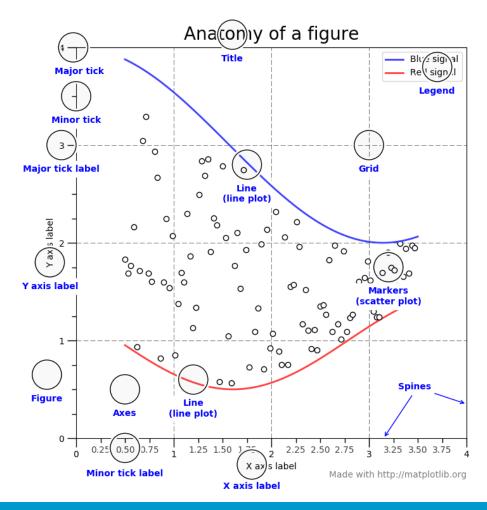
Cuando hay varios gráficos en los Axes (como en la figura), cada uno de ellos tiene su propia etiqueta y todas estas etiquetas se representan como una leyenda. En la figura anterior, la leyenda se coloca en la esquina superior derecha de la figura.

#### Title



Es el nombre que se le da a cada uno de los Axes. La figura también puede tener su propio título, cuando la figura tiene varios axes con sus propios títulos. La figura anterior tiene solo un Axes, por lo que solo hay un título para los Axes y la figura

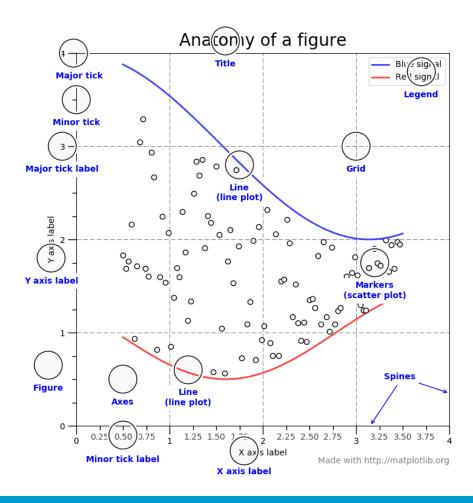
#### Tick Labels



Cada uno de los ejes (X, Y o Z) tendría un rango de valores que se dividen en muchos contenedores (bins) iguales. Los contenedores se eligen en dos niveles. En la figura, la escala del eje X varía de 0 a 4, dividida en 4 contenedores principales (0, 1, 2, 3 y 4) y cada uno de los contenedores principales se divide en 4 contenedores menores (0, 0.25, 0.5, 0,75). Los bins principales se denominan Major Tick y los Minor Bins se denominan Minor Tick, y los nombres que se les dan se denominan etiqueta de tick principal y etiqueta de tick menor.

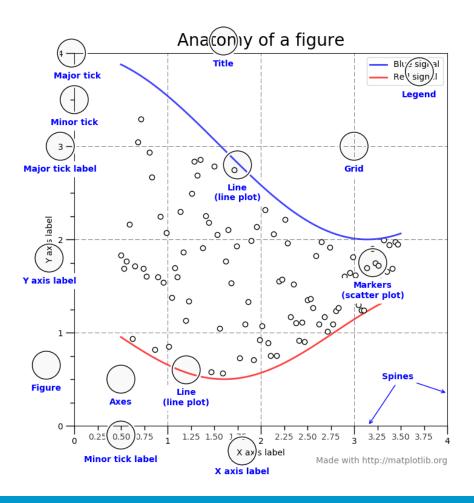


#### Spines



Los límites de la figura se llaman espinas. Habrá 4 espinas para cada figura (superior, inferior, izquierda y derecha).

#### Grid



Para facilitar la lectura de las coordenadas de varios puntos del gráfico, el área del gráfico se divide en una cuadrícula. Por lo general, esta cuadrícula se dibuja a lo largo de las marcas principales de los ejes X e Y. En la figura anterior, Grid se muestra en líneas discontinuas.

#### Modos interactivo y no interactivo

- En el modo interactivo, el gráfico se actualiza en el backend después de cada declaración. En el modo no interactivo, el gráfico no se muestra hasta que se le solicite explícitamente que lo haga. Con los siguientes comandos, el modo interactivo se puede activar o desactivar, y también se puede comprobar el modo actual en cualquier momento.
  - matplotlib.pyplot.ion() to set the interactive mode ON
  - matplotlib.pyplot.ioff() to switch off interactive mode
  - matplotlib.is\_interactive() to check whether interactive mode is ON(True) or OFF(False)



www.unir.net