



Visualización en Data Science

Diplomatura CDAAyA 2018



OUR BRAIN PROCESSES VISUALS **60,000x** FASTER THAN TEXT



90%

OF INFO TRANSMITTED
TO THE BRAIN IS VISUAL



70%

OF YOUR SENSORY RECEPTORS
ARE IN YOUR EYES



50%

OF YOUR BRAIN IS ACTIVE
IN VISUAL PROCESSING



40%

OF PEOPLE RESPOND
BETTER TO VISUALS

- Our brain process visuals 60000 faster than text
- 90% of the information transmitted to the brain is visual
- 70% of your sensory receptors are in your eyes
- 50% of your brain is active in visual processing
- 40% of people respond better to visuales



Herramienta para la comunicación

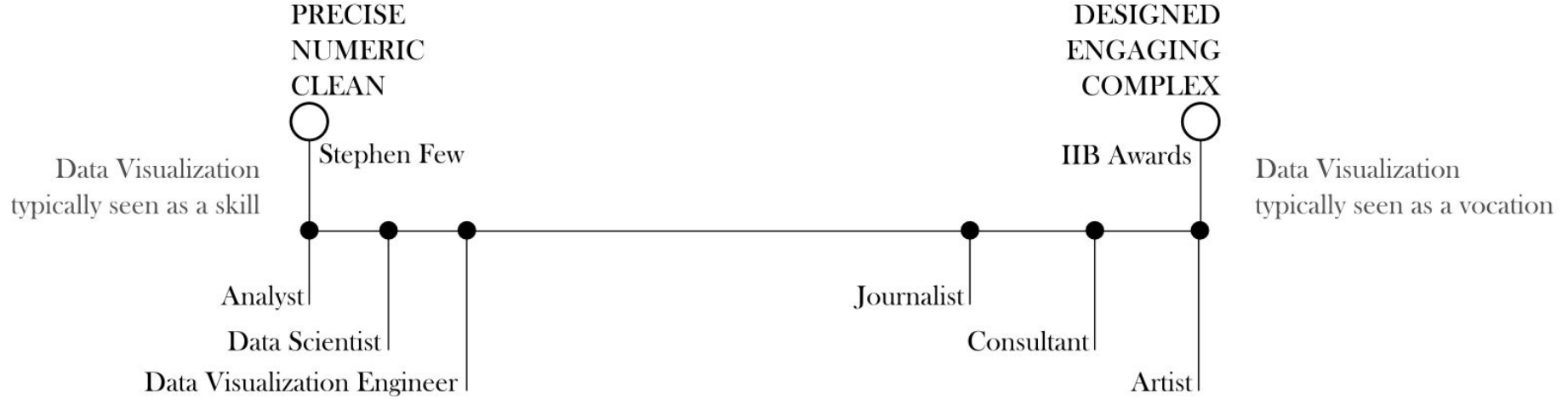




Herramienta para la compresión

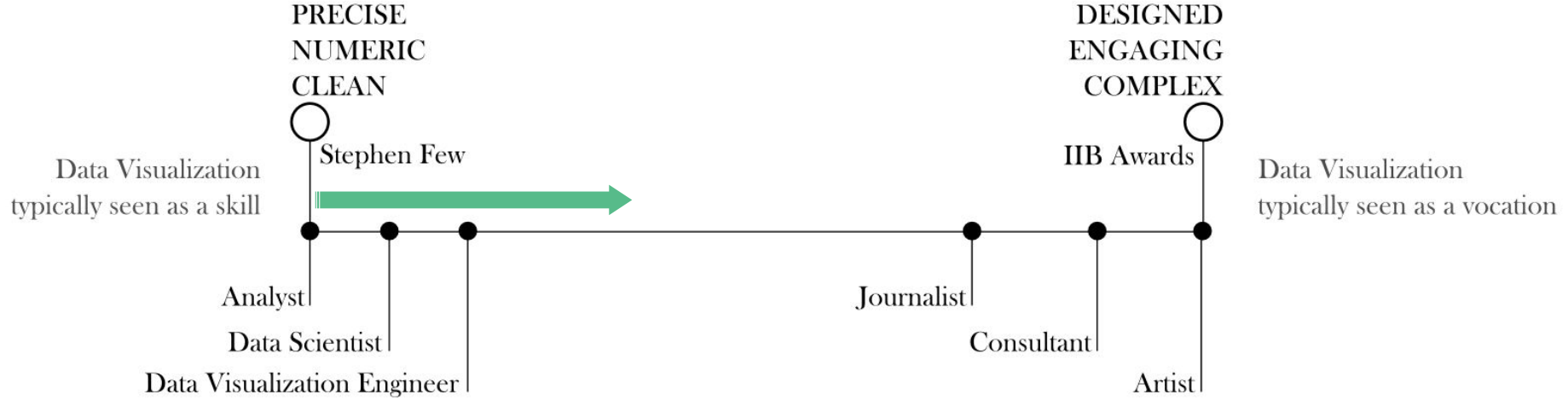


enrollment_id	username	course_id
1	9Uee7oEuuMmgPx2IzPffkHgkHZyPbWr0	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
3	1qXC7Fjbwp66GPQc6pHLfEu08WKozxG4	7GRhBDsirIGkRZBtSMEzNTyDr2JQm4xx
4	FIHlppZyoq8muPbdVxS44gfvceX9zvU7	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
5	p1Mp7WkVfzUijX0peVQKSHbgd5pXyl4c	7GRhBDsirIGkRZBtSMEzNTyDr2JQm4xx
6	dpK33RH9yepUAnyoywRwBt1AJzxGlaJa	AXUJZGmZ0xaYSWazu8RQ1G5c76ECT1Kd
7	I1KwJ6EdCZnEPLfC8Q7yWpIkLOHn7h02	7GRhBDsirIGkRZBtSMEzNTyDr2JQm4xx
9	J1oRHoSJ0InehnrxVdh32dK7QnDuCJWo	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
12	9tsGjrRgtMZ6V7yrA0yf0QPZHa1tDHAp	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
13	hDbSkVrFRj9Ryk3c5E1JYJQLyxm4jLRb	5X6FeZozNMgE2VRi3MJYjkkFK8SETtu2
14	X0hIczT5nEe052jMq1vN7QziDk8L2jnI	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
16	mPSPvu82Gr17tV9GJ95bDC7exvsVnwDE	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
18	b0Hk5D3sJulvyuC4JEm5kvAv0LAXswgQ	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
20	BoK7CAUaCFqnLgmWLxe0Hg8YkXUSeCtc	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
22	dPBUV0FPFjTZZK079rPAeq0WXhW4DUkF	7GRhBDsirIGkRZBtSMEzNTyDr2JQm4xx
23	BoK7CAUaCFqnLgmWLxe0Hg8YkXUSeCtc	AXUJZGmZ0xaYSWazu8RQ1G5c76ECT1Kd
26	vcAiZWU2sfUK00mnfjDwm0iTzACrKr78	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
28	BoK7CAUaCFqnLgmWLxe0Hg8YkXUSeCtc	TAYxxh39I2LZnftBpL0Lff2NxzrCKpkx
30	JPkczY0xyoDZBjwZAAQHmjpSvnPQzwV0	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila



Credit: [Medium Article](#)

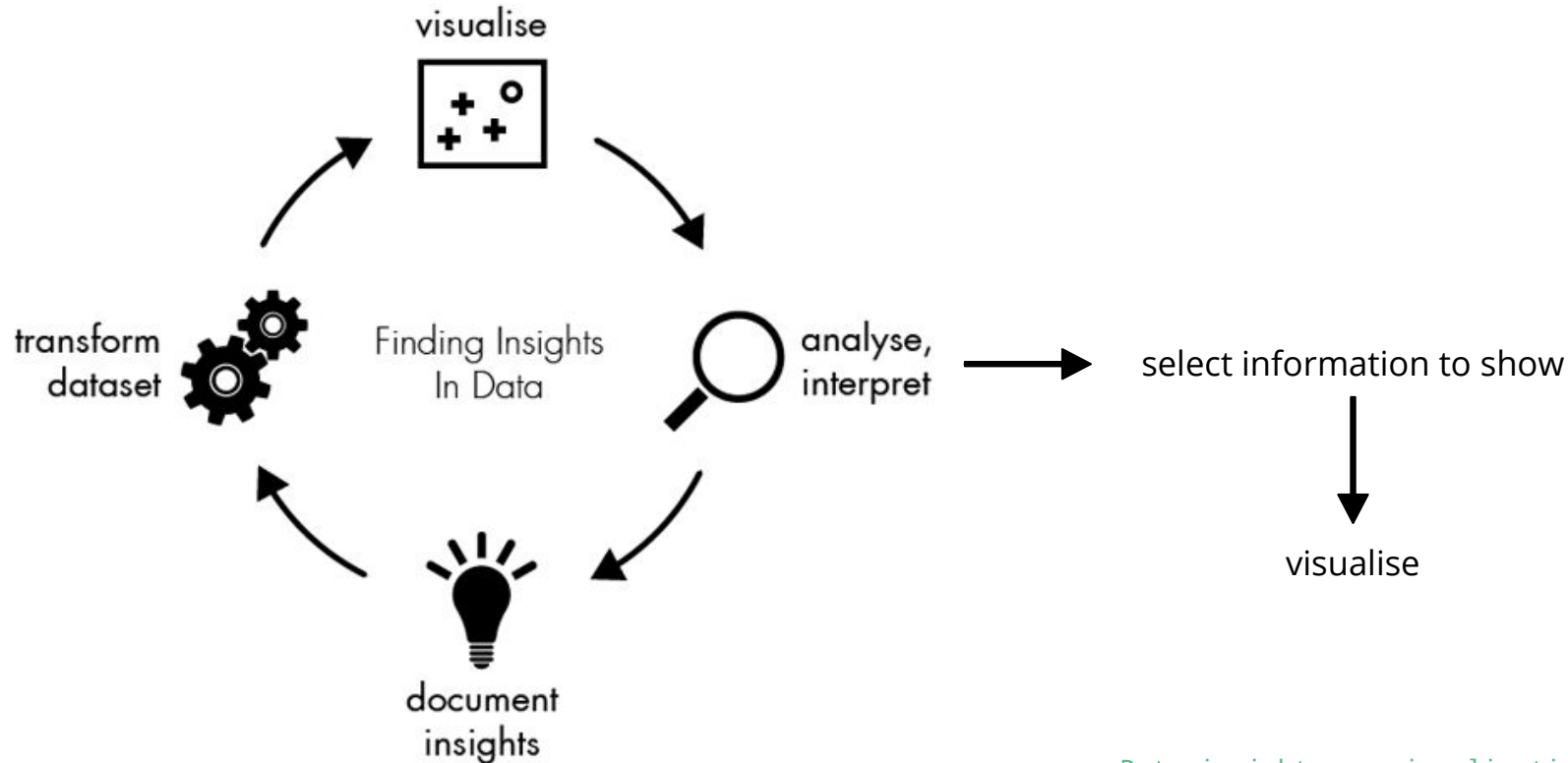
¿Podés identificar tu rol en esta línea? ¿A dónde querés llegar?



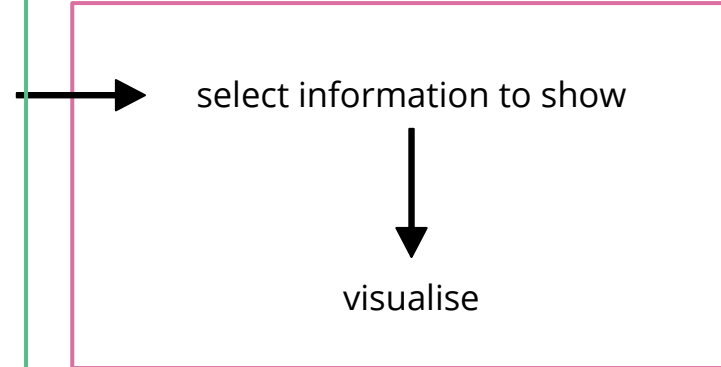
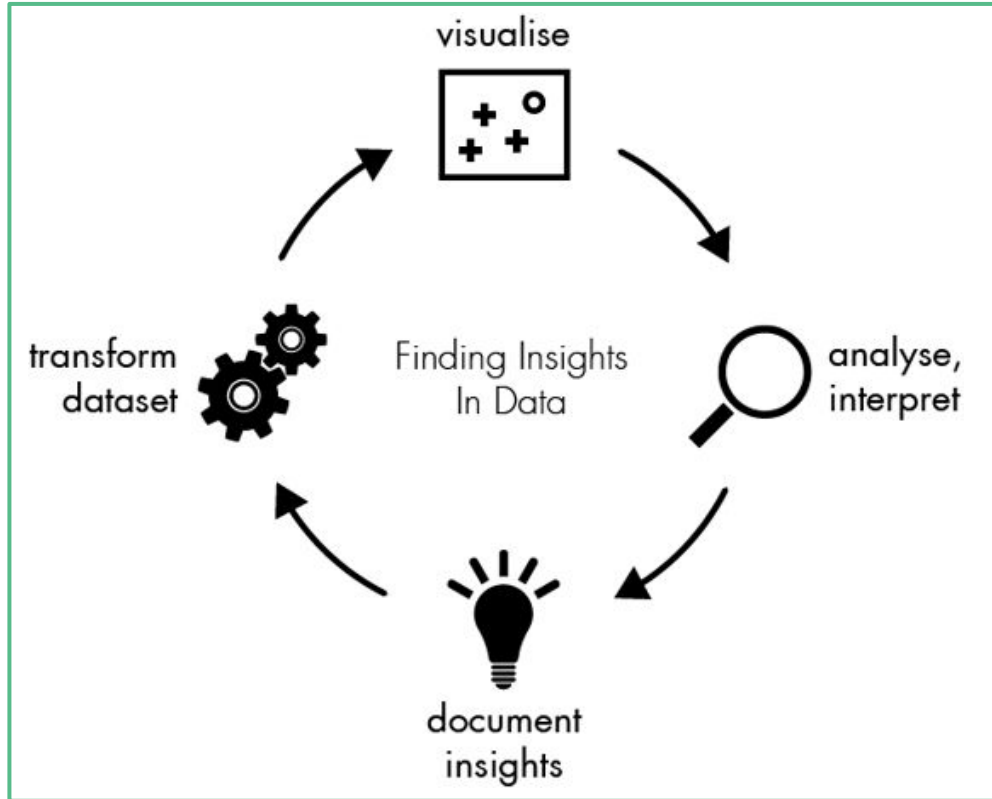
Credit: [Medium Article](#)

¿Podés identificar tu rol en esta línea? ¿A dónde querés llegar?

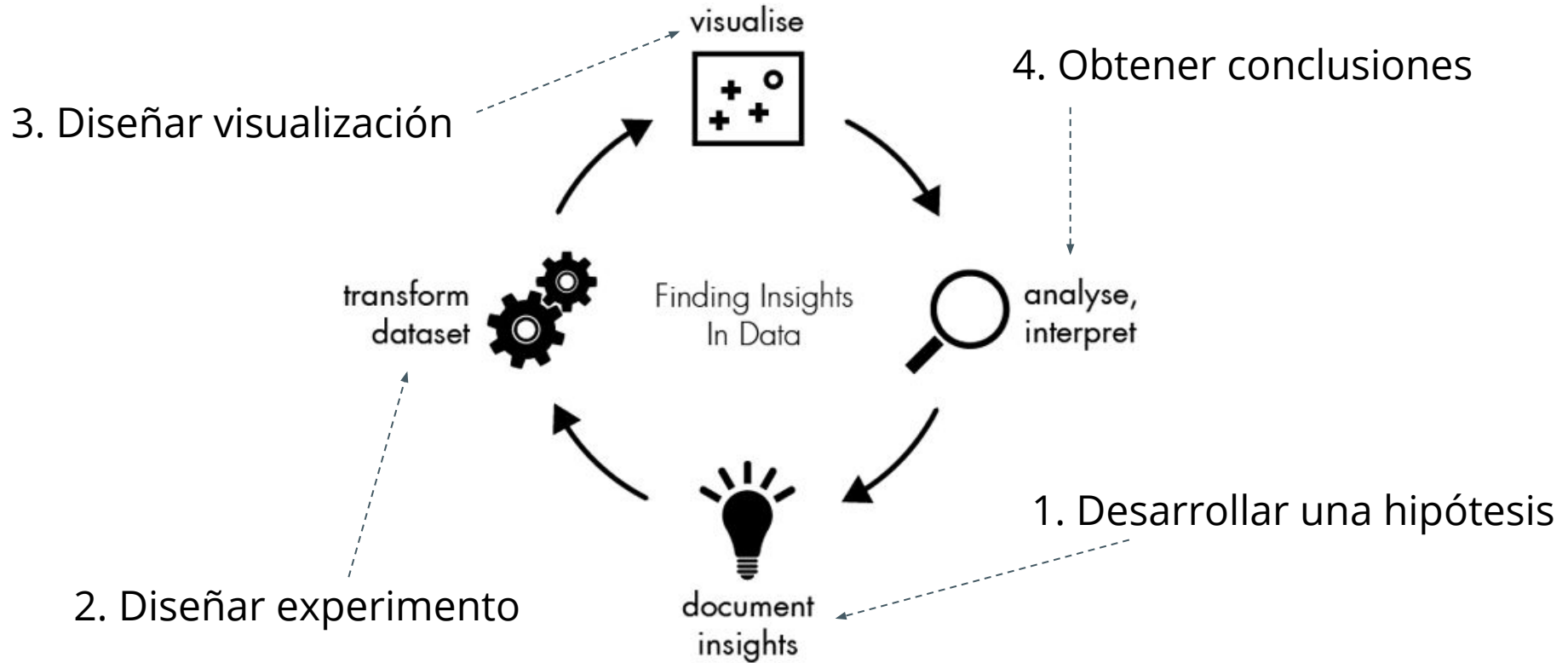
Exploración vs presentación



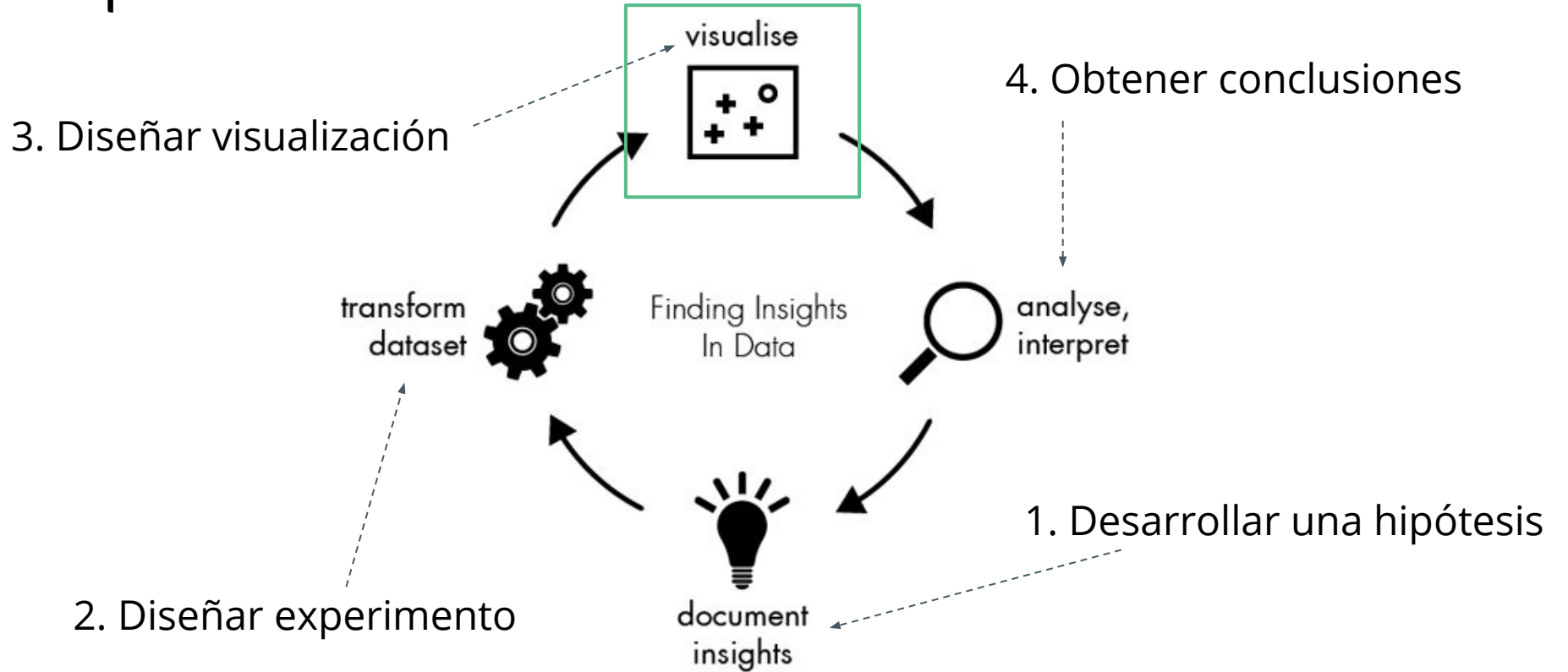
Exploración vs presentación



Exploración



Exploración



Configurar la notebook

Tipos básicos de gráficos

Tablas

- Muestran cantidades **exactas**
- Representan cualquier tipo de datos
- Son de acceso **universal**
- Son **fáciles** de crear
- Permiten comprar muchas variables

	Provincia	Población 2001	Población 2010	Variación absoluta	Variación relativa (%)
0	Ciudad de Buenos Aires	2.776.138	2.890.151	114.013	4,1
1	Buenos Aires	13.827.203	15.625.084	1.797.881	13,0
2	Catamarca	334.568	367.828	33.260	9,9
3	Chaco	984.446	1.055.259	70.813	7,2
4	Chubut	413.237	509.108	95.871	23,2
5	Córdoba	3.066.801	3.308.876	242.075	7,9
6	Corrientes	930.991	992.595	61.604	6,6
7	Entre Ríos	1.158.147	1.235.994	77.847	6,7
8	Formosa	486.559	530.162	43.603	9,0
9	Jujuy	611.888	673.307	61.419	10,0

Tablas

Las tablas en general no son buenas **resaltando patrones**, pero con un poco de formato se vuelven más **legibles** (y cautivantes)

Ticker	Name	Value	Change
	Dow Jones	15,988.08	↓ -2.39%
	S&P 500	1,880.33	↓ -2.16%
	Technology		↑ 2.10%
IBM	IBM	130.00	↓ -2.19%
AAPL	Apple	97.05	↓ -2.48%
MSFT	Microsoft	50.99	↓ -3.99%

<https://www.r-bloggers.com/formatting-table-output-in-r/>

Tablas

Table 1. Grading rubric for writing assignments. Lists categories, evaluation criteria, and point values for each criterion.

Possible grade	Length	Topic	Argument	Mechanics	Citations
A	The paper meets the page length requirement and is formatted correctly. 10 points	Topic fits the scope of the project, makes a clear argument. 20 points	Project includes in-depth discussion and elaboration in all sections. 20 points	No spelling and/or grammar mistakes. 5 points	Cites all information from out of class discussion sources. APA citation style is used in both text and bibliography. 10 points
B	The paper meets the length requirement but has inconsistent citation formatting. 8.5 points	The paper is focused but does not make a clear argument. 17 points	Project includes in-depth discussion and elaboration in most sections. 17 points	Minimal spelling and/or grammar mistakes. 4.25 points	Cites most information obtained from other sources. 8.5 points
C	The paper is up to 1 page too short or too long or is incorrectly formatted. 7.5 points	Topic is either too broad or too narrow. 15 points	Project has omissions of content or content runs-on excessively. Paper relies heavily on quotations for content. 15 points	Several spelling and grammar mistakes. 3.75 points	Cites some information from other sources. Citation style is either inconsistent or incorrect. 7.5 points
D	The paper is more than 1 page longer or shorter than assigned. 6.5 points	Paper does not stay on topic. 13 points	Project has cursory discussion in all the sections of the paper or brief discussions in only a few sections. 13 points	Many spelling and grammar mistakes that make the paper hard to understand. 3.25 points	Does not cite sources. 6.5 points

Gráficos de barra

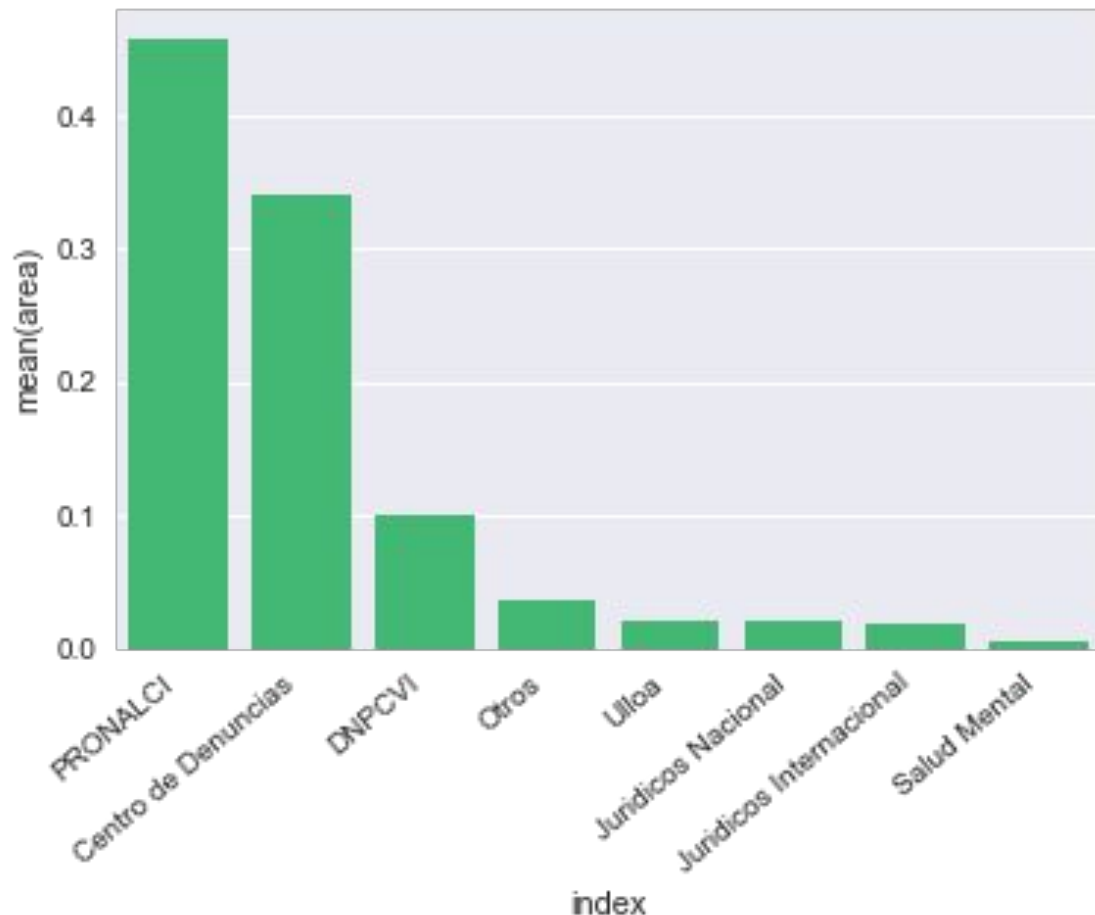
Comparan cantidades

numéricas entre variables

categorías

Son uno de los encodings más

fieles y fáciles de percibir



Gráficos de barra

Permiten comparar cantidades
en **grupos**

Grouped vs stacked

Stacked at 100%

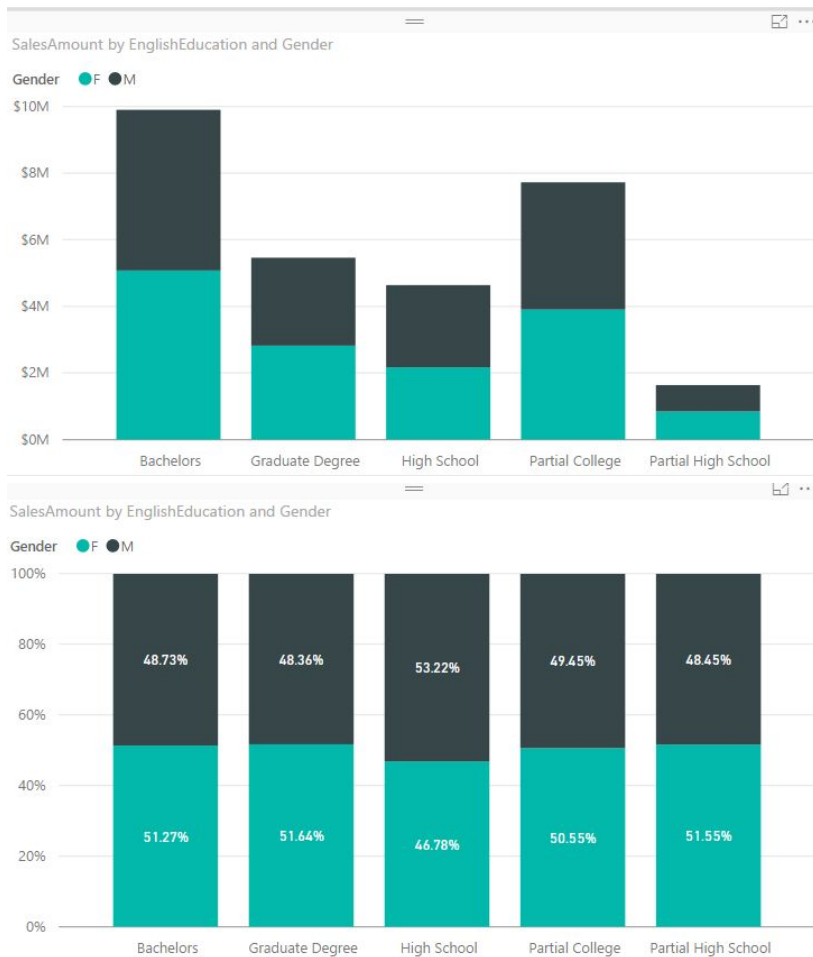


Gráficos de barra

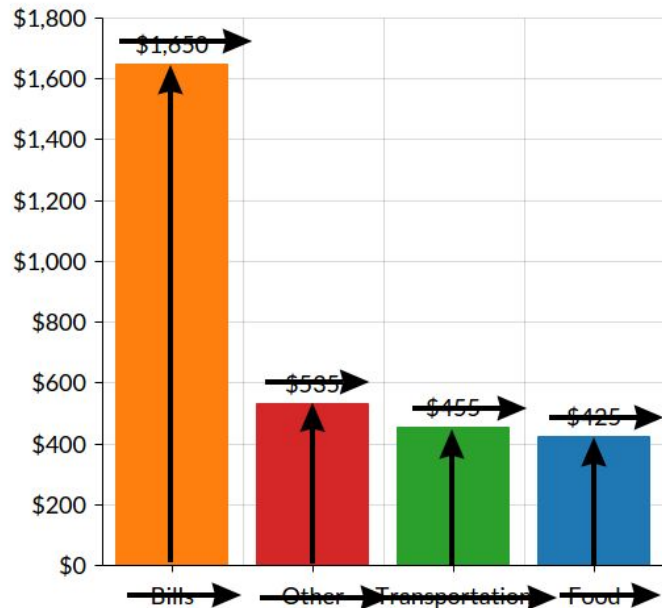
Permiten comparar cantidades
en **grupos**

Grouped vs stacked

Stacked at 100%



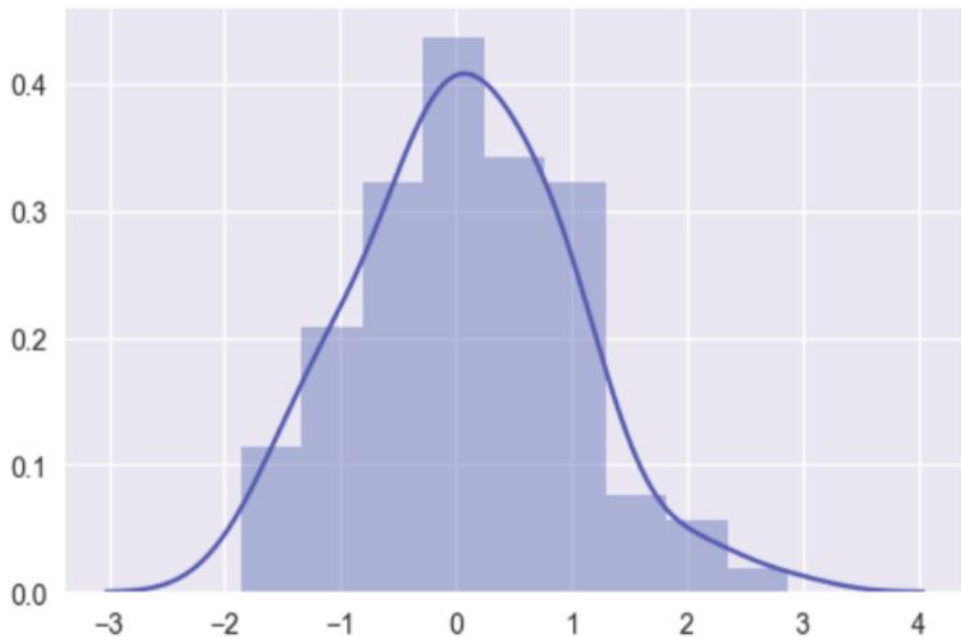
Gráficos de barra



Plotting univariate distributions

The most convenient way to take a quick look at a univariate distribution in seaborn is the `distplot()`

```
x = np.random.normal(size=100)
sns.distplot(x);
```

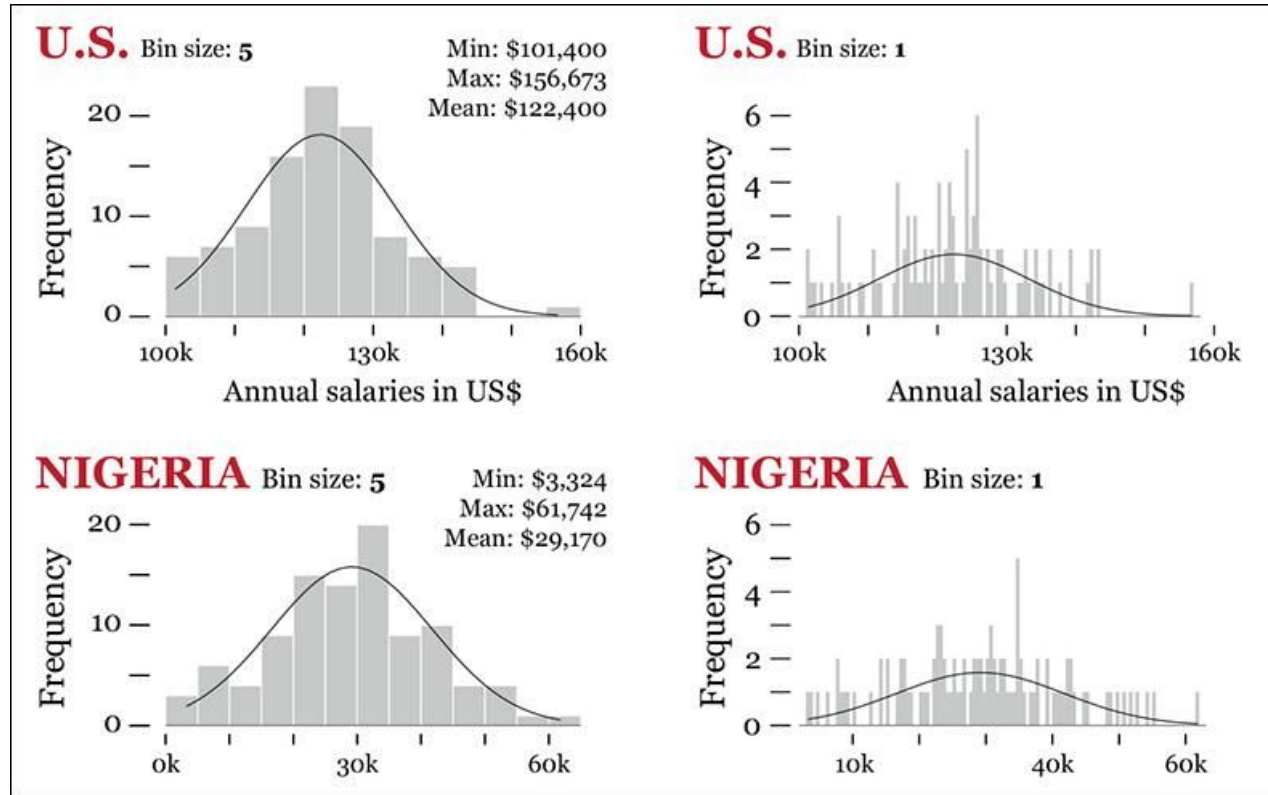


Histograma

- ¡No es lo mismo que un gráfico de barras!
- Muestra la **distribución** de **una** variable **numérica**
- Divide los datos en **bins**

Histograma

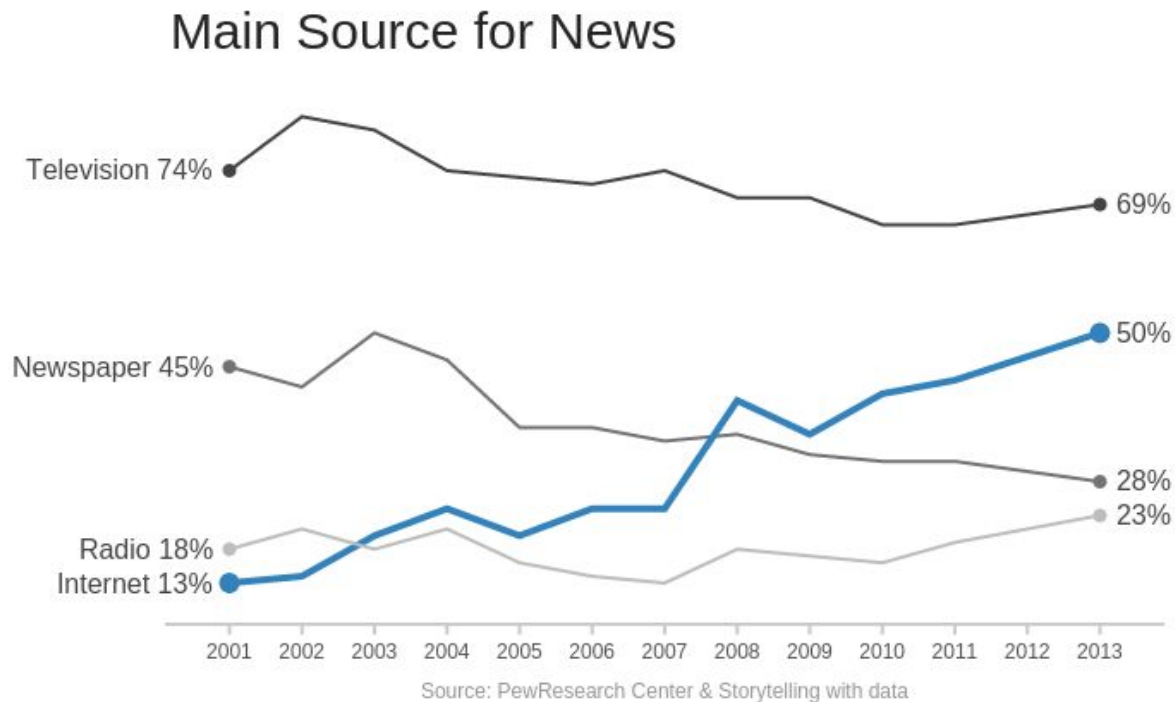
La **binarización** es muy importante en el histograma resultante.



Gráficos de línea

Cada línea representa la
variación de dos **variables**
numéricas

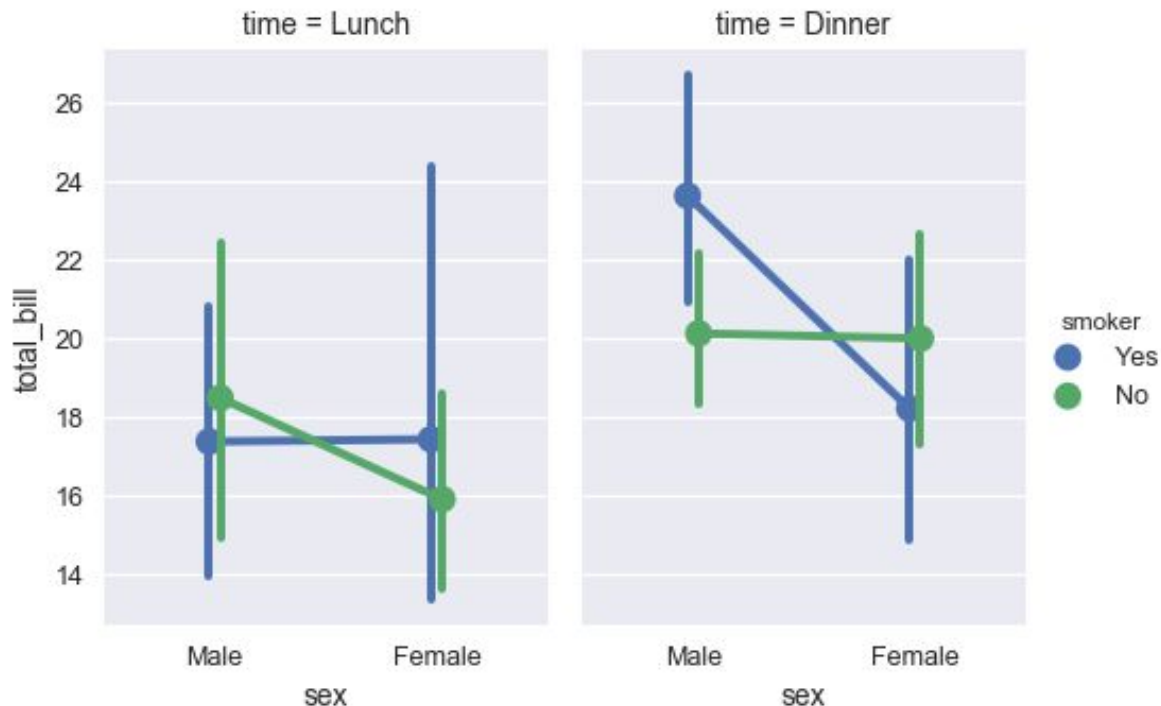
Múltiples líneas permiten
comparar distintas categorías



Gráficos de línea

Son muy versátiles

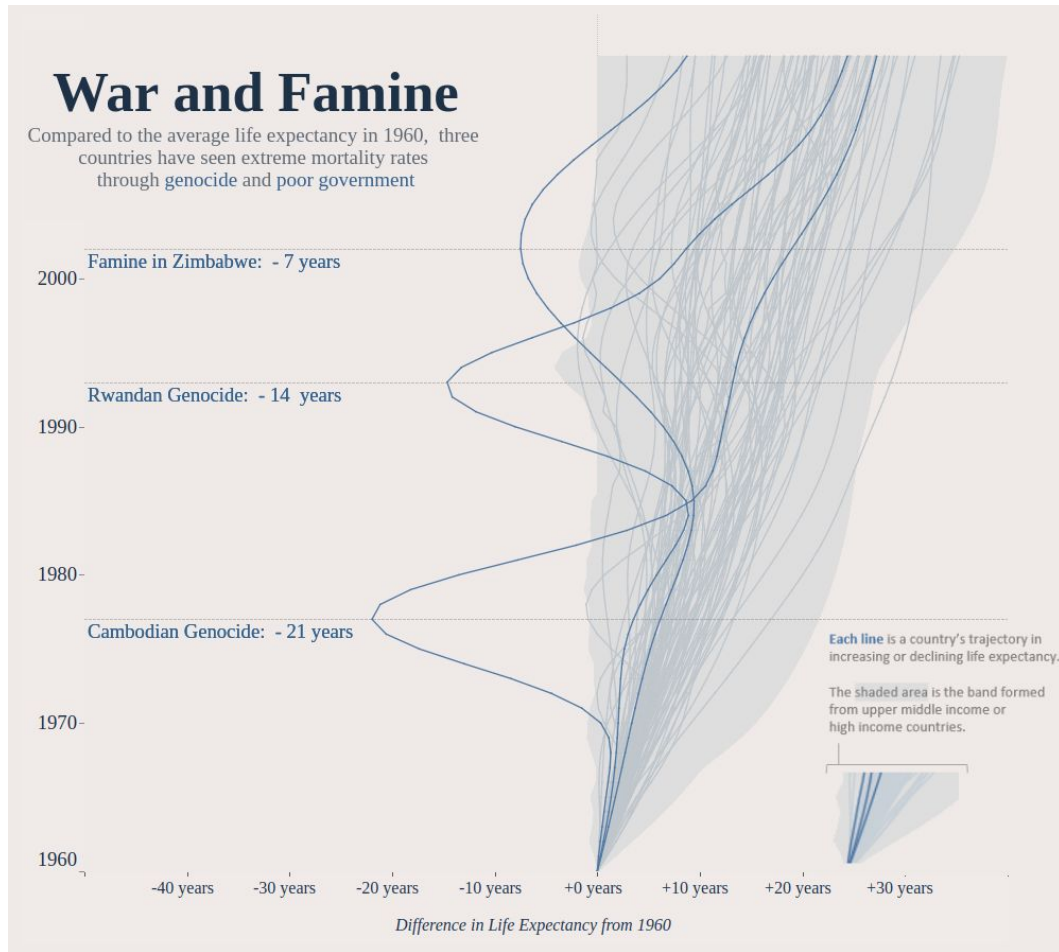
Generan gráficos visualmente
simples, por lo que pueden
contener mucha información



Gráficos de línea

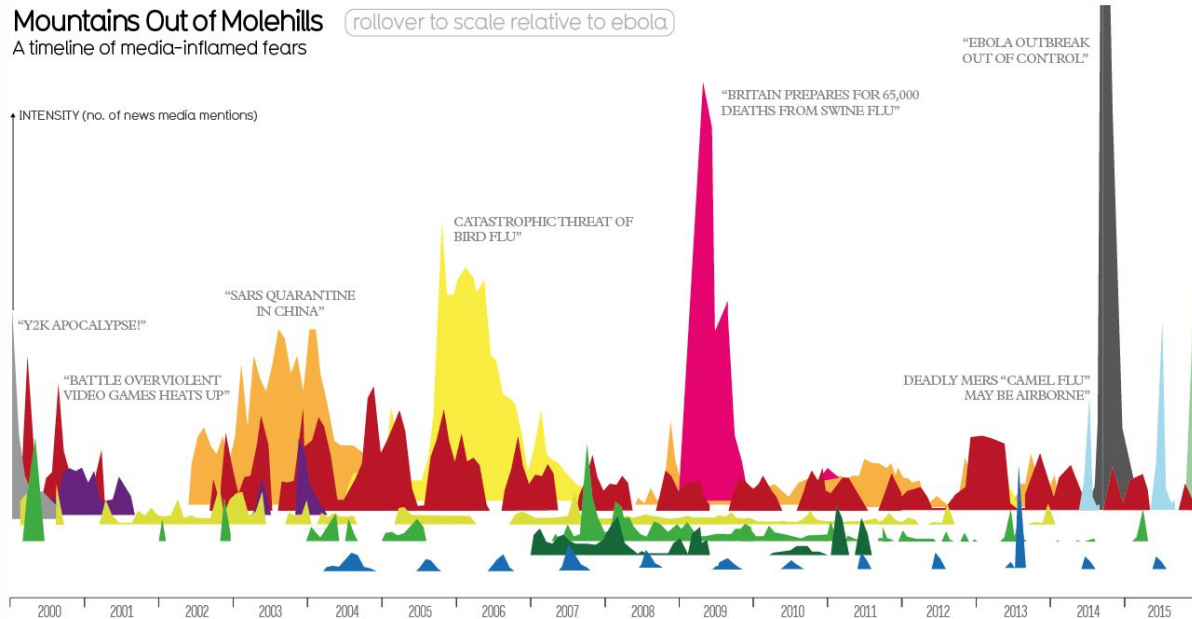
Son muy versátiles

Generan gráficos visualmente
simples, por lo que pueden
contener mucha información



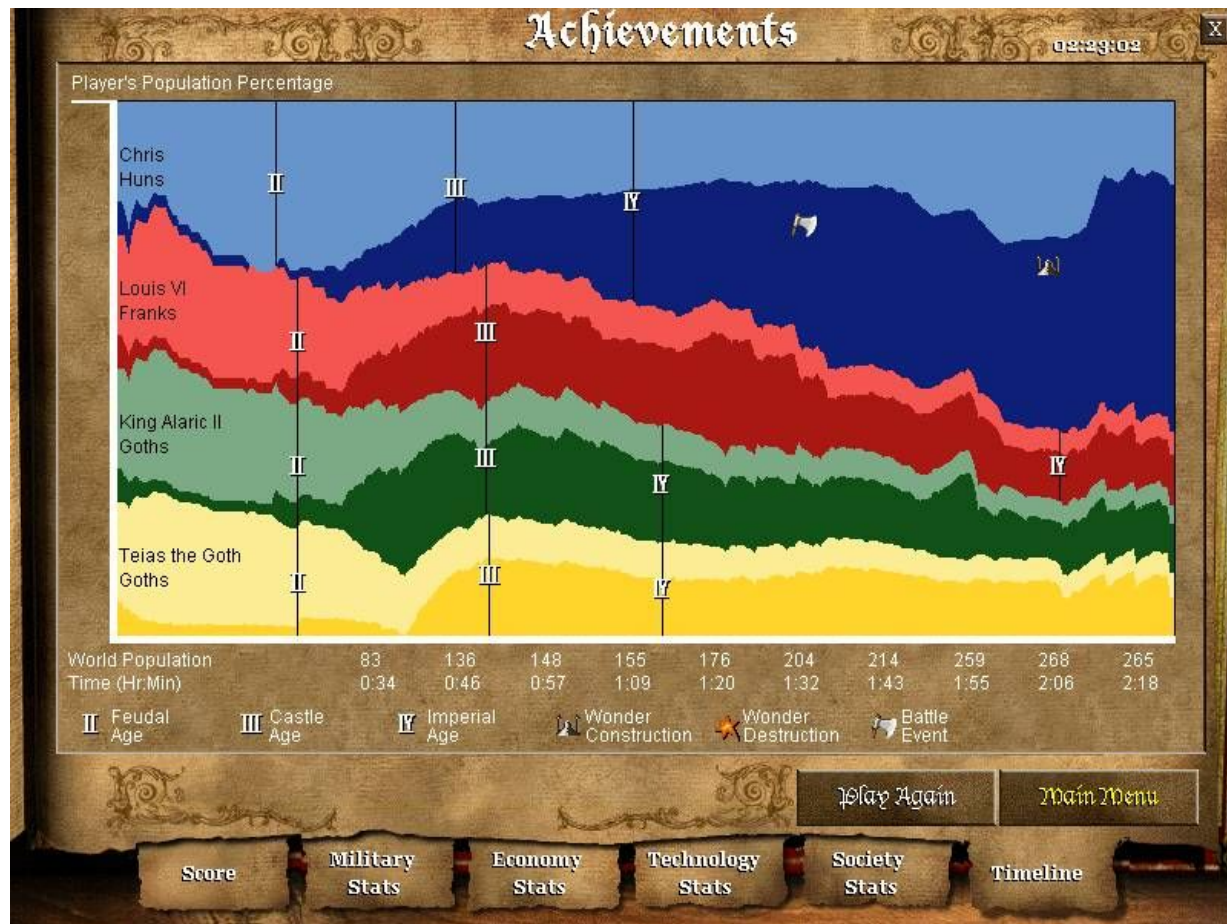
Gráficos de área

Igual a los gráficos de línea,
pero tienen más **impacto**
visual.



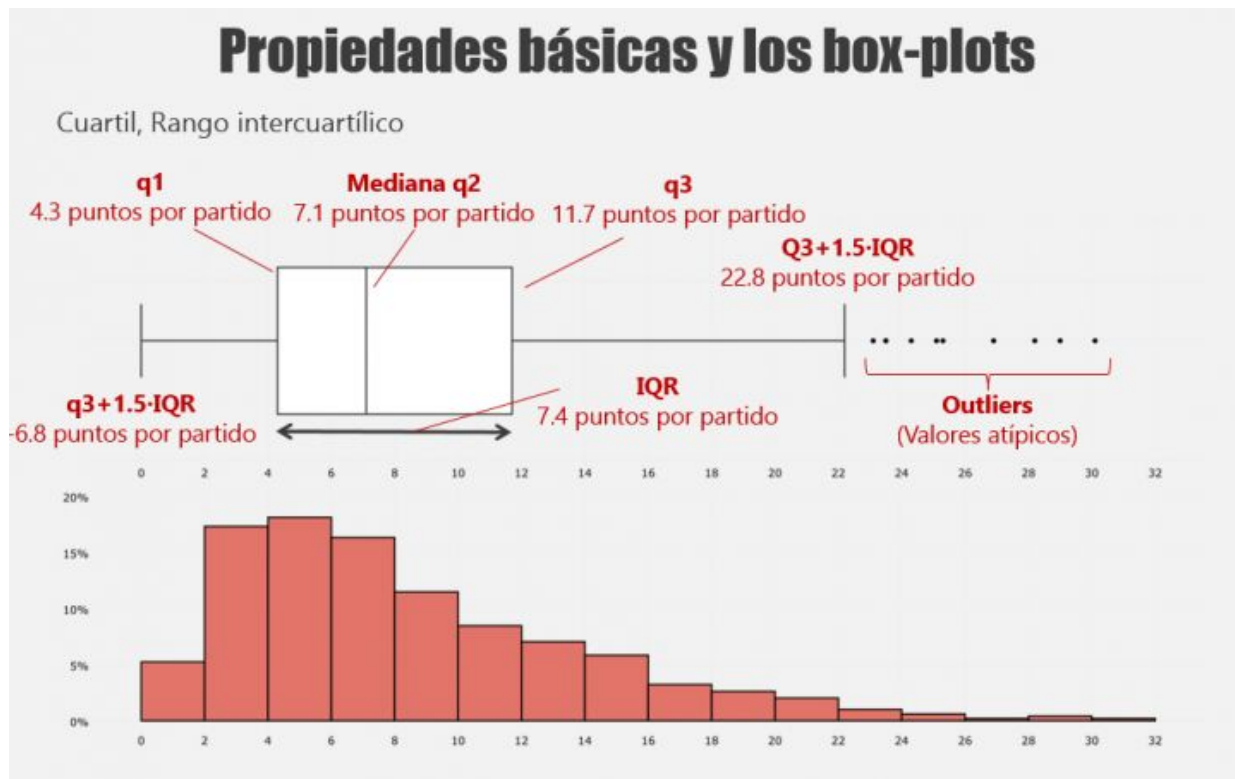
Gráficos de área

Pueden ser **apilados**, tomando propiedades similares a los gráficos de barra apilados

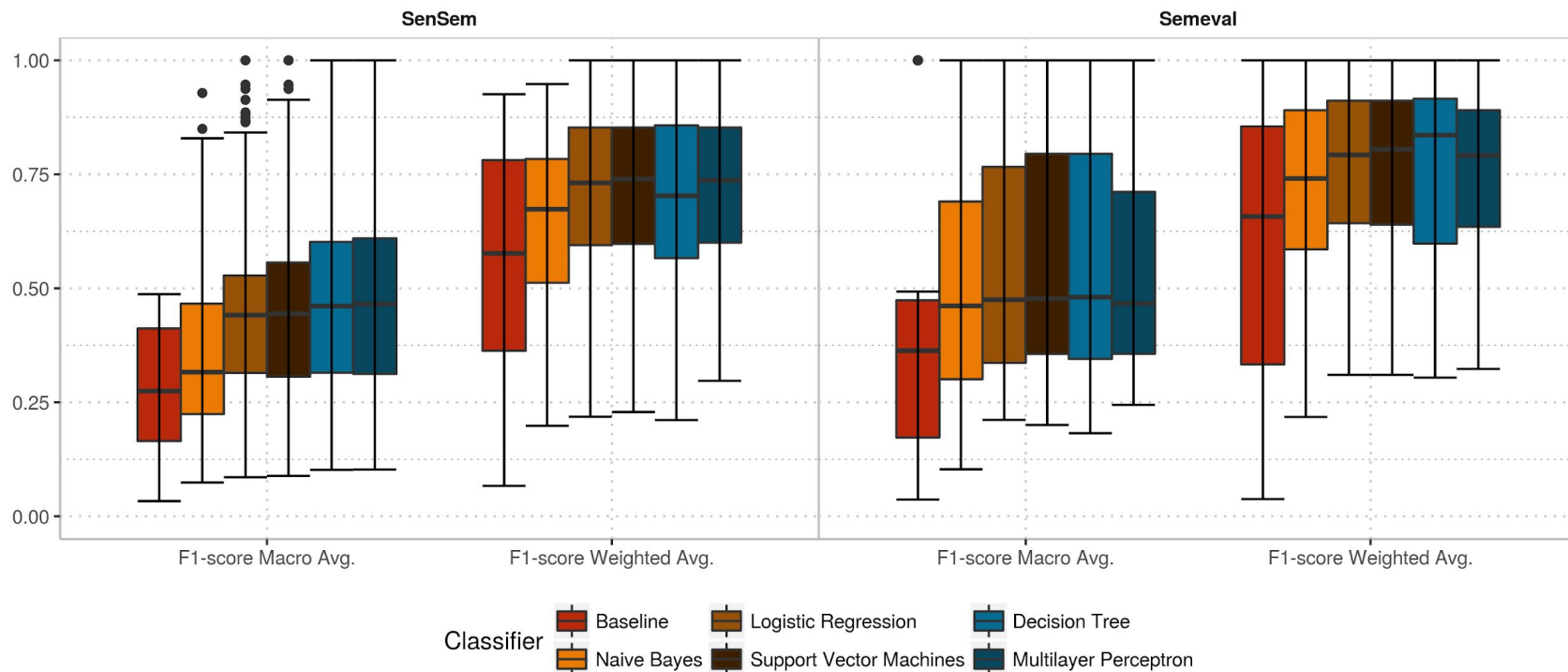


Gráficos de cajas

- Muestra la **distribución** de una **variable numérica continua**.
- Muestra información de forma más **condensada** que un histograma.

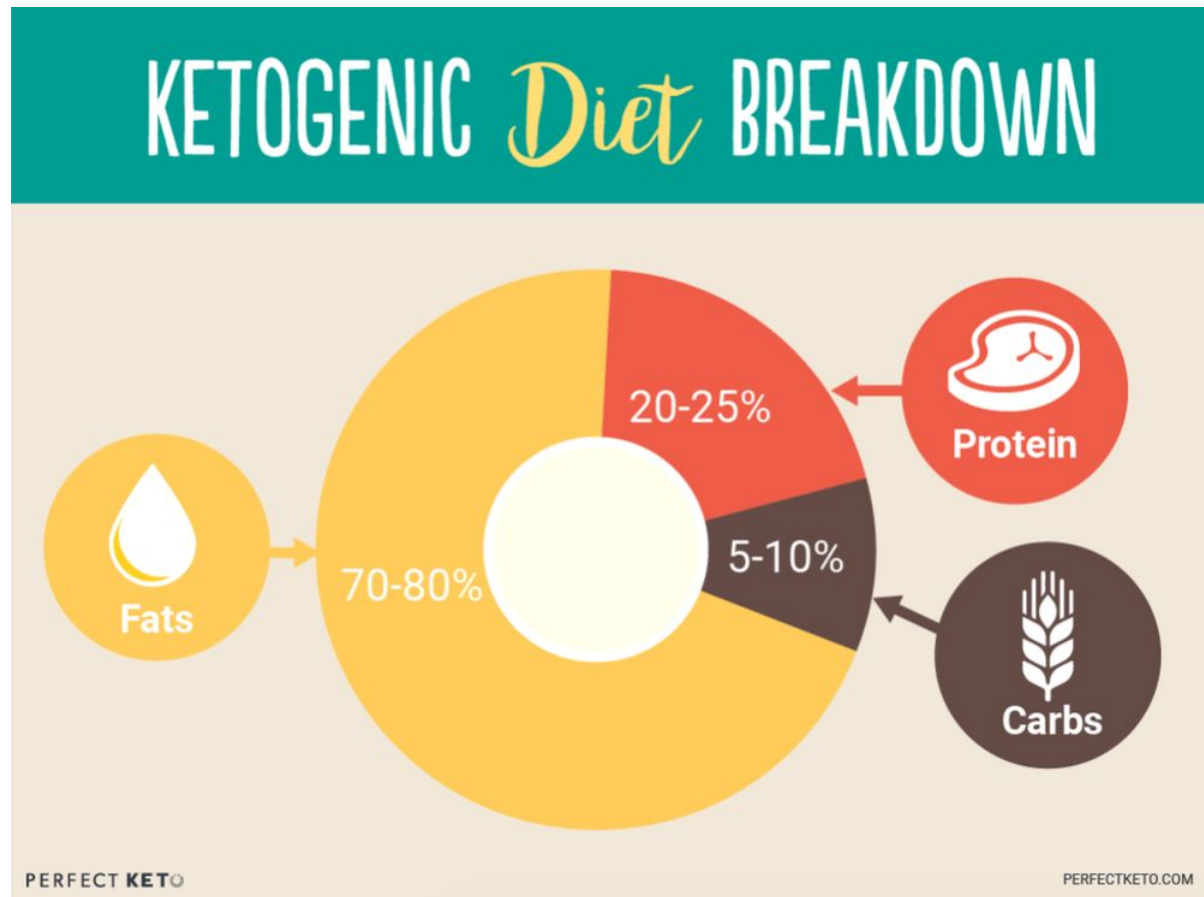


Gráficos de cajas



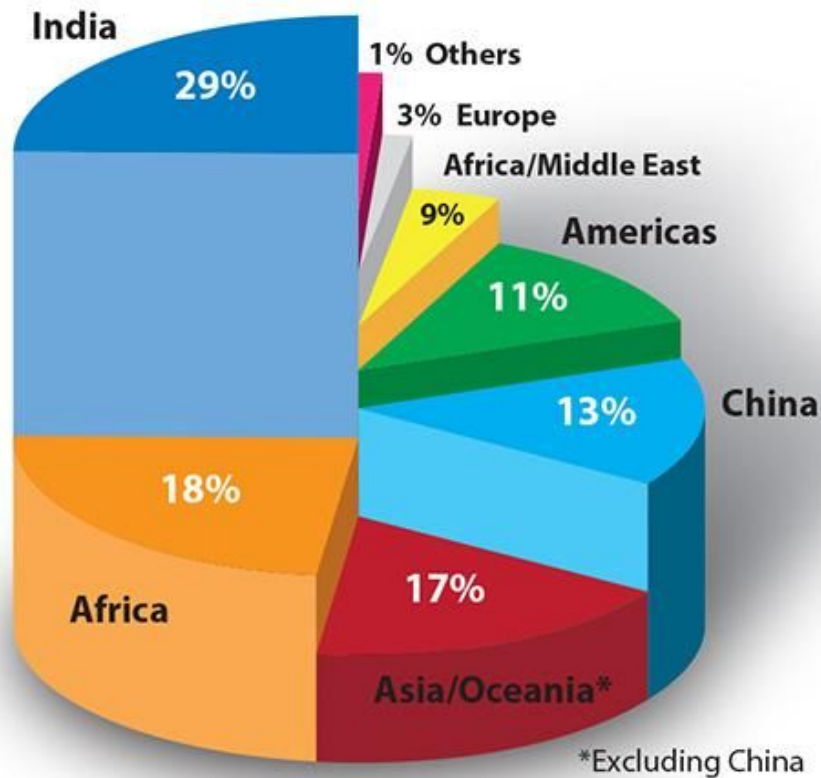
Gráficos de torta

Ilustra la distribución de la población con respecto a una **variable categórica**.



The horror of pie charts

Share of worldwide urban population growth 2010-2050

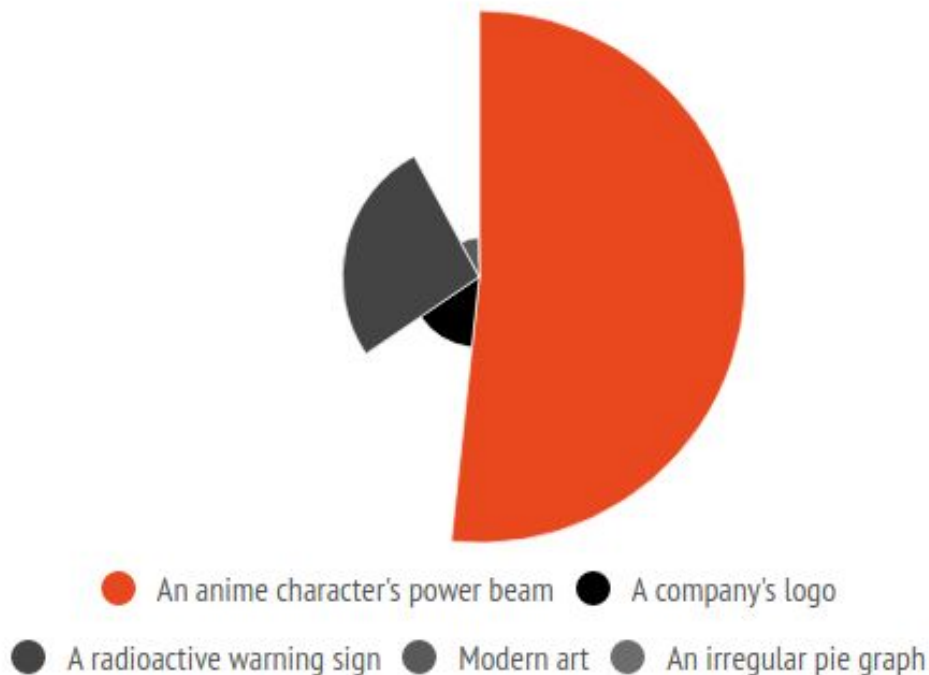


Son tan malos, que seaborn
no los tiene!

Gráficos de torta

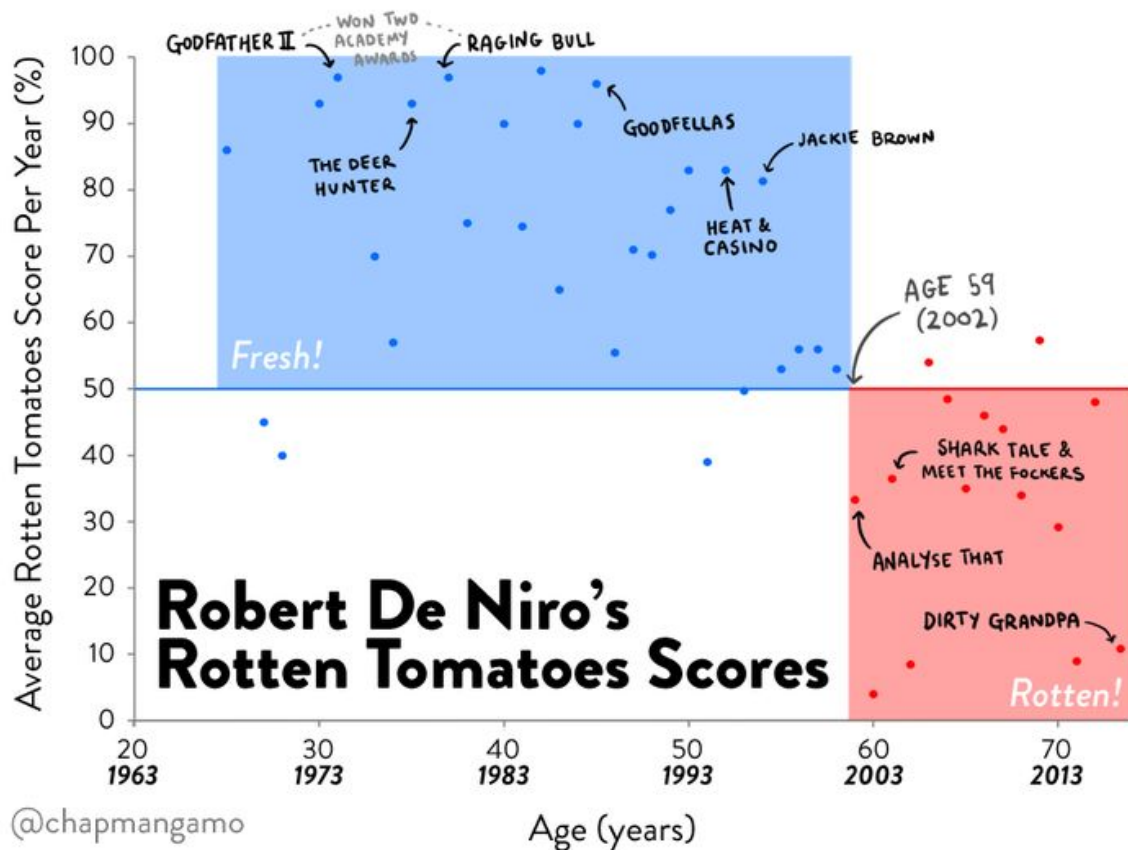
What people think this chart type is

Elementos visuales adicionales
facilitan la comparación entre
elementos del gráfico



Gráficos de puntos

- Muestra la relación entre 2 o 3 **variables numéricas continuas**
- Puede usar color, forma de los puntos para variables categóricas, y tamaño para variables numéricas

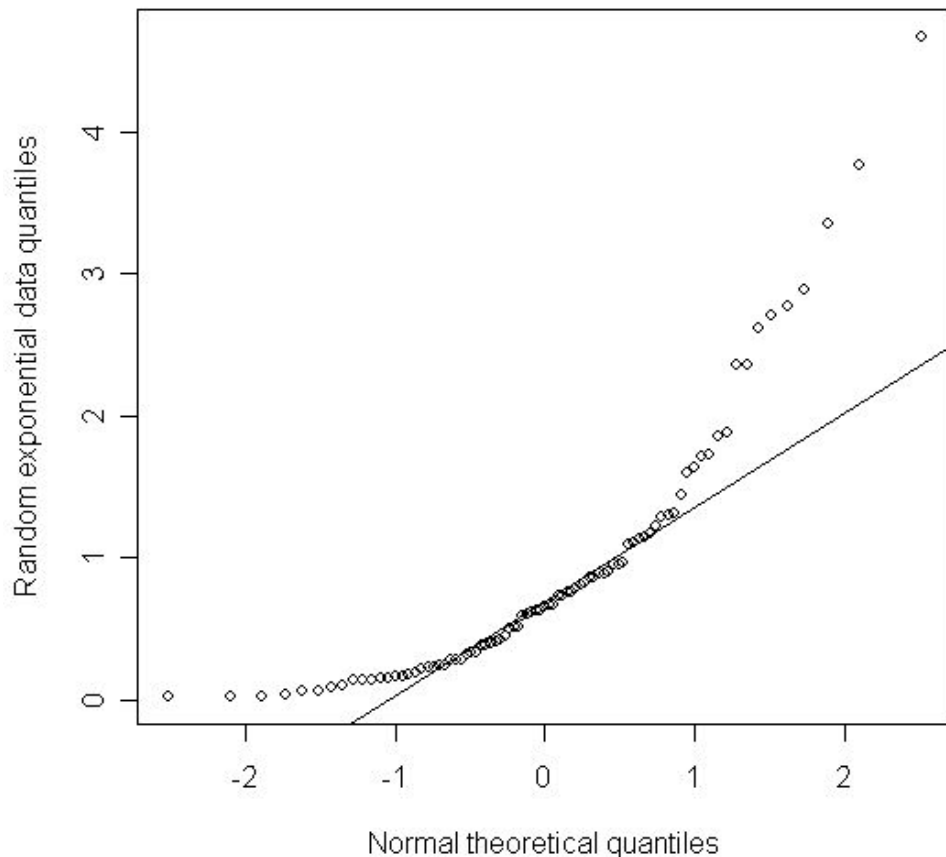


[Scatter plot shows exactly when Robert de Niro stopped making good films](#)

Gráficos QQ

- Compara los cuartiles de dos muestras
- Sirve para ver que tan parecidas son las dos distribuciones de las que provienen las muestras

Normal Q-Q Plot with exponential data



Tutoriales de Seaborn

- Visualizar datos categóricos:
<https://seaborn.pydata.org/tutorial/categorical.html>
- Visualizar datos lineales
<https://seaborn.pydata.org/tutorial/relational.html#relational-tutorial>
- Encontrar relaciones entre variables
<https://seaborn.pydata.org/tutorial/regression.html>

¿preguntas?