



# Diplomado para Acceder a Grado

## Módulo Ciencia de Datos

### Visualización en Ciencia de Datos

**Dr. José Ramón Iglesias**

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingeniería Electronica

Universidad Popular del Cesar

Herramienta para la comunicación

seaborn.boxplot?

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

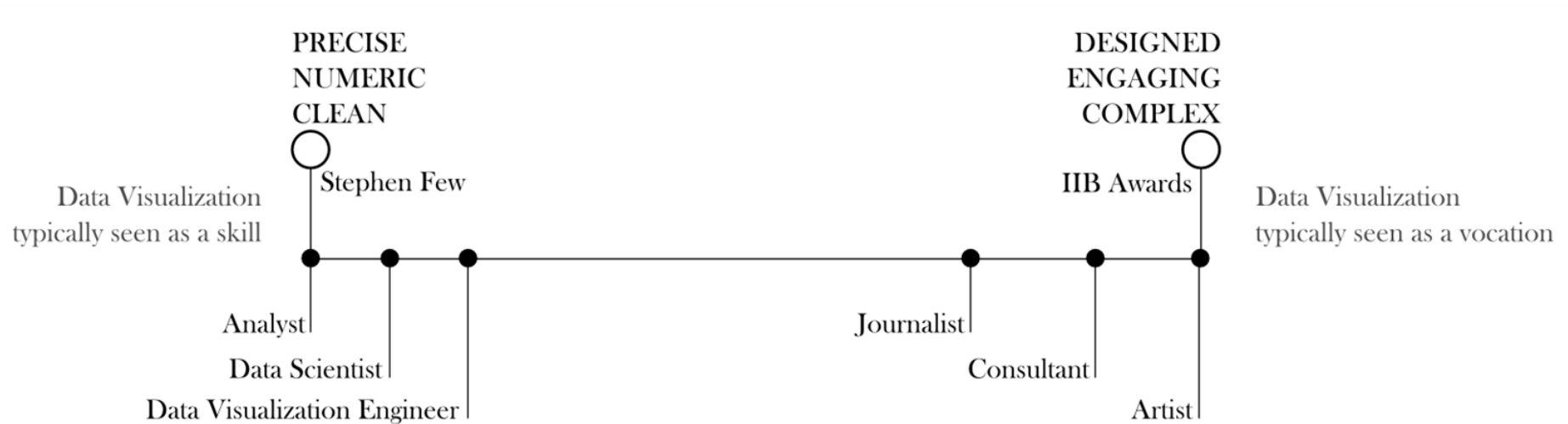
Herramienta para la compresión

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL		
1							Optimizer			Config																														
2	Results	Log	Trucar	Diff	Optim	LR	Cell	batch	dropc	epoch	embedc	hiddr	max_s	Filter	Merge	Pretra	Finet	AUC R	rmse			AUC R	rmse	Accurac	R2	AUC R	rmse	Accurac	R2	AUC R	rmse	Accurac	R2	AUC R	rmse	Accurac	R2	Eval all		
3																																								
4	LSTM																																							
5	../results/kddcup/lstm/pre	14648	N					100	0.3	500	-		100	50	> 5	Y	-	-					0.787	0.388	0.795	0.278														
6	/home/mteruel/edm/resul	14662	N					100	0.3	500	-		100	50	> 5	Y	-	-					0.879	0.359	0.831	0.439	0.794	0.345	0.850	0.275	0.605	0.533	0.672	-0.308	0.498	0.456	0.770	-0.431		
7	/home/mteruel/edm/resul	14663	N					100	0.3	500	-		50	50	> 5	Y	-	-					0.880	0.359	0.827	0.438	0.802	0.339	0.858	0.296	0.656	0.511	0.685	-0.183	0.597	0.414	0.800	-0.199		
8	/home/mteruel/edm/resul	14664	N					100	0.3	500	-		50	100	> 5	Y	-	-		0.814	0.374		0.881	0.361	0.822	0.431	0.804	0.336	0.864	0.315	0.666	0.490	0.713	-0.108	0.626	0.405	0.795	-0.094		
9	../results/kddcup/lstm/pre	14735	Y					50	0.3	500	-		50	50	> 5	Y	-	-					0.759	0.467	0.617	0.050	0.657	0.415	0.784	-0.053	0.611	0.500	0.613	-0.164	0.634	0.398	0.783	-0.074		
10	../results/kddcup/lstm/pre	14885	N					50	0.3	500	-		50	50	N	Y	-	-		0.837	0.335		0.871	0.341	0.855	0.450	0.842	0.290	0.895	0.369	0.741	0.386	0.796	0.153	0.662	0.418	0.771	-0.195	14939	
11		N						100	0.3	500	-		50	100	N	Y	-	-					0.871	0.339	0.851	0.458	0.837	0.289	0.896	0.375	0.748	0.390	0.800	0.138	0.633	0.416	0.797	-0.181	14937	
12																																								
13	Embeddings																																							
19	/home/mteruel/edm/resul	14661	N					100	0.3	500		50	50	100	> 5	N	N						0.885	0.352	0.836	0.457	0.814	0.335	0.856	0.326	0.703	0.466	0.723	0.006	0.724	0.393	0.797	-0.047		
20	/home/mteruel/edm/resul	14667	N					100	0.3	500		50	50	50	> 5	N	N						0.880	0.359	0.829	0.439	0.813	0.332	0.867	0.330	0.687	0.478	0.705	-0.053	0.668	0.405	0.810	-0.142		
21	../results/kddcup/embeddc	14706	Y					50	0.2	500		50	50	20	> 5	N	Y	Y					0.728	0.481	0.614	-0.010	0.683	0.390	0.798	0.069	0.641	0.467	0.687	0.004	0.599	0.404	0.773	-0.113		
22	../results/kddcup/embeddc	14716	Y					100	0.3	500		50	50	50	> 5	Y	N						0.756	0.459	0.645	0.079	0.674	0.430	0.746	-0.153	0.649	0.462	0.673	0.036	0.534	0.422	0.770	-0.207		
23	../results/kddcup/embeddc	14821	N					100	0.3	500		50	50	100	> 5	N	N			0.830	0.363		0.884	0.357	0.830	0.443	0.810	0.334	0.868	0.319	0.740	0.455	0.727	0.050	0.685	0.379	0.815	0.029		
24	../results/kddcup/embeddc	14823	N					100	0.3	500		50	50	100	> 5	Y	N			0.834	0.361		0.884	0.358	0.825	0.442	0.813	0.333	0.865	0.322	0.731	0.456	0.715	0.043	0.688	0.375	0.831	0.051		
25	../results/kddcup/embeddc	14828	N					50	0.2	500		50	100	20	> 5	Y	Y	Y		0.808	0.378		0.871	0.365	0.816	0.417	0.801	0.344	0.846	0.279	0.701	0.479	0.711	-0.053	0.575	0.407	0.813	-0.122		
26	../results/kddcup/embeddc	14832	N					50	0.3	500		50	100	200	> 5	Y	Y	Y		0.818	0.370		0.879	0.359	0.825	0.436	0.788	0.340	0.861	0.296	0.707	0.471	0.711	-0.018	0.672	0.388	0.811	-0.021		
27	../results/kddcup/embeddc	14858	N					50	0.3	500		50	100	200	> 5	Y	Y	N		0.825	0.365		0.875	0.365	0.827	0.420	0.806	0.338	0.854	0.302	0.714	0.449	0.739	0.076	0.650	0.383	0.824	0.007		
28	../results/kddcup/embeddc	14873	N					50	0.3	500		20	100	200	> 5	Y	Y	Y		0.831	0.363		0.879	0.361	0.827	0.433	0.815	0.335	0.857	0.316	0.733	0.438	0.742	0.122	0.675	0.381	0.826	0.020		
29	../results/kddcup/embeddc	14875	N					50	0.3	500		20	100	200	> 5	Y	Y	N		0.835	0.362		0.880	0.361	0.822	0.432	0.815	0.340	0.846	0.293	0.722	0.445	0.736	0.089	0.712	0.371	0.823	0.069		
30	../results/kddcup/embeddc	14877	N					50	0.3	500		20	50	200	> 5	Y	Y	N		0.841	0.360		0.880	0.364	0.818	0.423	0.819	0.334	0.859	0.320	0.753	0.432	0.735	0.145	0.715	0.372	0.826	0.065	14941	
31	../results/kddcup/embeddc	14886	N					100	0.3	500		50	50	100	N	Y	N			0.850	0.330		0.887	0.338	0.853	0.461	0.850	0.291	0.895	0.366	0.783	0.379	0.807	0.184	0.699	0.396	0.813	-0.069	14922 +14	
32	../results/kddcup/embeddc	14887	N					50	0.3	500		20	100	200	N	Y	Y	Y		0.846	0.328		0.879	0.339	0.852	0.458	0.843	0.289	0.896	0.373	0.788	0.371	0.820	0.219	0.652	0.399	0.804	-0.089 x		
33		N						50	0.3	500		20	50	200	N	Y	Y	N					0.881	0.339	0.851	0.456	0.843	0.290	0.892	0.366	0.804	0.362	0.831	0.258	0.740	0.379	0.814	0.020	14940	
41	../results/kddcup/embeddc	15167	N				adam	0.01	gru		100	0.3	500	20	50	200	N	Y	Y	Y		0.818	0.343	0.890	0.334	0.854	0.472	0.839	0.294	0.896	0.349	0.732	0.391	0.807	0.130	0.576	0.437	0.787	-0.305	
42	../results/kddcup/embeddc	15177	N				adam	??	gru		100	0.3	500	20	50	200	N	Y	Y	Y		0.811	0.345	0.886	0.336	0.858	0.466	0.823	0.304	0.890	0.307	0.683	0.438	0.773	-0.088	0.627	0.419	0.801	-0.199	
43	../results/kddcup/embeddc	15235	N				adam	0.01	lstm		100	0.3	500	50	50	100	N	Y	N			0.811	0.347	0.882	0.345	0.841	0.439	0.814	0.304	0.892	0.305	0.684	0.425	0.799	-0.026	0.621	0.411	0.810	-0.153	
44	../results/kddcup/embeddc	15236	N				adam	0.01	lstm		50	0.3	500	20	100	200	N	Y	Y	Y		0.812	0.345	0.884	0.340	0.850	0.452	0.826	0.300	0.892	0.325	0.700	0.405	0.800	0.071	0.589	0.431	0.786	-0.265	
45	../results/kddcup/embeddc	26285	N					100	0.3	500		20	20	300	N	Y	N			0.853	0.325		0.881	0.335	0.854	0.469	0.841	0.291	0.895	0.362	0.790	0.369	0.814	0.228	0.675	0.397	0.796	-0.076		
46	../results/kddcup/embeddc	26286	N					100	0.3	500		20	20	300	N	Y	Y	Y		0.857	0.322		0.883	0.334	0.857	0.474	0.845	0.288	0.895	0.375	0.783	0.364	0.826	0.250	0.757	0.371	0.820	0.062		

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL							
1							Optimizer																																						
2	Results	Log	Trucar	Diff	Optim	LR	Cell	batch	dropc	epoch	embed	hider	max_s	Filter	Merge	Pretra	Finet	AUC	R	rmse																									
3																																													
4	LSTM																																												
5	../results/kddcup/lstm/pre	14648	N					100	0.3	500	-		100	50	> 5	Y	-	-																											
6	/home/mteruel/edm/resul	14662	N					100	0.3	500	-		100	50	> 5	Y	-	-																											
7	/home/mteruel/edm/resul	14663	N					100	0.3	500	-		50	50	> 5	Y	-	-																											
8	/home/mteruel/edm/resul	14664	N					100	0.3	500	-		50	100	> 5	Y	-	-		0.814	0.374																								
9	../results/kddcup/lstm/pre	14735	Y					50	0.3	500	-		50	50	> 5	Y	-	-																											
10	../results/kddcup/lstm/pre	14885	N					50	0.3	500	-		50	50	N	Y	-	-		0.837	0.335																								
11			N					100	0.3	500	-		50	100	N	Y	-	-																											
12																																													
13	Embeddings																																												
19	/home/mteruel/edm/resul	14661	N					100	0.3	500		50	50	100	> 5	N	N																												
20	/home/mteruel/edm/resul	14667	N					100	0.3	500		50	50	50	> 5	N	N																												
21	../results/kddcup/embedc	14706	Y					50	0.2	500		50	50	20	> 5	N	Y	Y																											
22	../results/kddcup/embedc	14716	Y					100	0.3	500		50	50	50	> 5	Y	N																												
23	../results/kddcup/embedc	14821	N																																										
24	../results/kddcup/embedc	14823	N																																										
25	../results/kddcup/embedc	14828	N																																										
26	../results/kddcup/embedc	14832	N																																										
27	../results/kddcup/embedc	14858	N																																										
28	../results/kddcup/embedc	14873	N																																										
29	../results/kddcup/embedc	14875	N																																										
30	../results/kddcup/embedc	14877	N																																										
31	../results/kddcup/embedc	14886	N																																										
32	../results/kddcup/embedc	14887	N																																										
33			N																				</																						

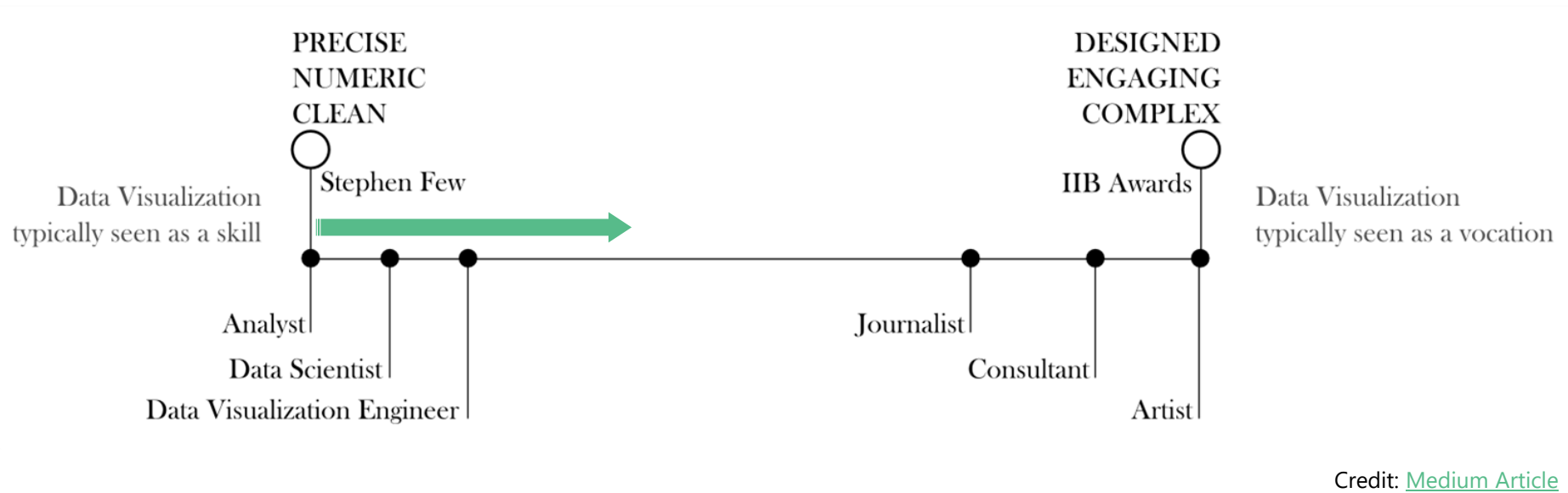


enrollment_id	username	course_id
1	9Uee7oEuuMmgPx2IzPffkHgkHZyPbWr0	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
3	1qXC7Fjbwp66GPQc6pHLfEu08WKozxG4	7GRhBDsirIGkRZBtSMEzNTyDr2JQm4xx
4	FIHlppZyoq8muPbdVxS44gfvceX9zvU7	DPnLzkJJq00PRJfBxIHbQEERiYHu5ila
5	p1Mp7WkVfzUijX0peVQKSHbgd5pXyl4c	7GRhBDsirIGkRZBtSMEzNTyDr2JQm4xx
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 quieres llegar?



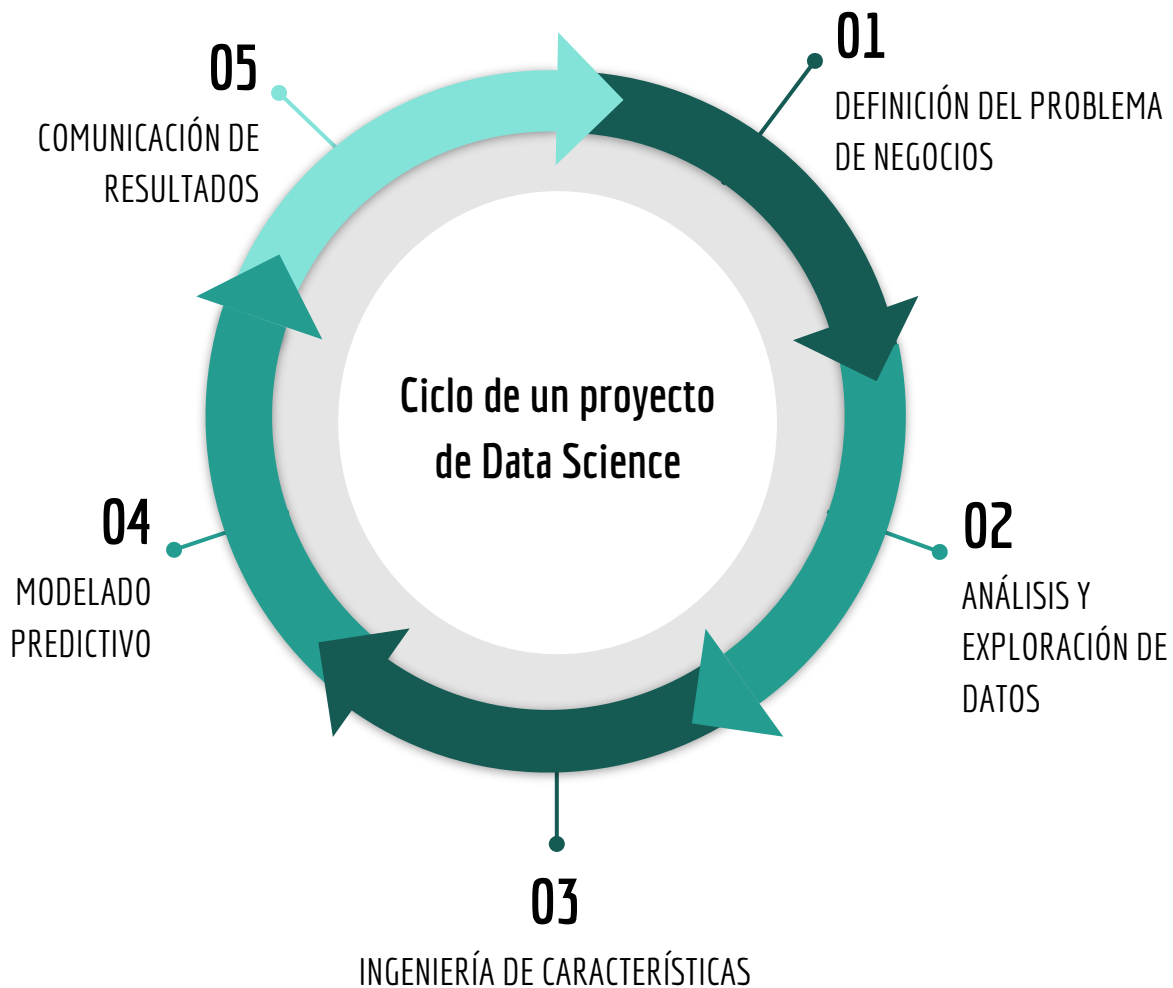
¿Podés identificar tu rol en esta línea? ¿A dónde quieres llegar?

# La realidad

¡Data science es un ciclo!

Durante esta materia, veremos conceptos involucrados en:

1. Herramientas estadísticas y visualizaciones para la etapa 02.
2. Herramientas estadísticas necesarias para interpretar los resultados de la etapa 04.
3. Visualización y comunicación efectiva para la etapa 05.

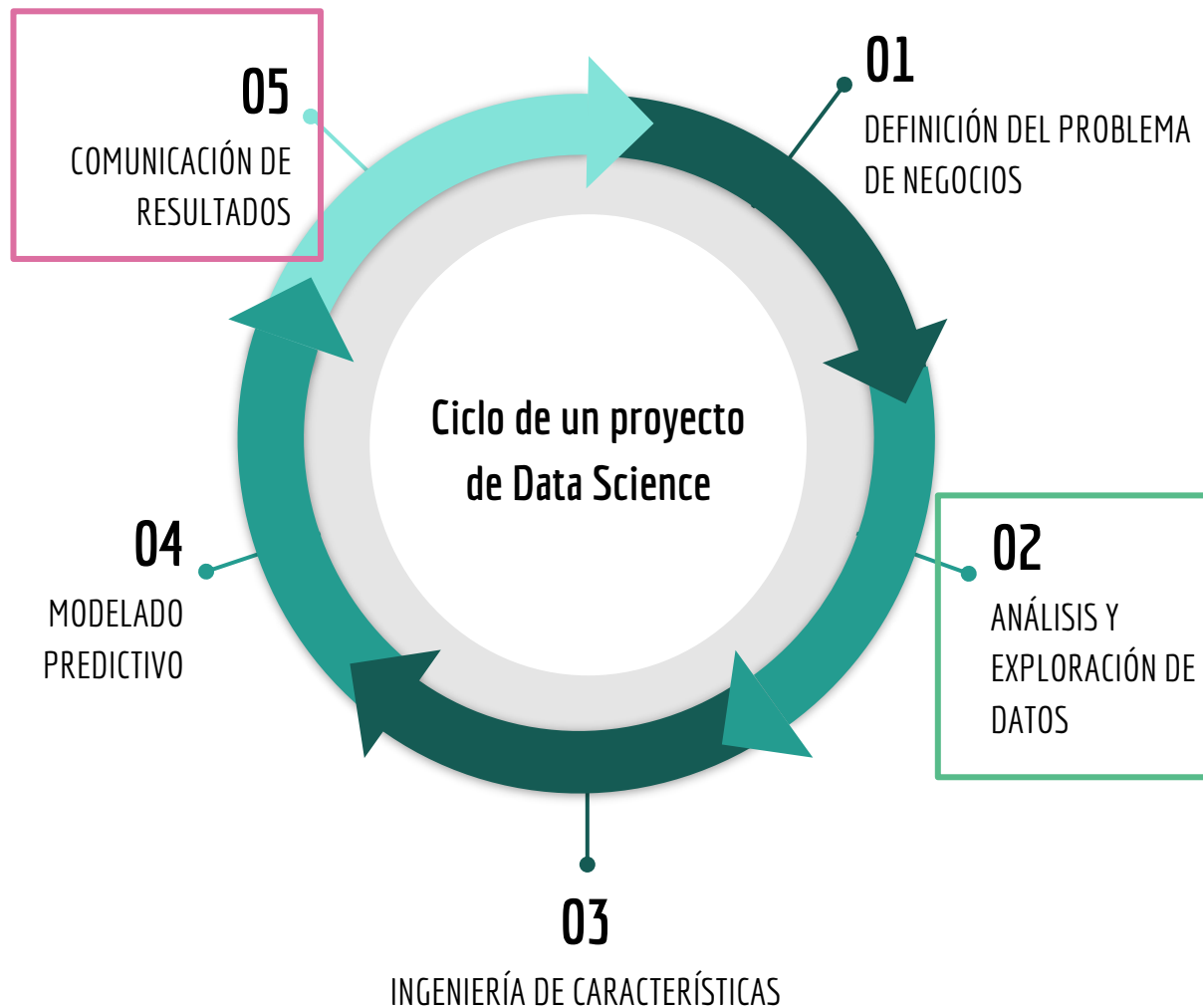


# La realidad

¡Data science es un ciclo!

Durante esta materia, veremos conceptos involucrados en:

1. Herramientas estadísticas y visualizaciones para la etapa 02.
2. Herramientas estadísticas necesarias para interpretar los resultados de la etapa 04.
3. Visualización y comunicación efectiva para la etapa 05.



To calculate is not in itself to analyze

*Edgar Allan Poe, Murders in the Rue Morgue*

# 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 Rios	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/>

# Gráficos de barra

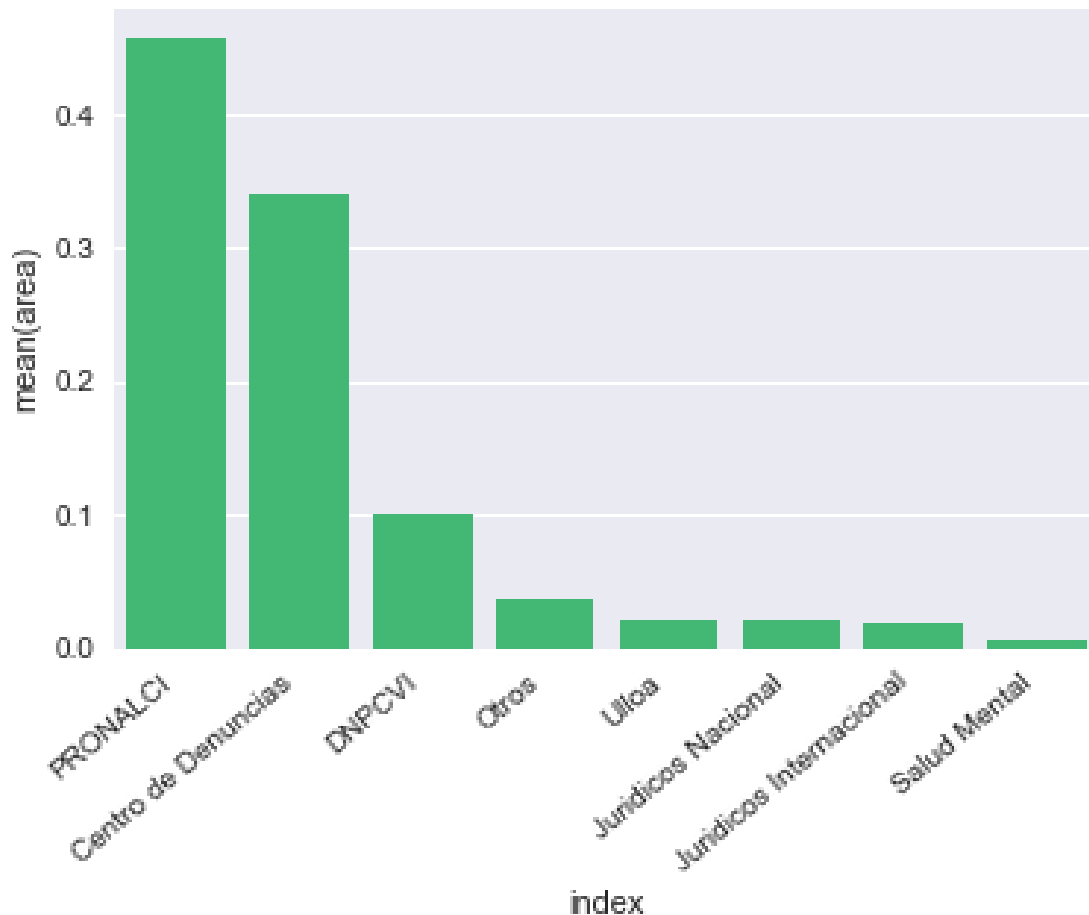
Comparan cantidades

**numéricas** entre variables

**categóricas**

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%

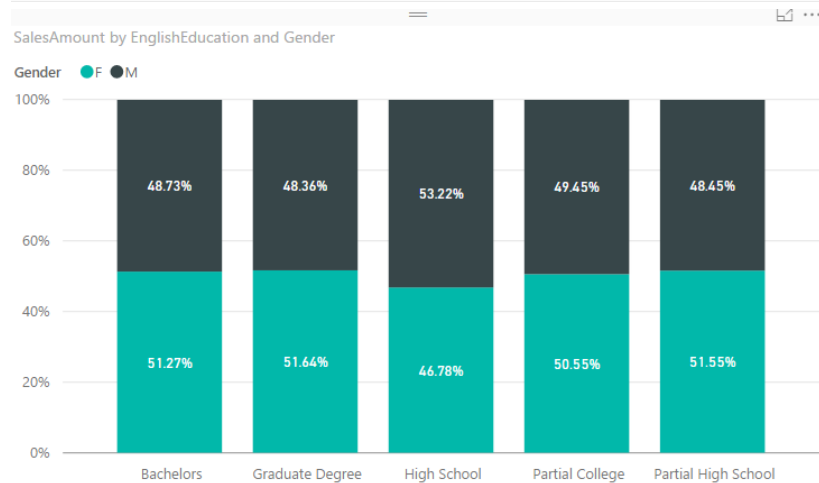
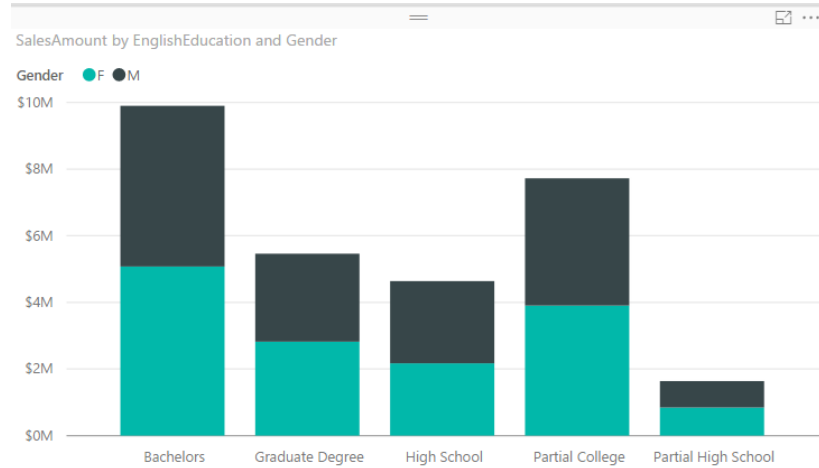


# Gráficos de barra

Permiten comparar cantidades  
en **grupos**

Grouped vs stacked

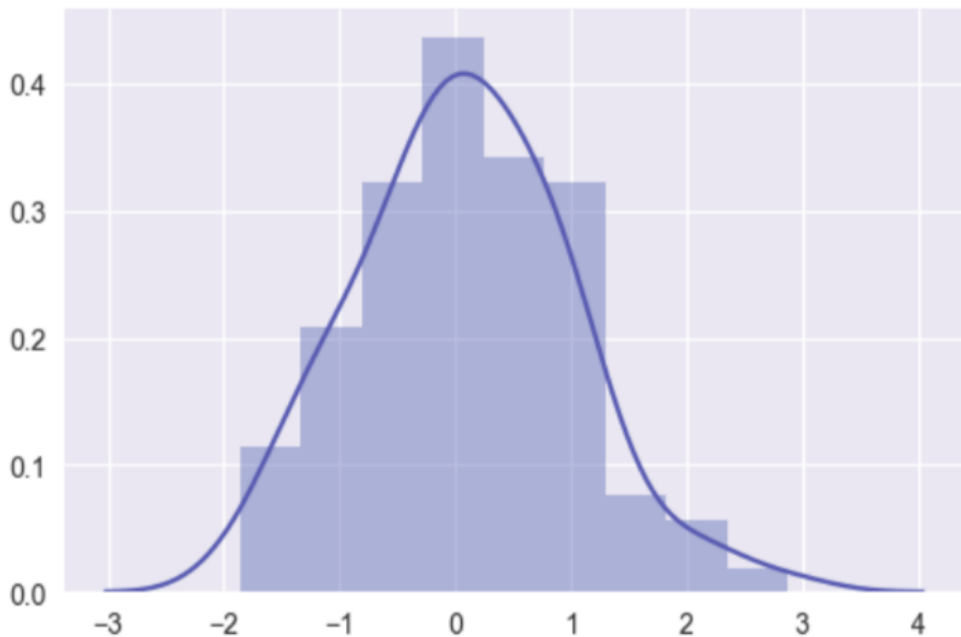
Stacked at 100%



# 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);
```



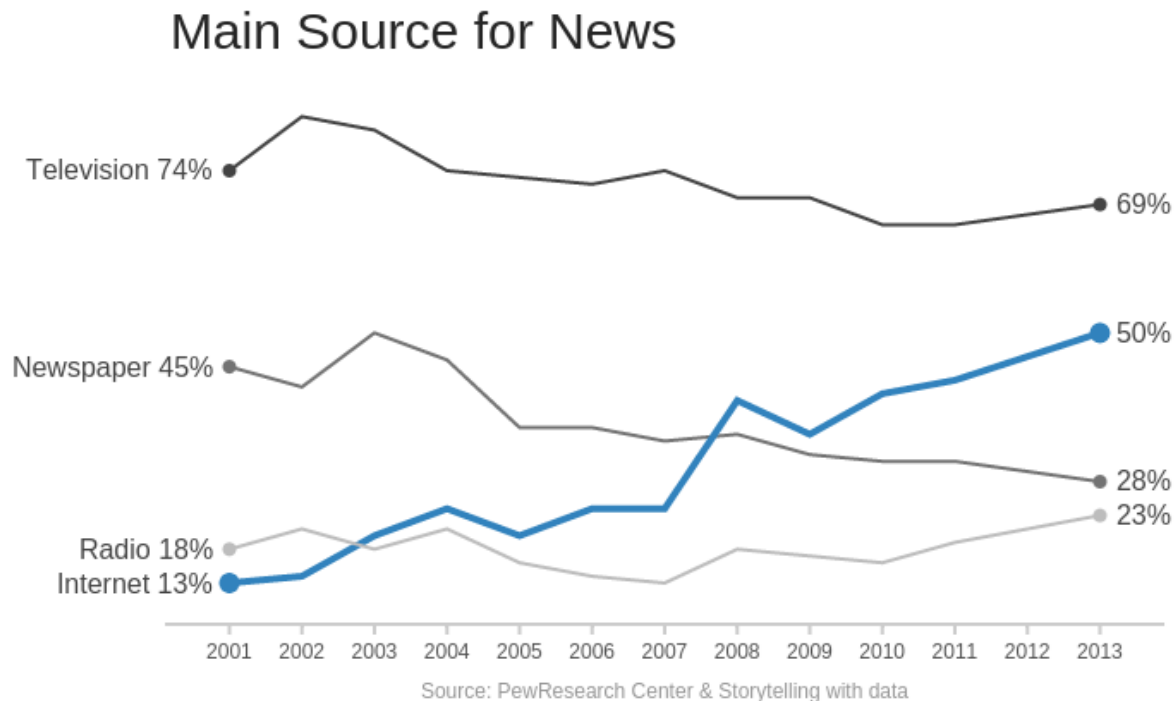
## Histograma

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

# Gráficos de línea

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

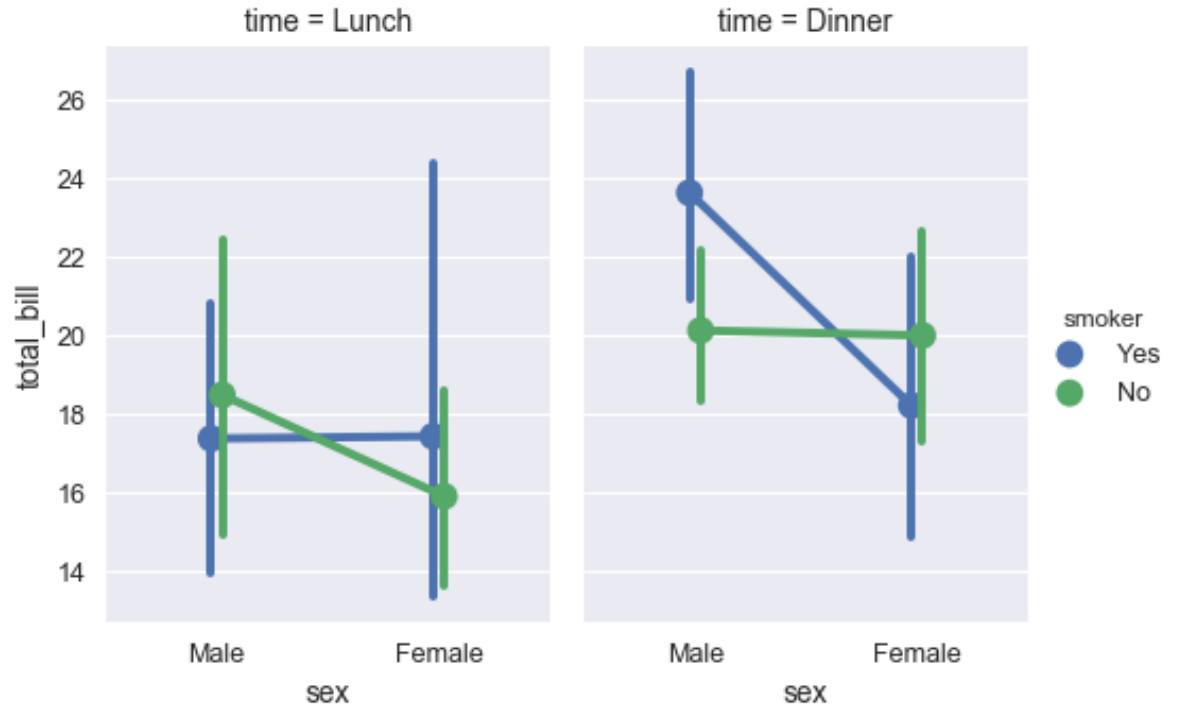
Múltiples líneas permiten  
comparar distintas categorías



# Gráficos de línea

Son muy versátiles

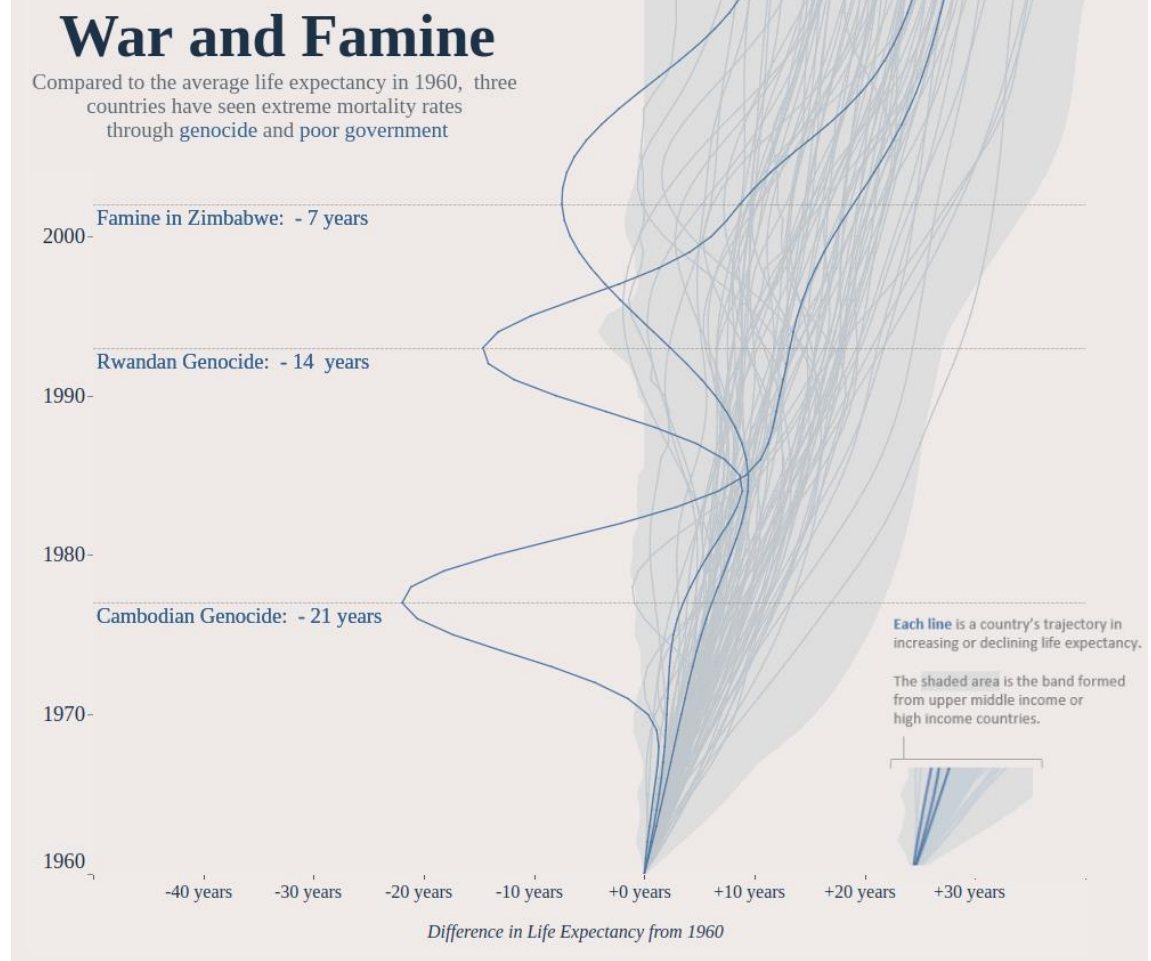
Generan gráficos visualmente  
simples, por lo que pueden  
contener mucha información



# Gráficos de línea

Son muy versátiles

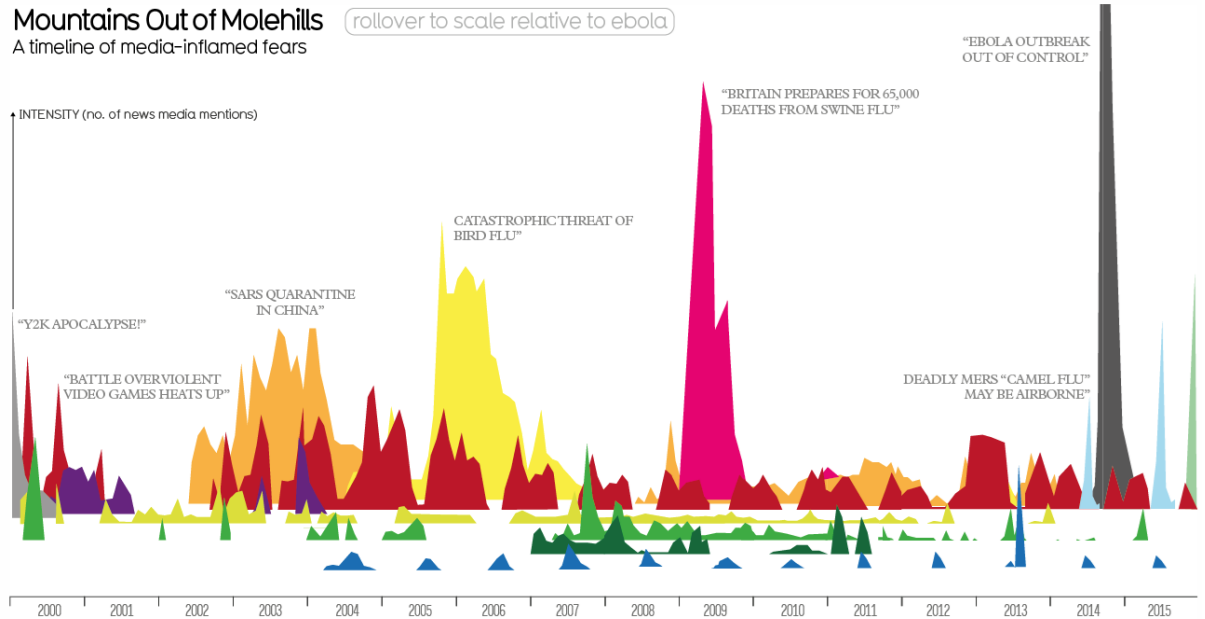
Generan gráficos visualmente  
simples, por lo que pueden  
contener mucha información





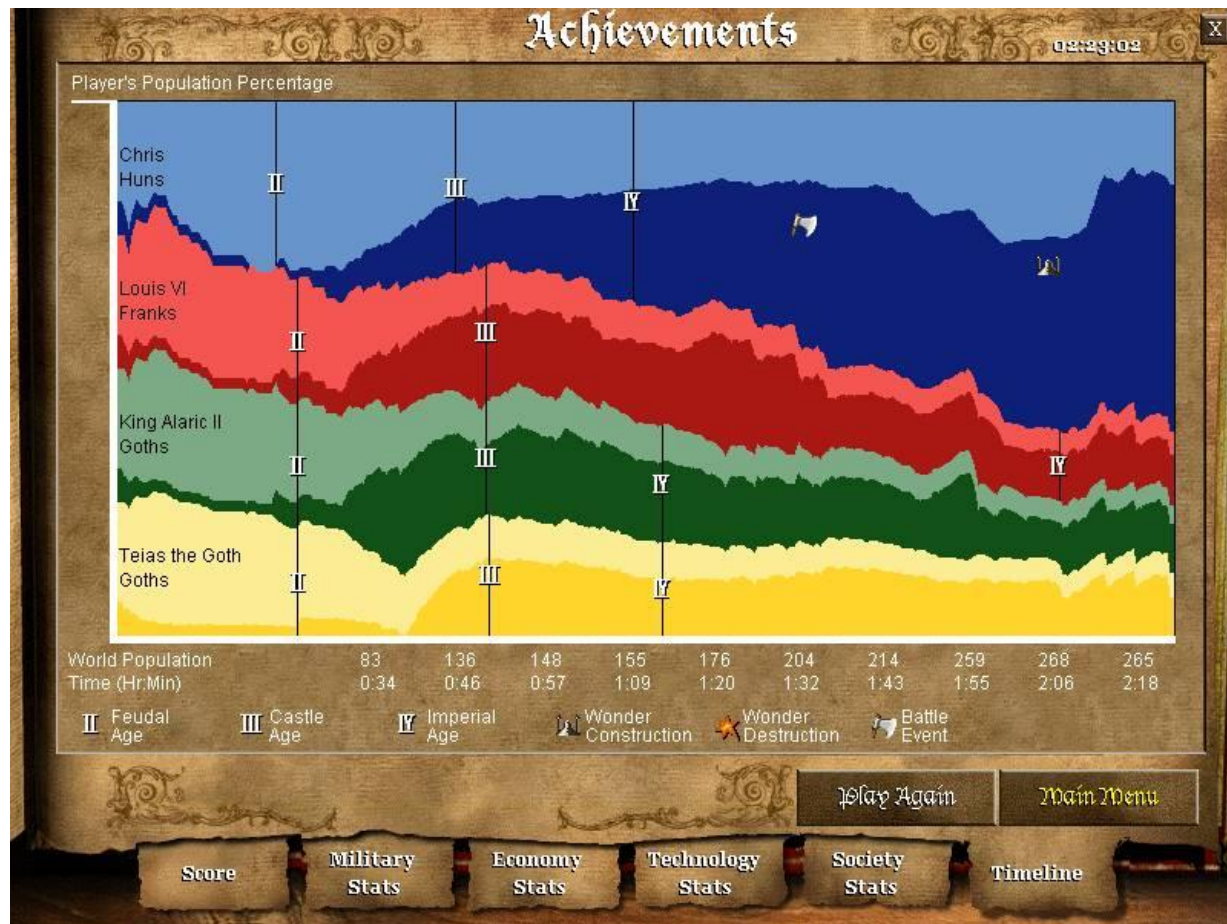
# Gráficos de área

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



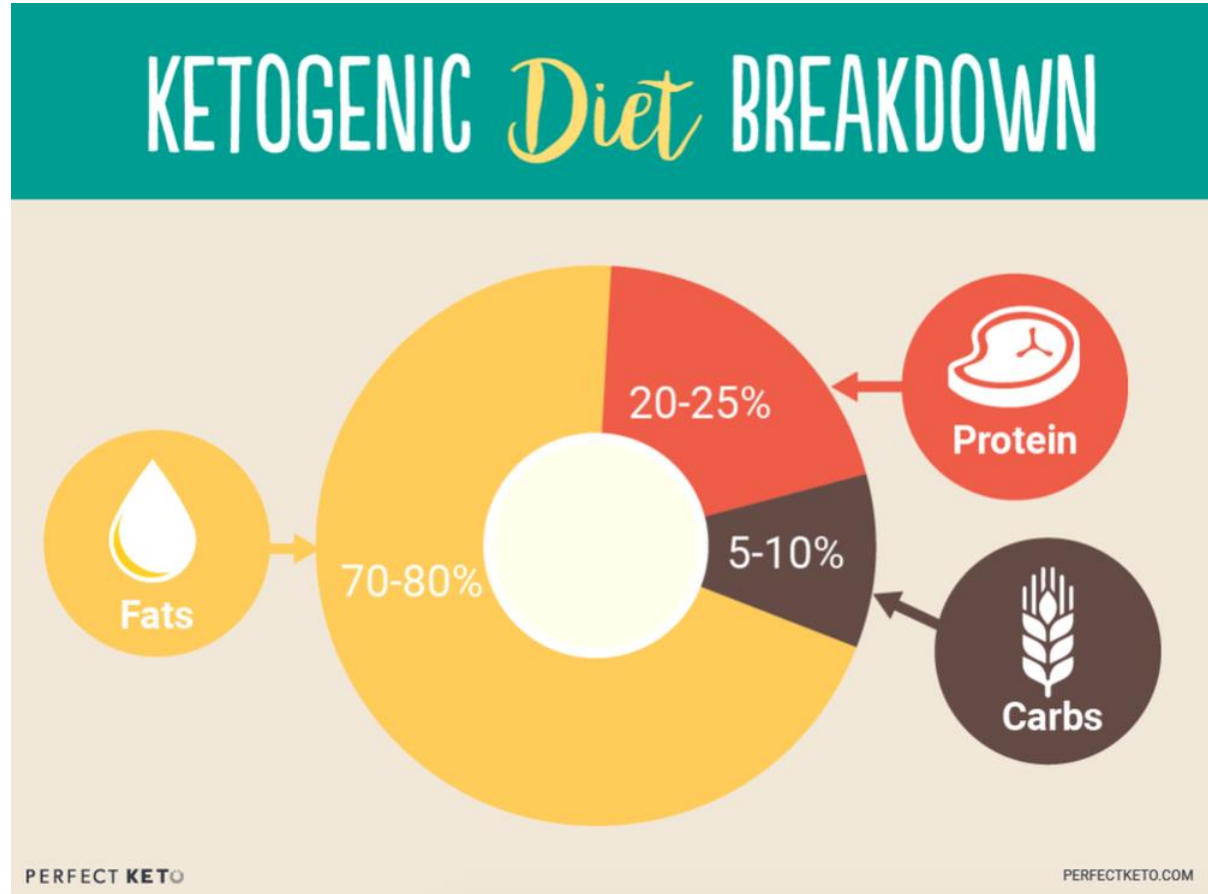
# Gráficos de área

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



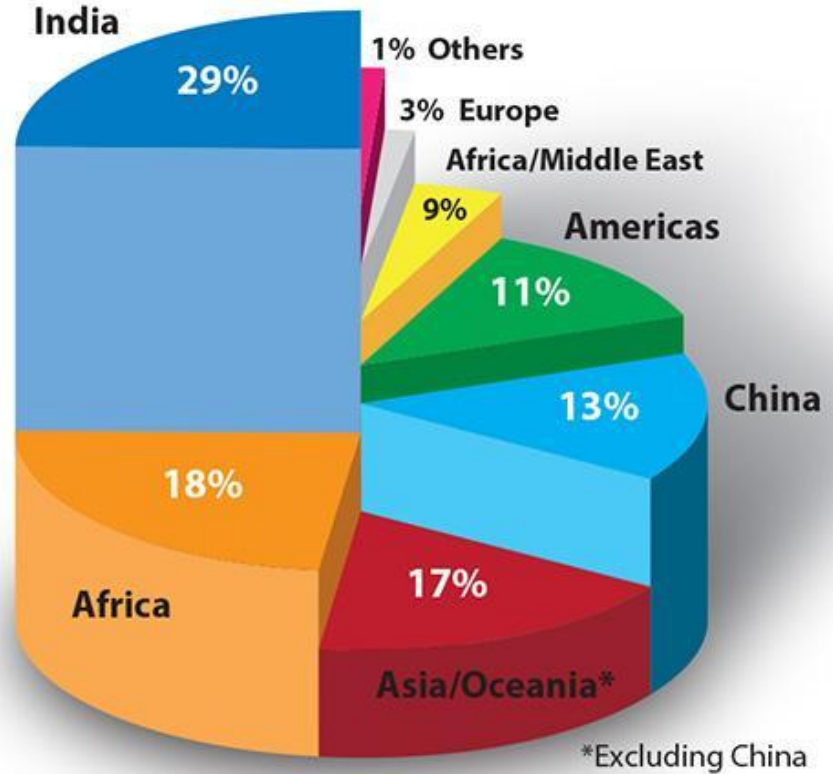
## Gráficos de torta

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



# The horror of pie charts

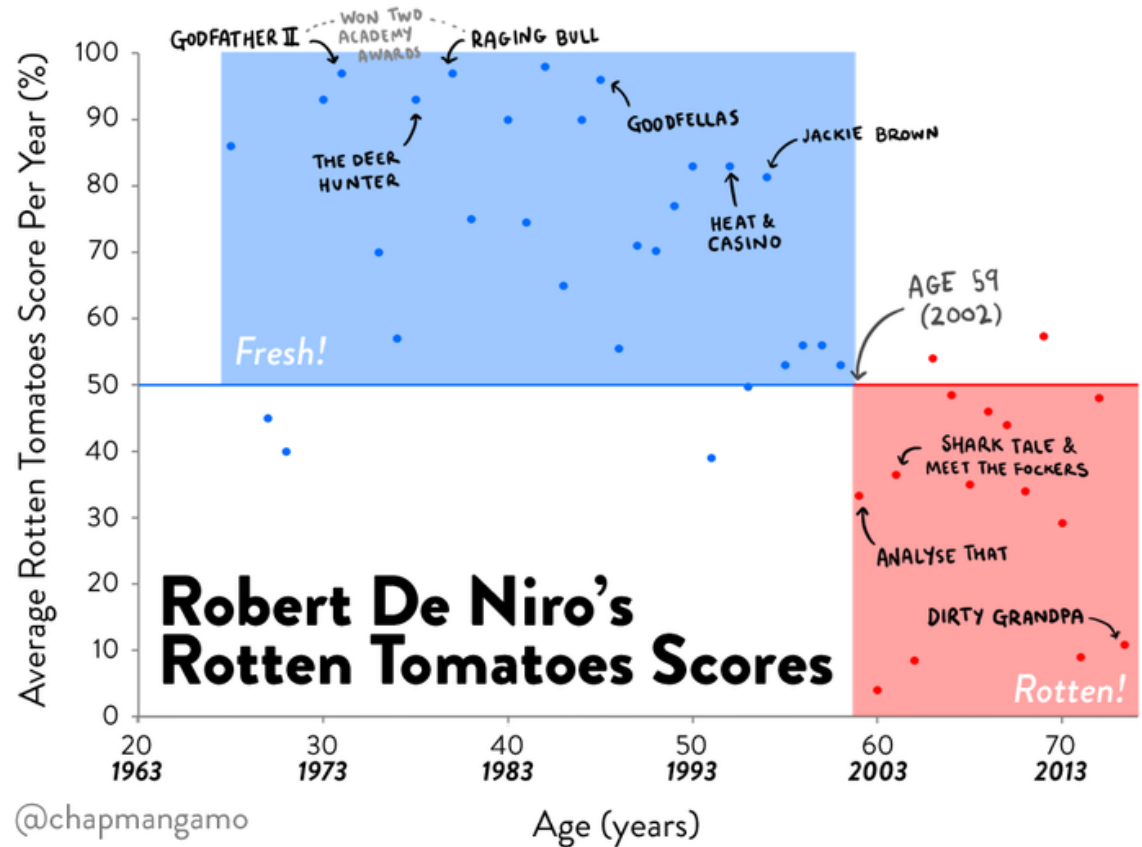
Share of worldwide urban population growth 2010-2050



Son tan malos, que seaborn  
no los tiene!

# Gráficos de 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](#)

# 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?