



Procesamiento de Señales II

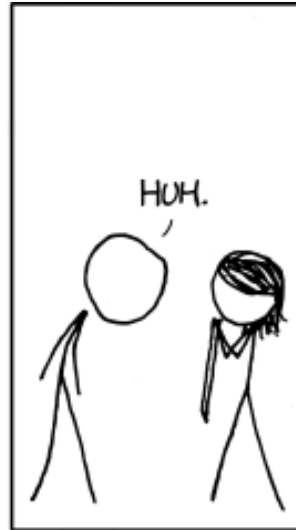
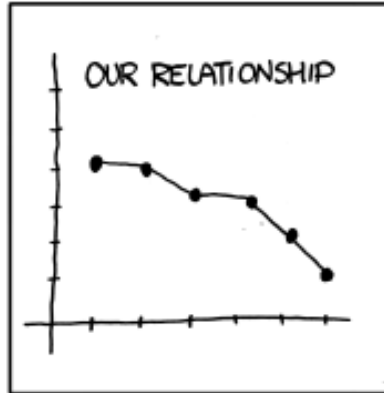
Ciencia de Datos II

**Análisis y Visualización de Datos - Visualización y
Comunicación**

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP
Director Semillero TRIAC
Ingeniería Electronica
Universidad Popular del Cesar

Visualización y comunicación



Visualización y comunicación





¿QUIÉNES SON LOS REFUGADOS?



80%

de todos los refugiados del mundo viven en países vecinos de su país de origen



2.3 millones

de los refugiados provienen de Sudán del Sur



80%

son mujeres y niños



50,000

de los niños son huérfanos



Los hijos de los refugiados tienen

5x

menos probabilidades de ir a la escuela que sus compañeros no refugiados



6

personas al día perdieron la vida tratando de llegar a Europa el año pasado



Doble

de personas desplazadas a nivel mundial en los últimos 20 años

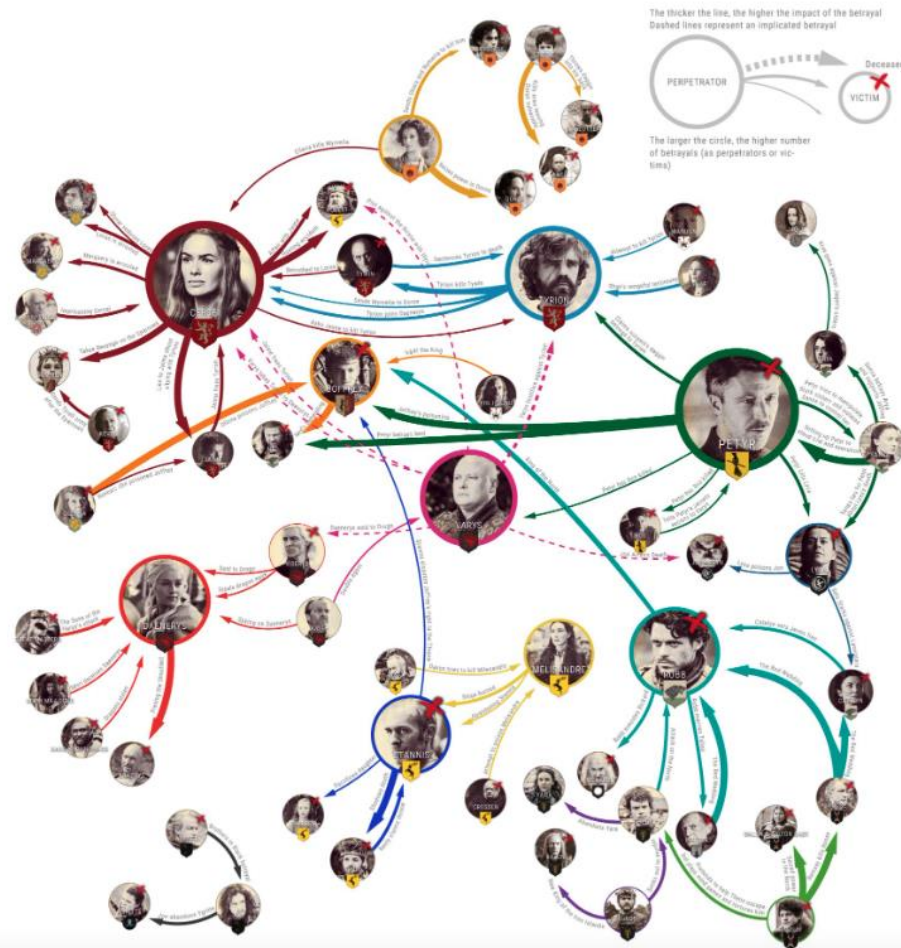


6.7 millones

de refugiados viven en los países más pobres del mundo

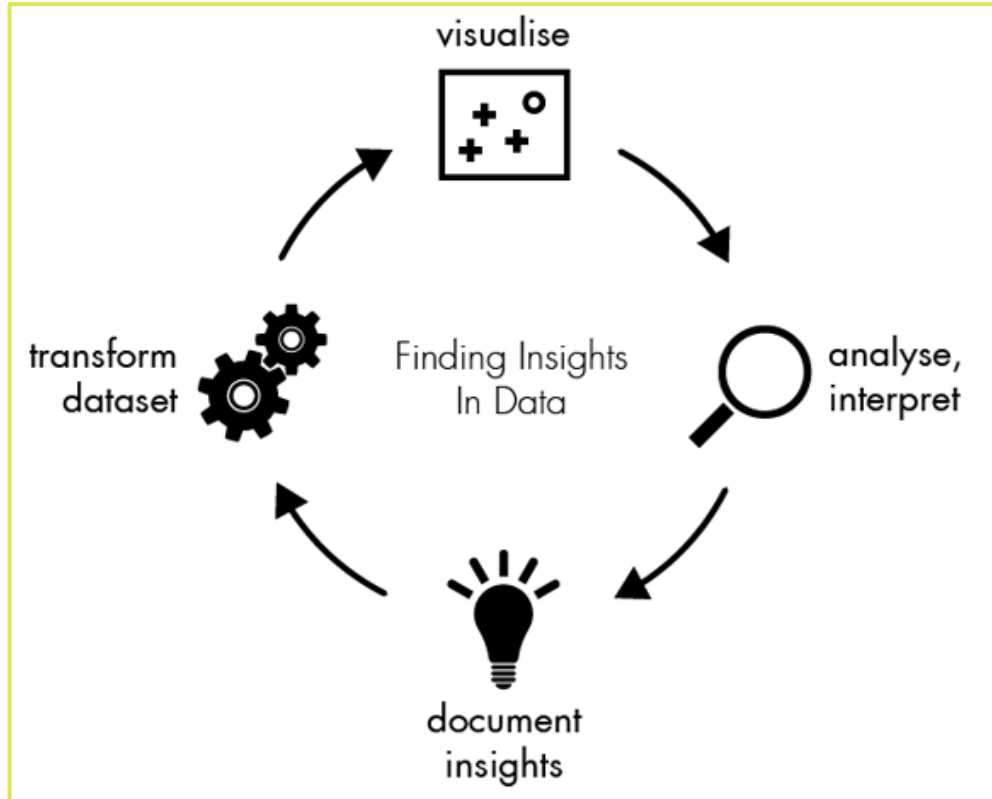
Crédito: [El arte de Medir](#)

Game of Thrones Web of Betrayals

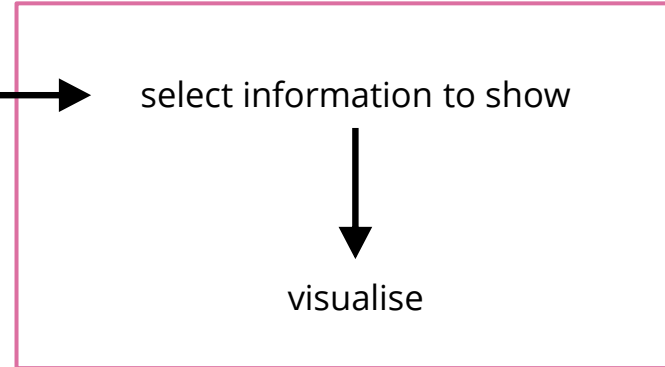


Crédito: [El arte de Medir](#)

Exploración vs presentación



Uso de la visualización de datos para encontrar información en los datos



Exploración

Visualizaciones para nosotros

“The way we represent a thing affects
the way we reason about a thing”

(La forma en que representamos una cosa afecta la forma en que razonamos sobre una cosa)

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

	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			Optimi	LR	Cell		batch_	dropc	epoch	embedc	hiddr	max_s	Filter	Merge	Pretra	Fineti	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/edmv/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/edmv/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/edmv/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/edmv/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/edmv/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	

Sesgos en la percepción

¿Qué pensamos cuando no nos paramos a pensar?

Patternicity bug!

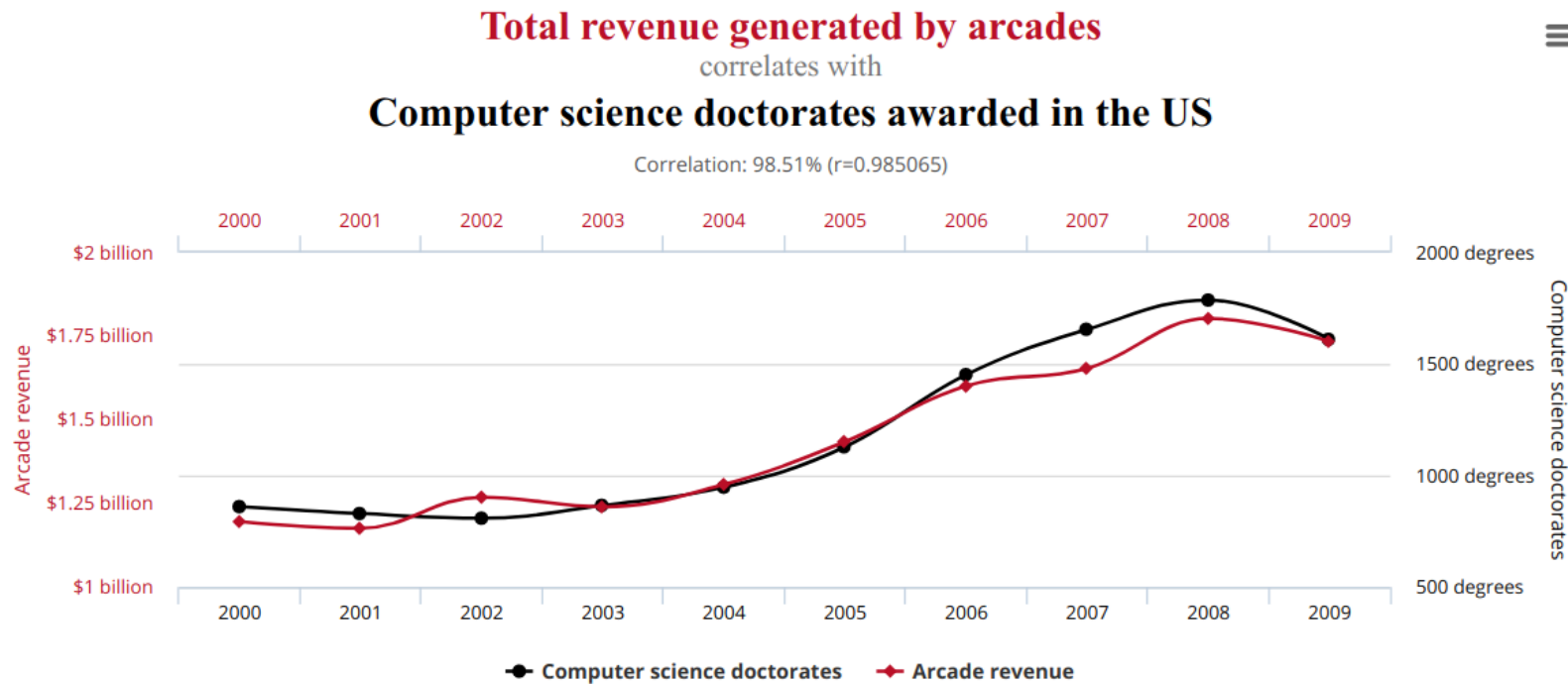
Tendencia a encontrar patrones en los objetos y a percibir el todo como algo más que la suma de sus partes.



The face of mars



Correlación no es causalidad



Data sources: U.S. Census Bureau and National Science Foundation

tylervigen.com

<https://www.tylervigen.com/spurious-correlations>

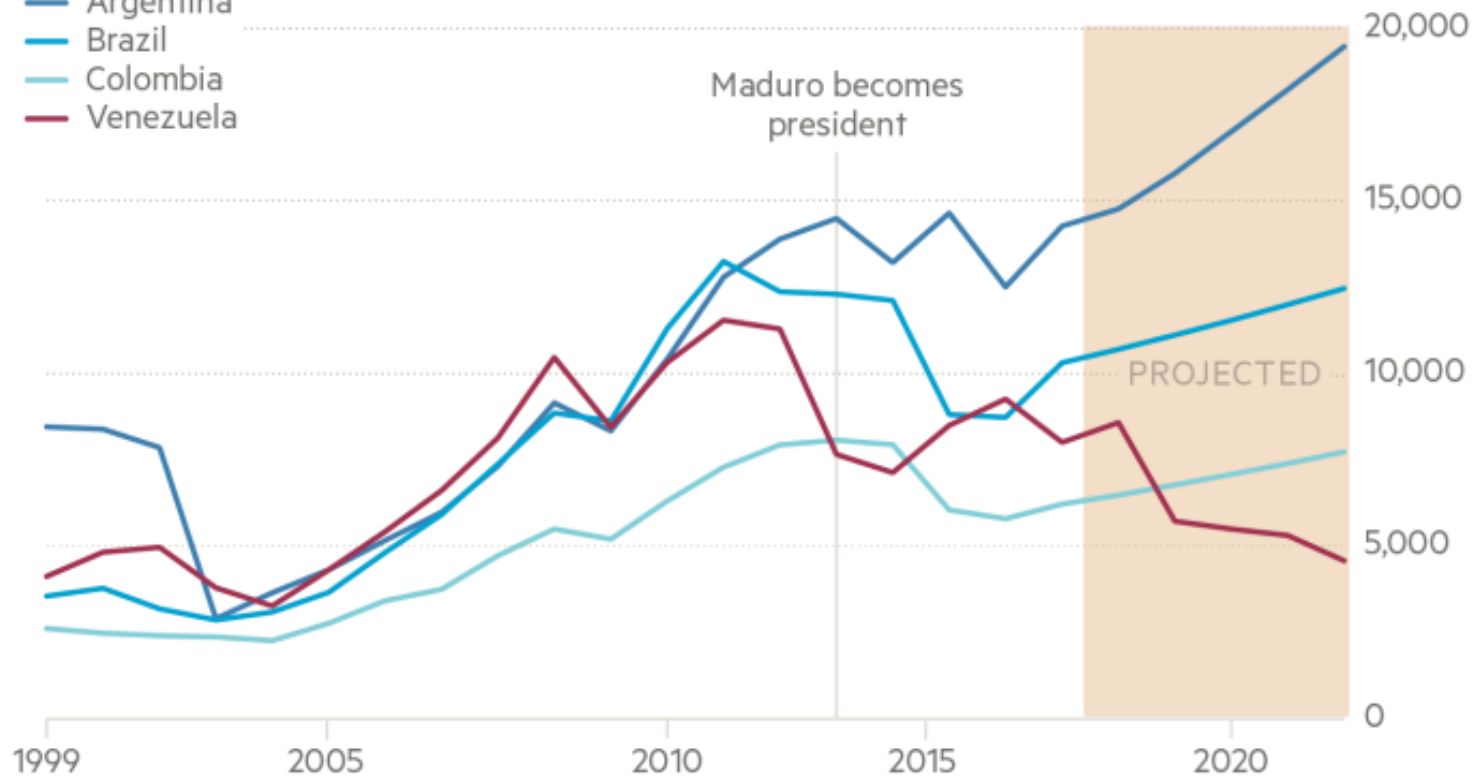
Storytelling bug!

Tendencia a encontrar racionalizaciones que expliquen la presencia de patrones en los objetos

Oil-rich Venezuela will have a lower per-capita GDP than its peers

GDP per capita in current US dollars

- Argentina
- Brazil
- Colombia
- Venezuela

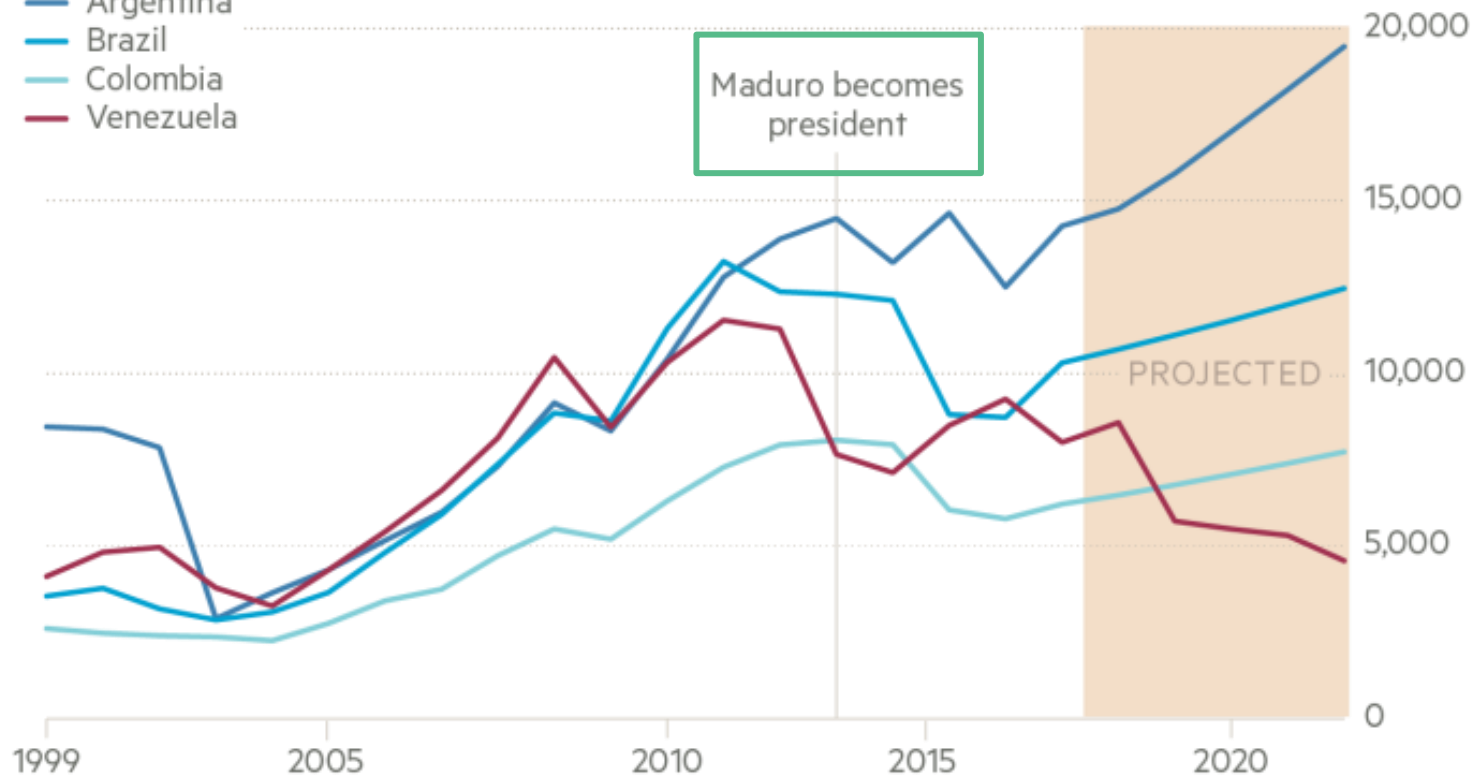


Source: IMF World Economic Outlook Database

Oil-rich Venezuela will have a lower per-capita GDP than its peers

GDP per capita in current US dollars

- Argentina
- Brazil
- Colombia
- Venezuela



Source: IMF World Economic Outlook Database

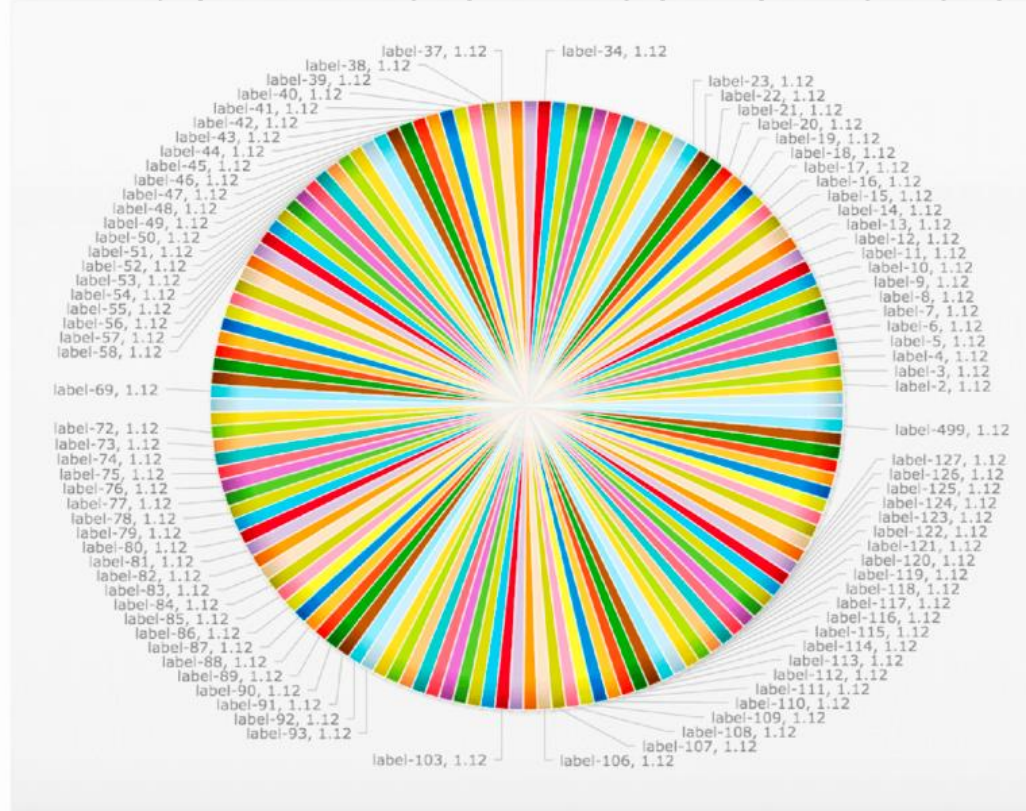
Confirmation bug!

Tendencia a creer (más) verdadera la información que sustenta nuestras creencias ya existentes.

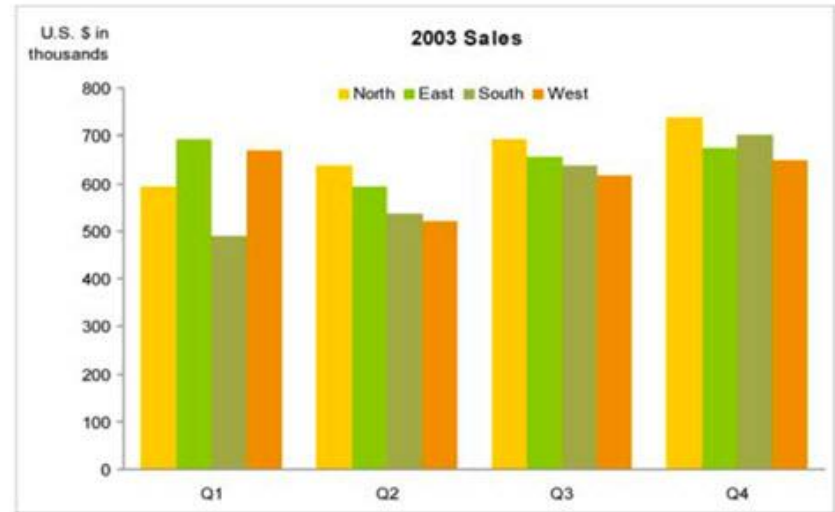
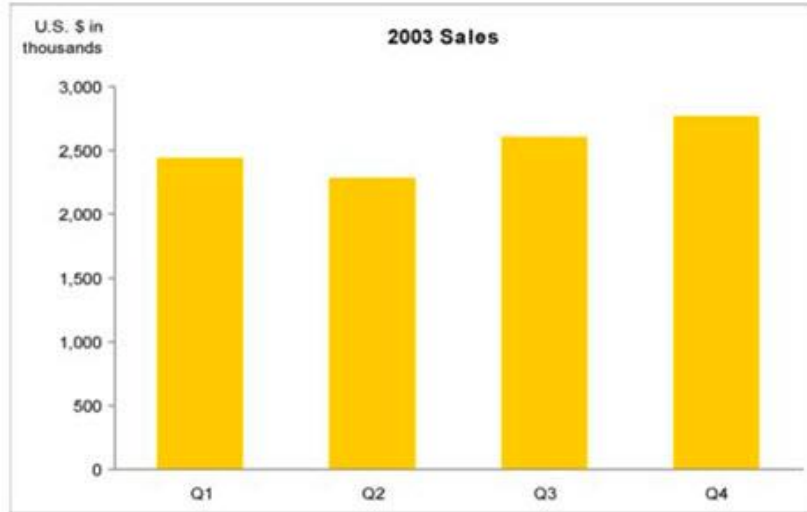
Presentación

Visualizaciones para otros

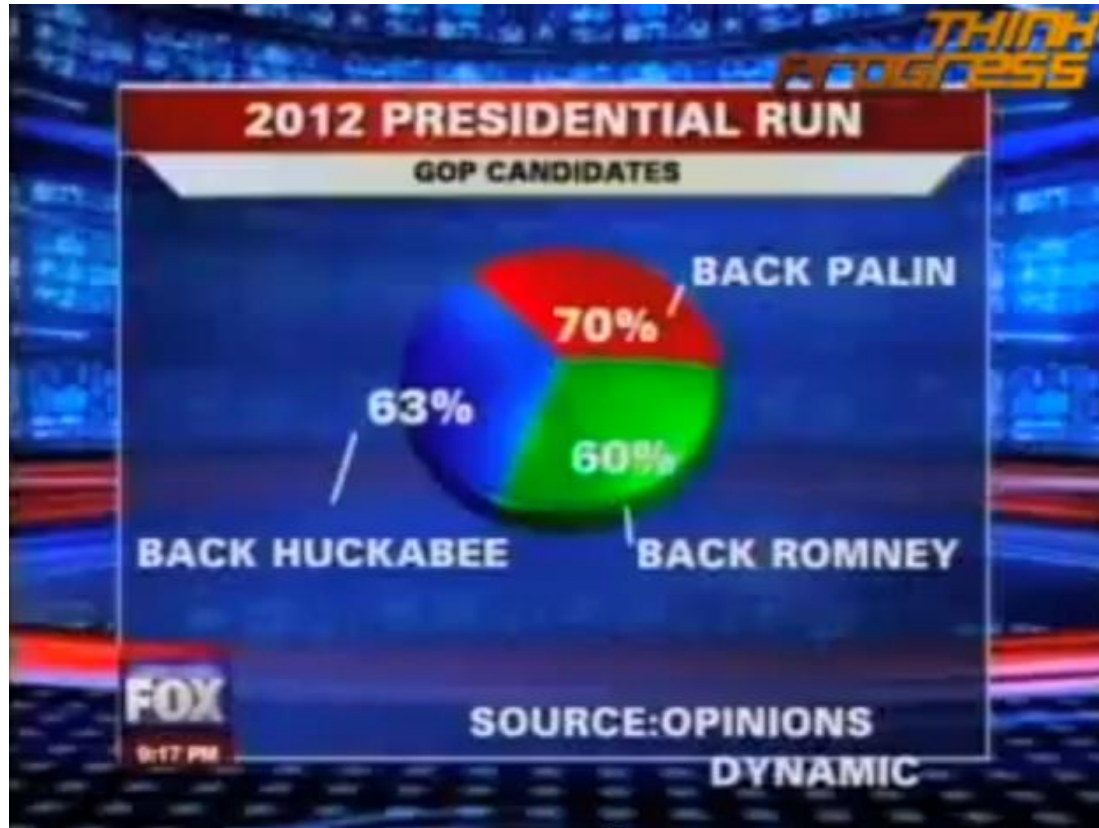
VISUALIZACIÓN Y COMUNICACIÓN



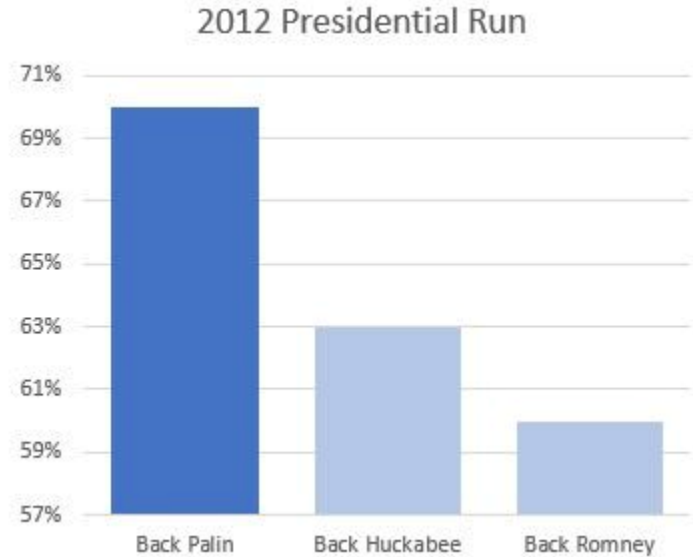
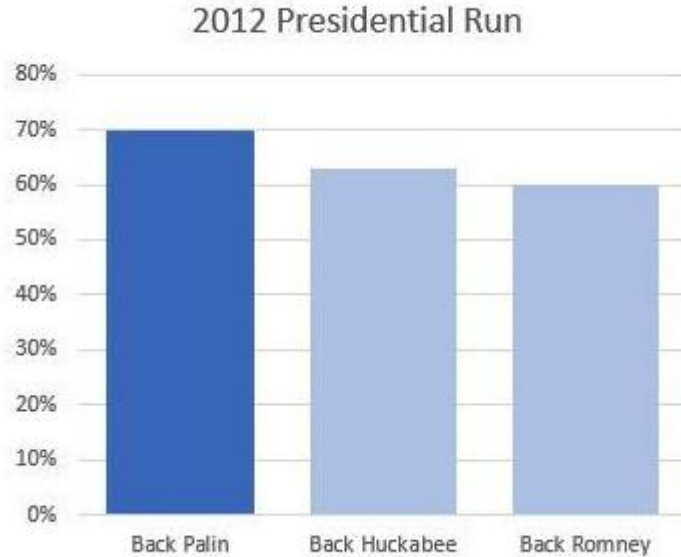
VISUALIZACIÓN Y COMUNICACIÓN



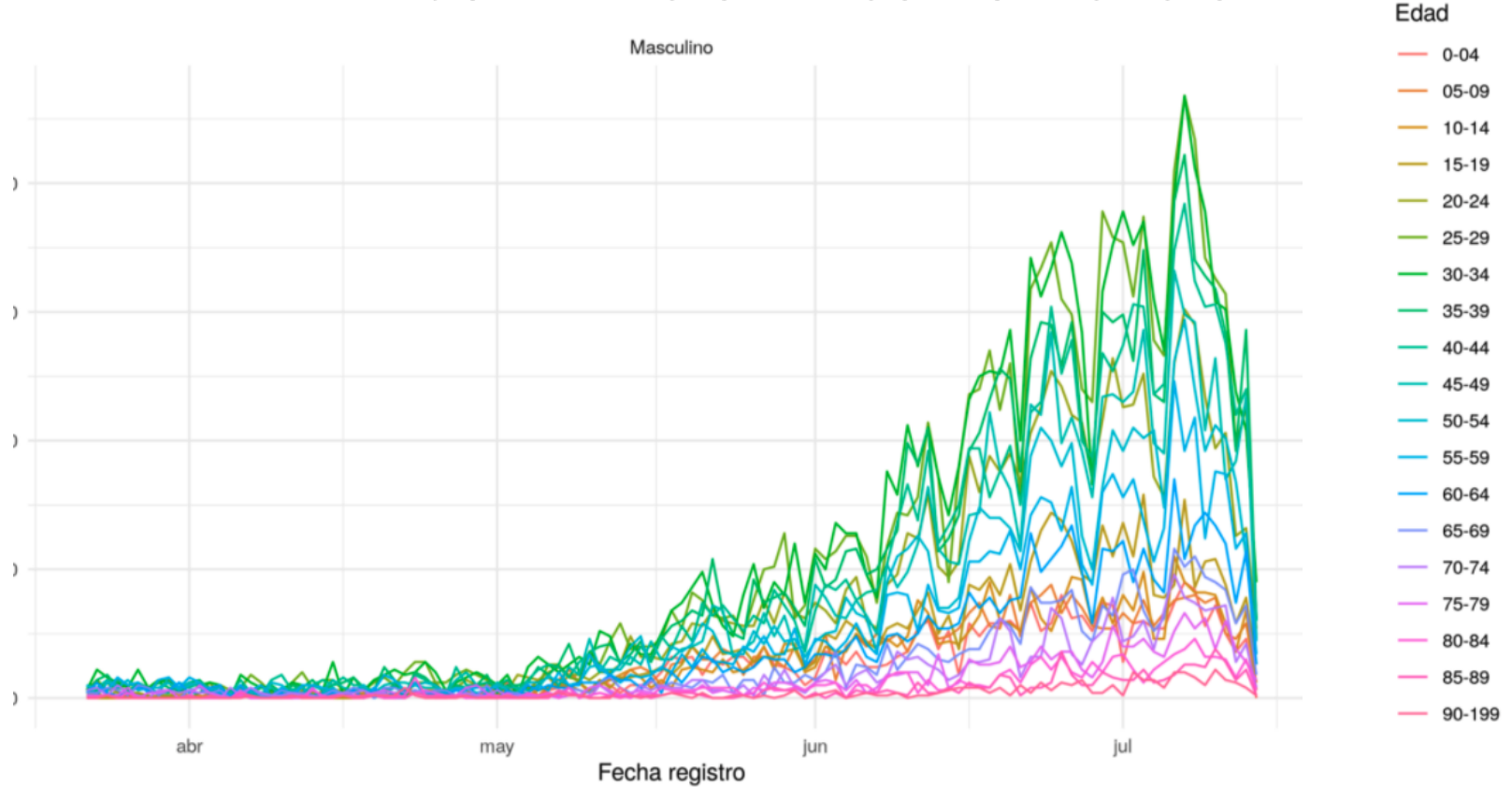
VISUALIZACIÓN Y COMUNICACIÓN



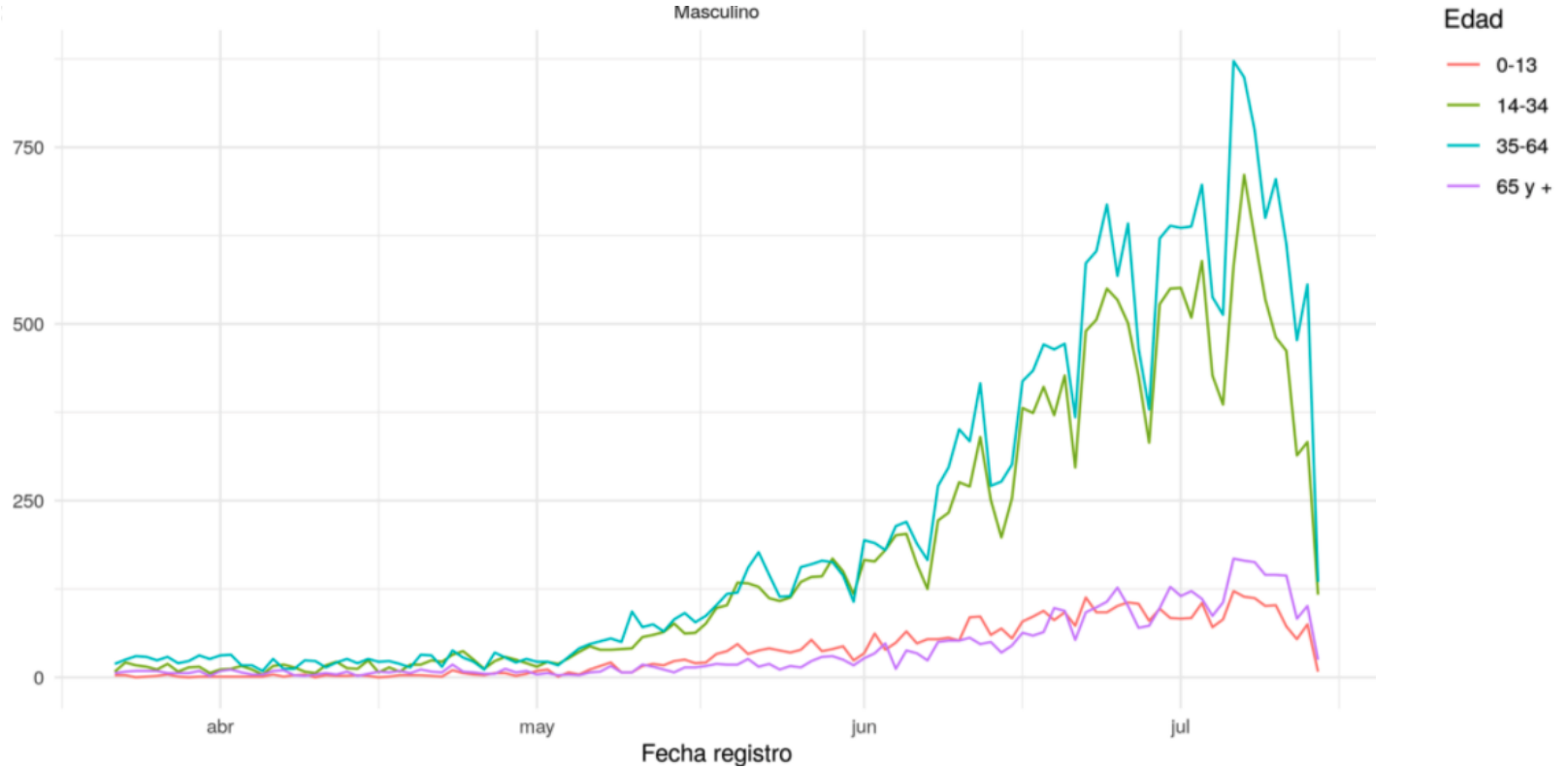
VISUALIZACIÓN Y COMUNICACIÓN



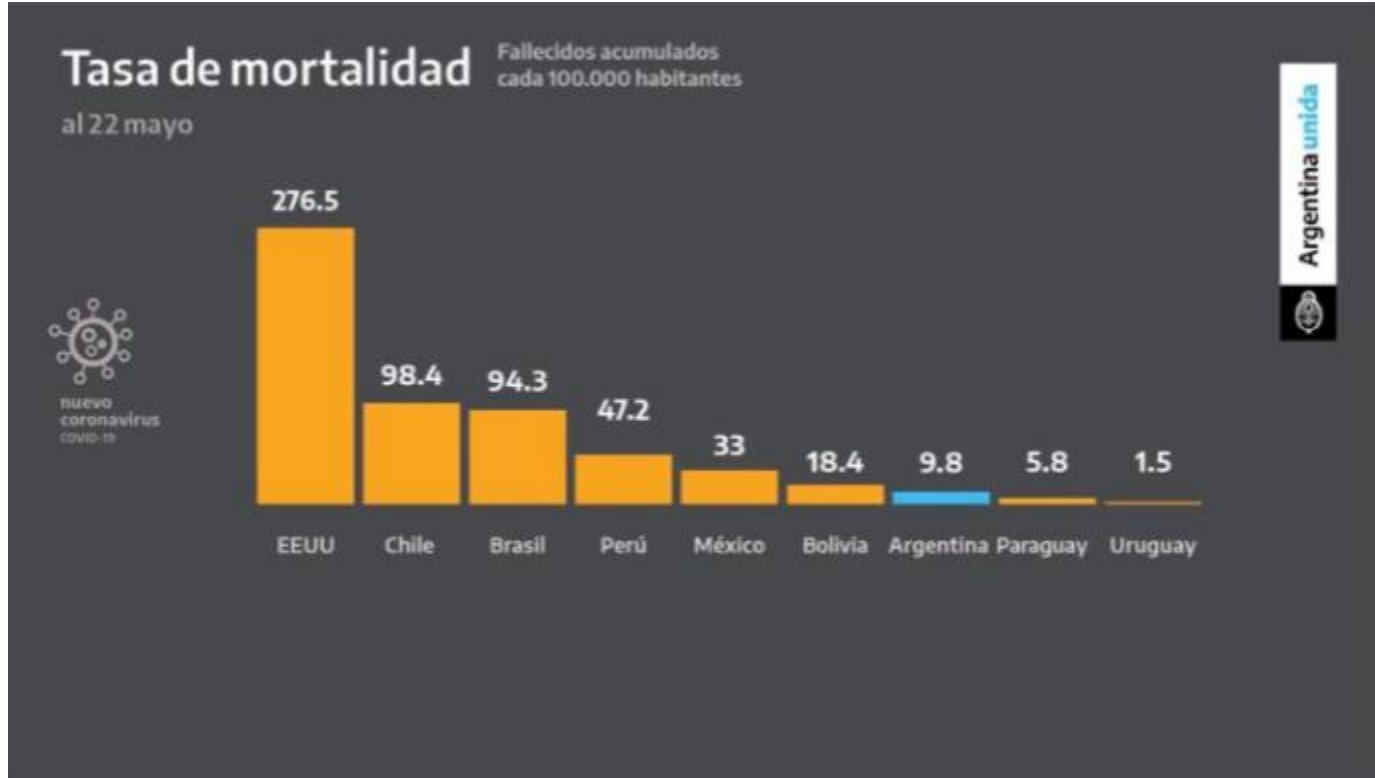
VISUALIZACIÓN Y COMUNICACIÓN



VISUALIZACIÓN Y COMUNICACIÓN



VISUALIZACIÓN Y COMUNICACIÓN



VISUALIZACIÓN Y COMUNICACIÓN

País	Población Total	Total Fallecimientos	Tasa de Mortalidad cada 100 mil hab.	Tasa de Mortalidad cada 1 millón hab.
Estados Unidos	331.002.647	94.702	28,61	286,11
Chile	19.116.209	589	3,08	30,81
Brasil	212.559.409	20.047	9,43	94,31
Perú	32.971.846	3.148	9,55	95,48
México	128.932.753	6.510	5,05	50,49
Bolivia	11.673.029	215	1,84	18,42
Argentina	45.195.777	416	0,92	9,20
Paraguay	7.132.530	11	0,15	1,54
Uruguay	3.473.727	20	0,58	5,76

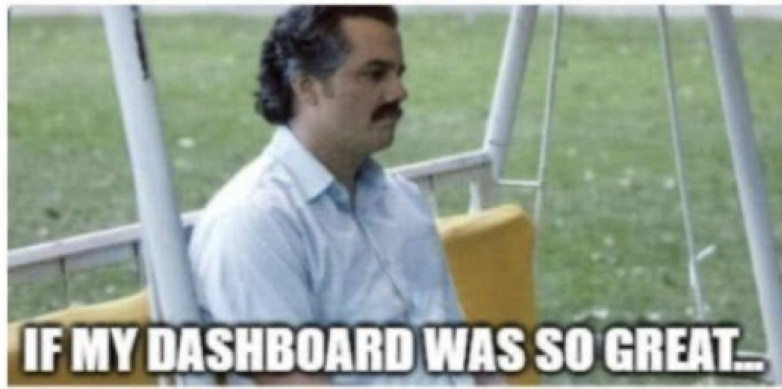
Optimizar el proceso
de comunicación



Generar un mensaje
que se decodifique
rápida y fielmente

Características de una buena visualización

- Honesta: representa datos que son correctos.
- Funcional: representa datos para que puedan ser interpretados adecuadamente.
- Estética
- Esclarecedora: Debe mostrar patrones que no serían fácilmente percibidos usando otros medios.
- Informativo



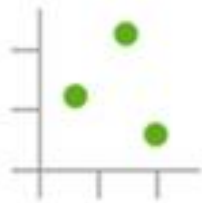
?



Encodings visuales

Mapeo datos  elementos visuales

¿Qué son los elementos visuales?



Position



Length



Angle/Slope



Area



Volume



Difference



Color hue



Color Saturation



Contrast



Texture

¿Cómo elegir los elementos visuales?

Principio de la **consistencia**: las propiedades de la imagen deben corresponderse con las propiedades de los datos.

- The lie factor

Principio del **ordenamiento por importancia**: la información más importante debe ser codificada de la forma más eficiente posible.

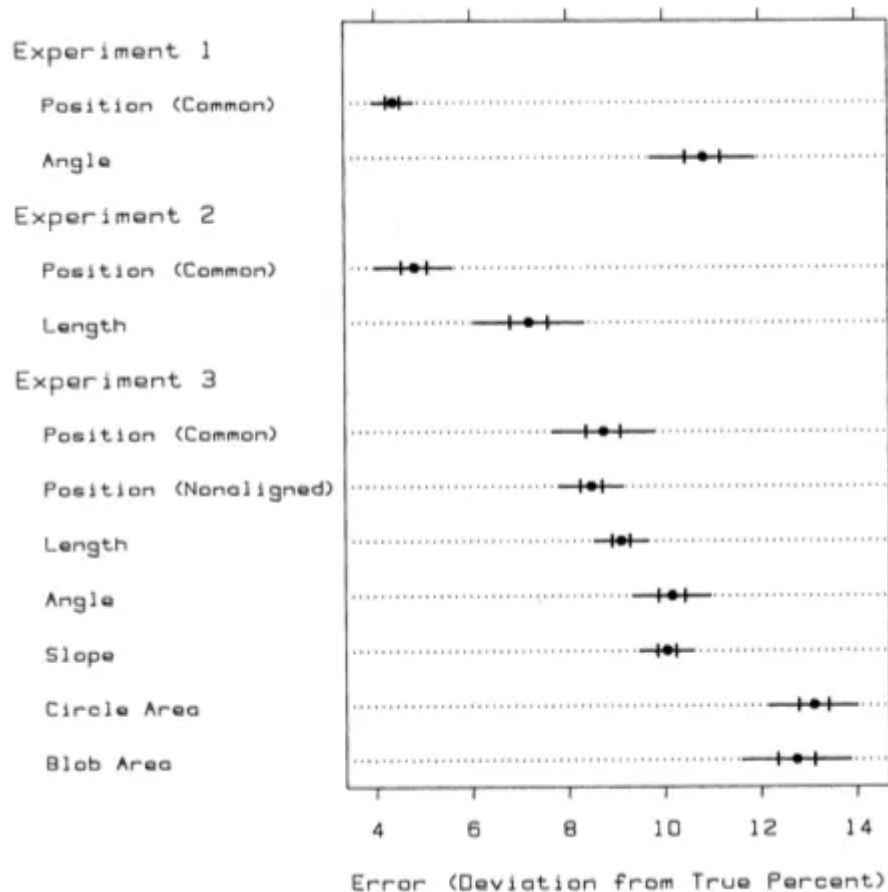
- ¿Cuál es la información más importante?
- ¿Cuáles son los encodings más efectivos?

No todos los elementos
visuales son igual de efectivos

¿Cómo medimos el error?

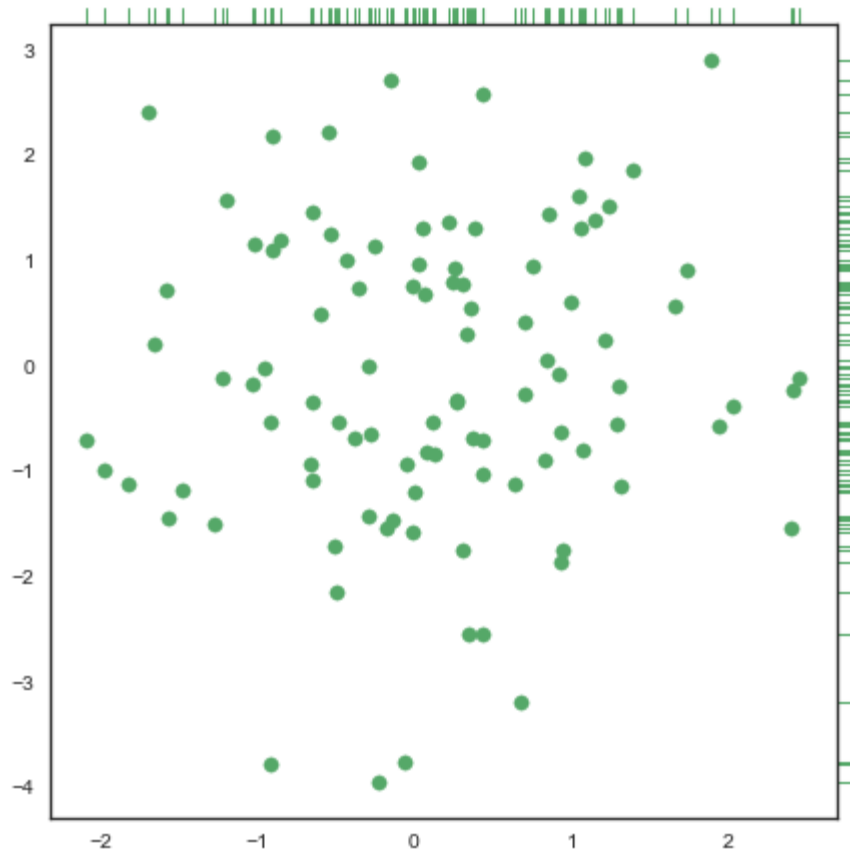
Distintos encodings nos permiten **estimar** mejor o peor la diferencia entre dos cantidades

William S. Cleveland and Robert McGill,
Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods, 1984



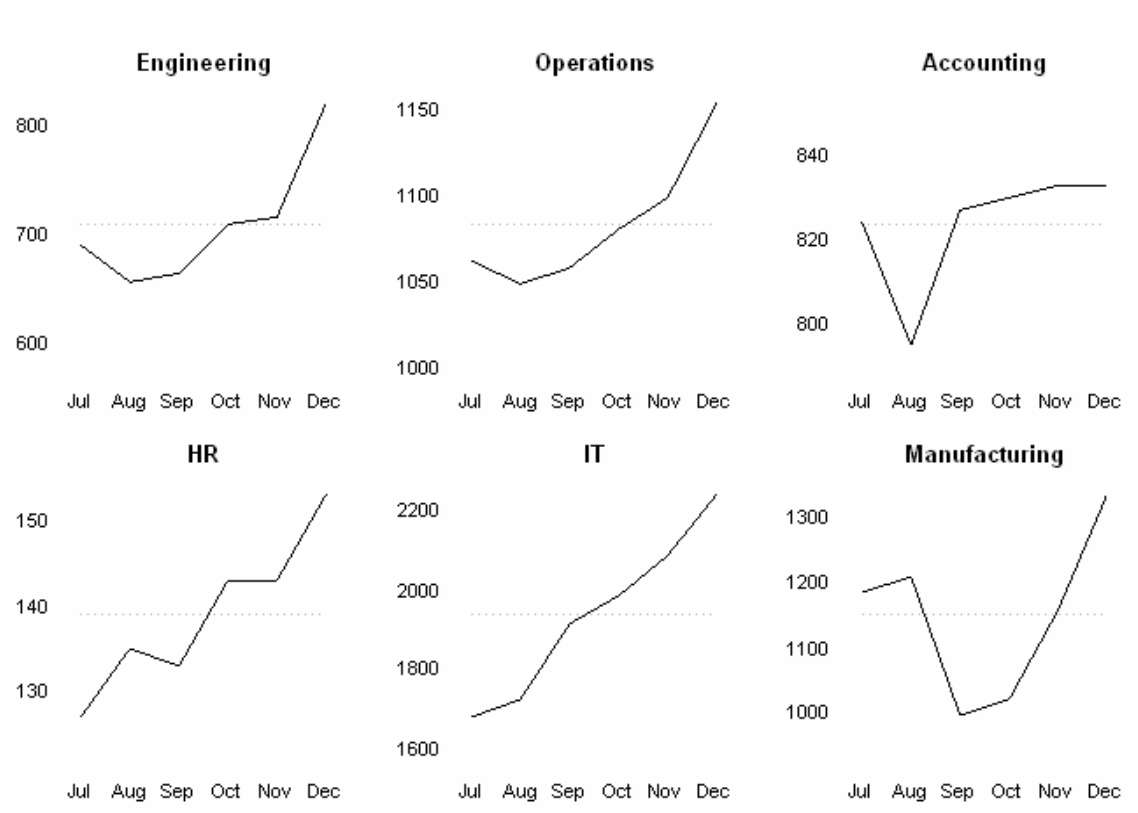
¿Cómo medimos el error?

1. **Posición en una escala común**
2. Posición en escalas no alineadas
3. Largo
4. Ángulo e inclinación (empate)
5. Área
6. Volumen, densidad y saturación de color (empate)
7. Tonalidad cromática



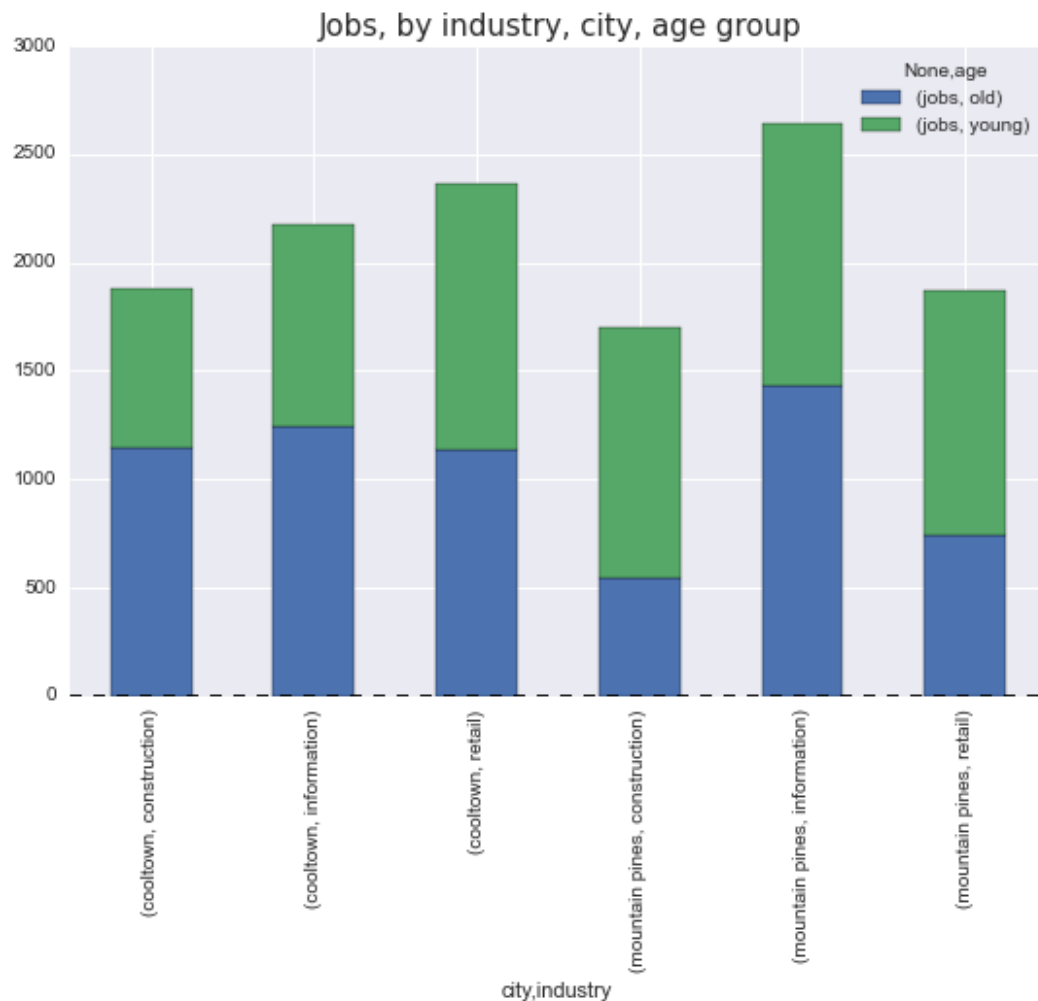
¿Cómo medimos el error?

1. Posición en una escala común
2. **Posición en escalas no alineadas**
3. Largo
4. Ángulo e inclinación
5. Área
6. Volumen, densidad y saturación de color
7. Tonalidad cromática



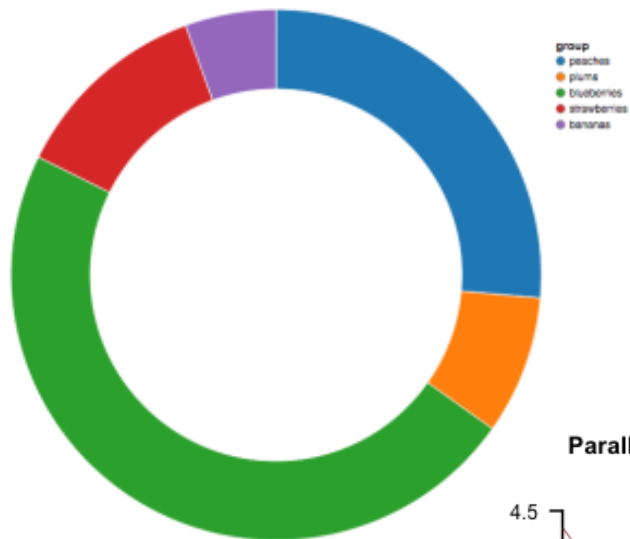
¿Cómo medimos el error?

1. Posición en una escala común
2. Posición en escalas no alineadas
3. **Largo**
4. Ángulo e inclinación
5. Área
6. Volumen, densidad y saturación de color
7. Tonalidad cromática

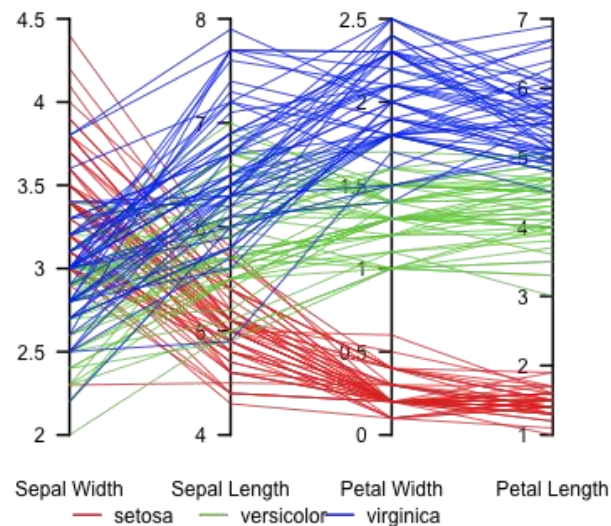


¿Cómo medimos el error?

1. Posición en una escala común
2. Posición en escalas no alineadas
3. Largo
4. **Ángulo e inclinación**
5. Área
6. Volumen, densidad y saturación de color
7. Tonalidad cromática

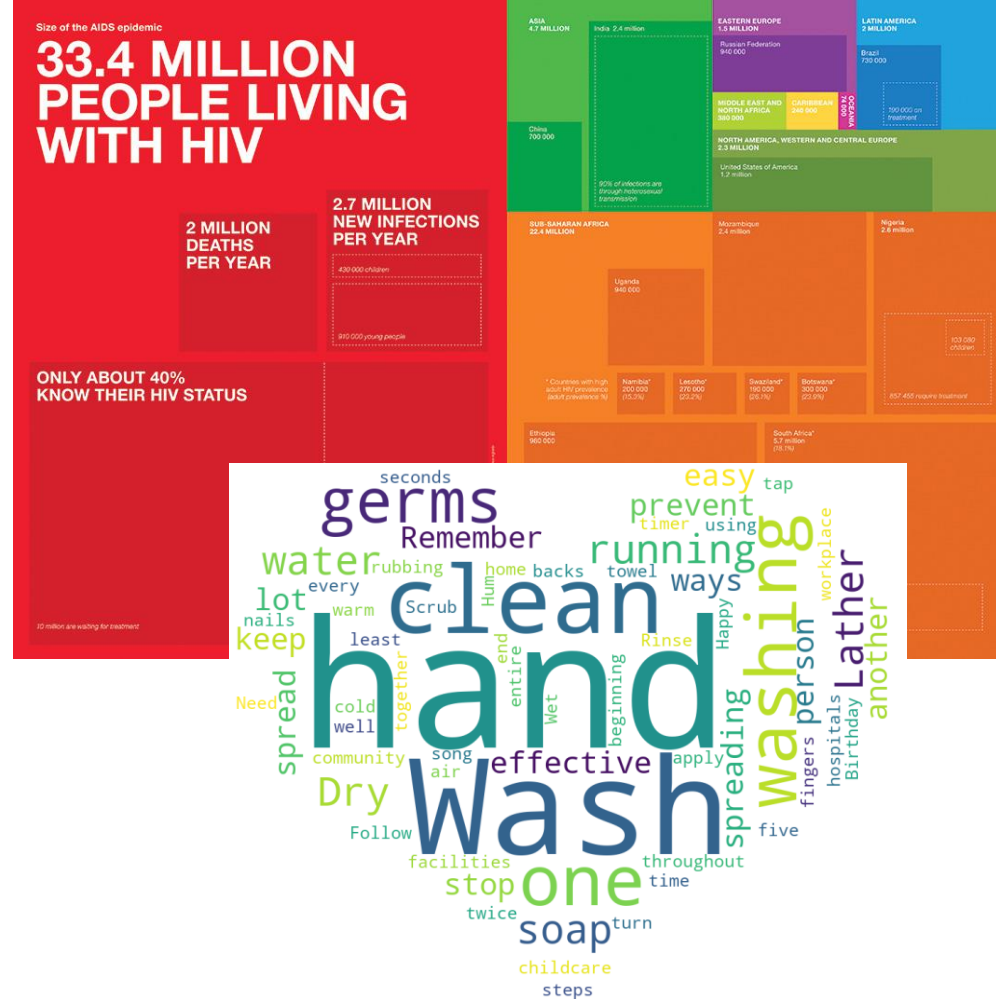


Parallel coordinate plot, Fisher's Iris data



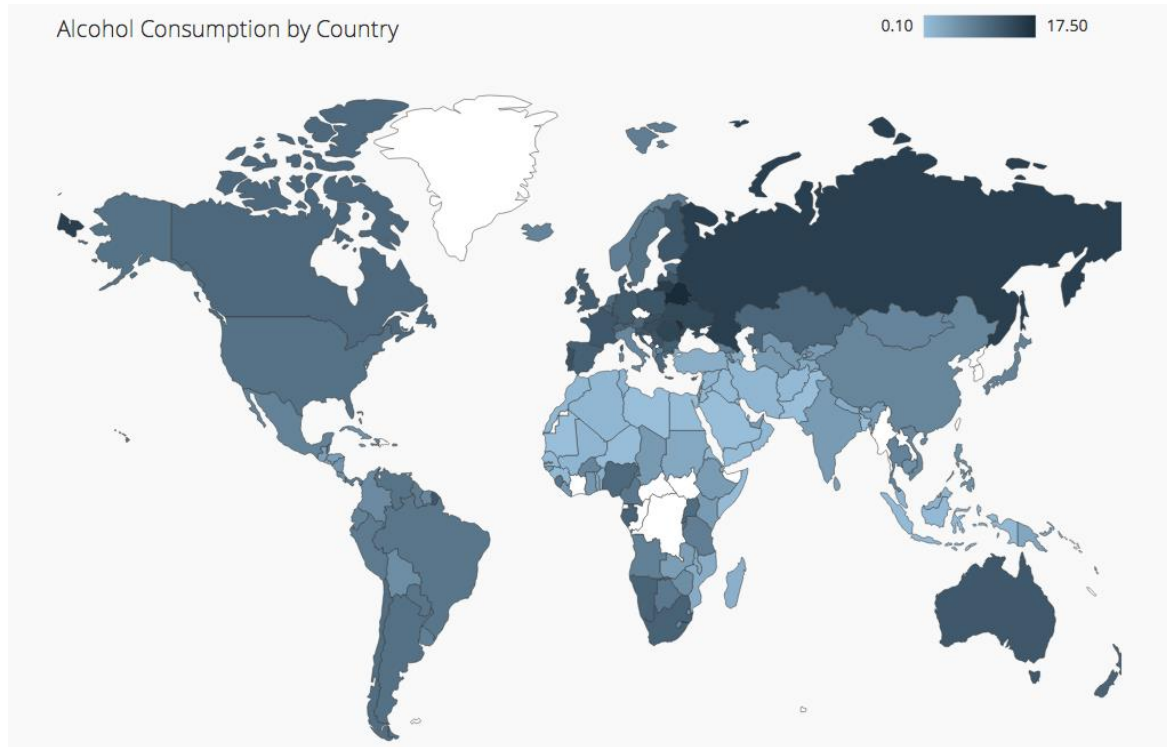
¿Cómo medimos el error?

1. Posición en una escala común
2. Posición en escalas no alineadas
3. Largo
4. Ángulo e inclinación (empate)
5. Área
6. Volumen, densidad y saturación de color (empate)
7. Tonalidad cromática



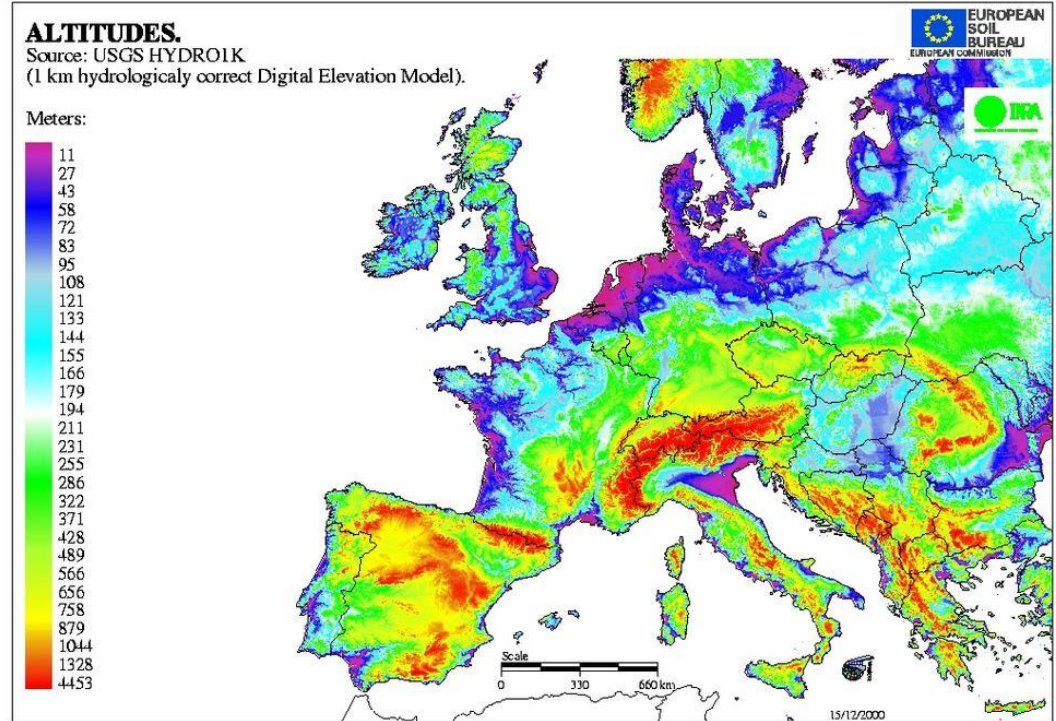
¿Cómo medimos el error?

1. Posición en una escala común
2. Posición en escalas no alineadas
3. Largo
4. Ángulo e inclinación
5. Área
6. **Volumen, densidad y saturación de color**
7. Tonalidad cromática



¿Cómo medimos el error?

1. Posición en una escala común
2. Posición en escalas no alineadas
3. Largo
4. Ángulo e inclinación
5. Área
6. Volumen, densidad y saturación de color
7. **Tonalidad cromática**



Encodings de acuerdo al tipo de datos

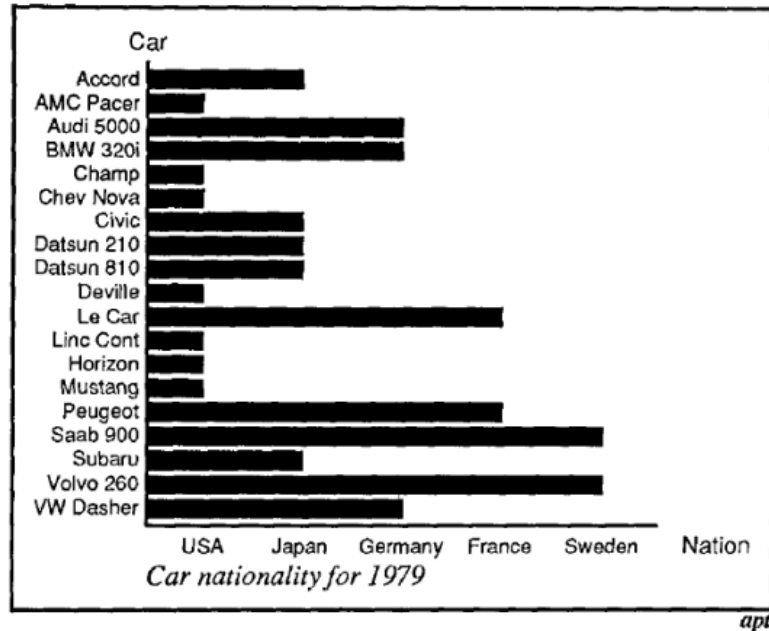


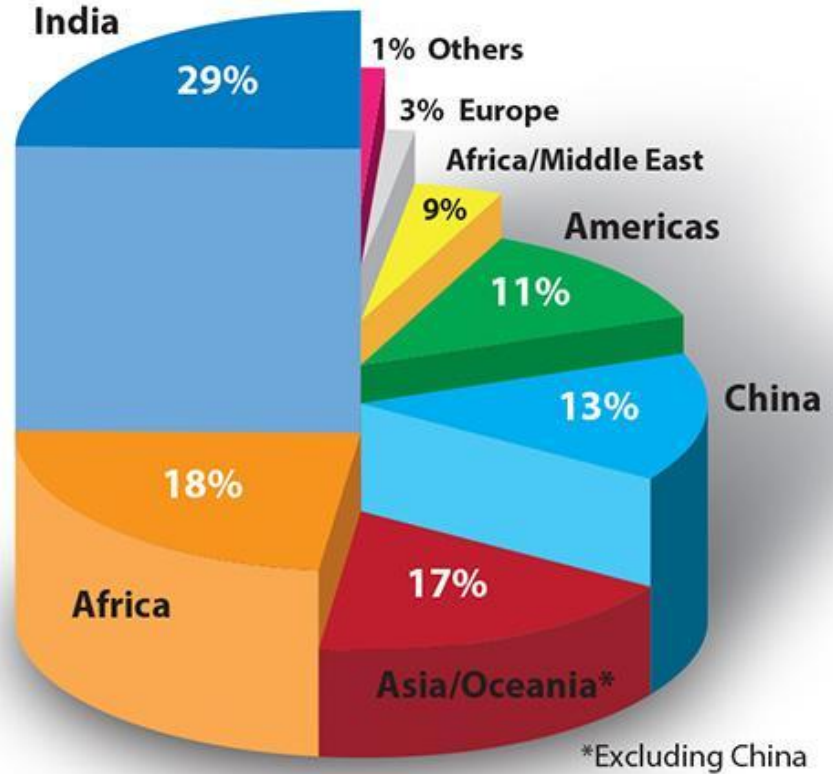
Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

El largo de las barras está utilizado para codificar una variable categórica, lo cual sugiere una jerarquía de valores.

Además, es difícil comparar los distintos países entre sí.

The horror of pie charts

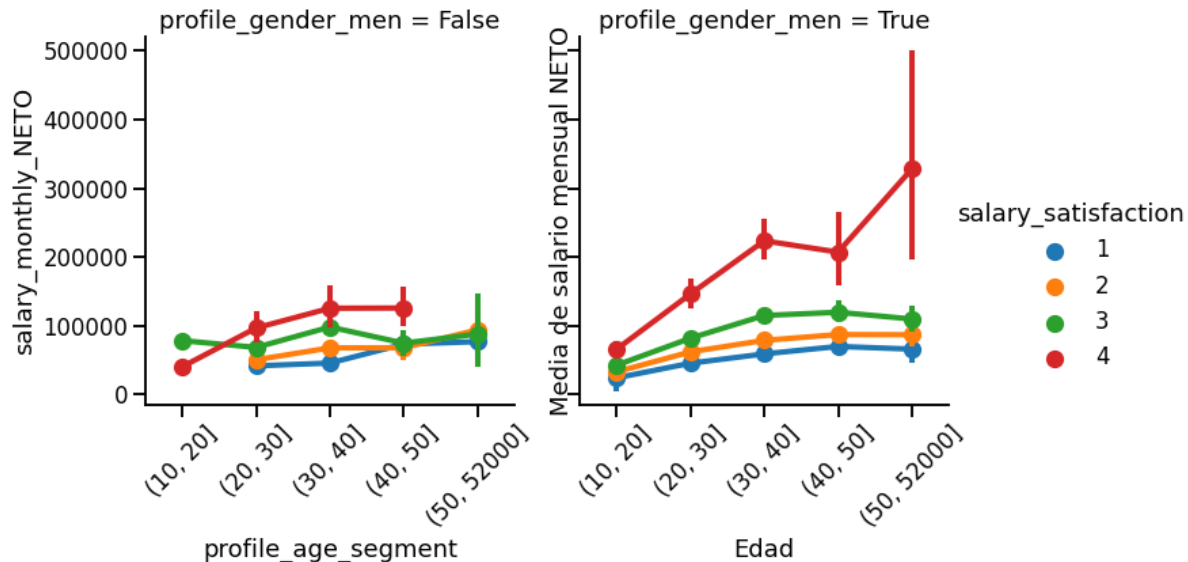
Share of worldwide urban population growth 2010-2050



Gráficos complejos

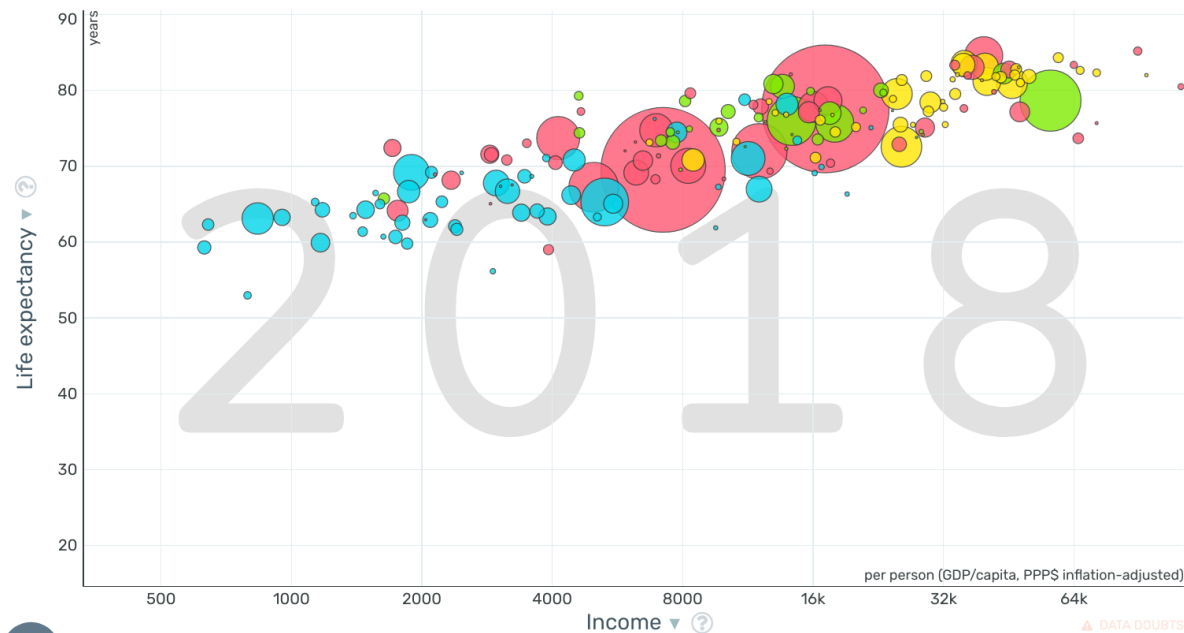
Agregando más variables

- Cada nueva variable que agregamos necesita un nuevo encoding.
- En seaborn la estructura es ligeramente distinta
 - [catplot](#)
- Otras librerías como Plotly permiten gráficos más complejos e interactivos



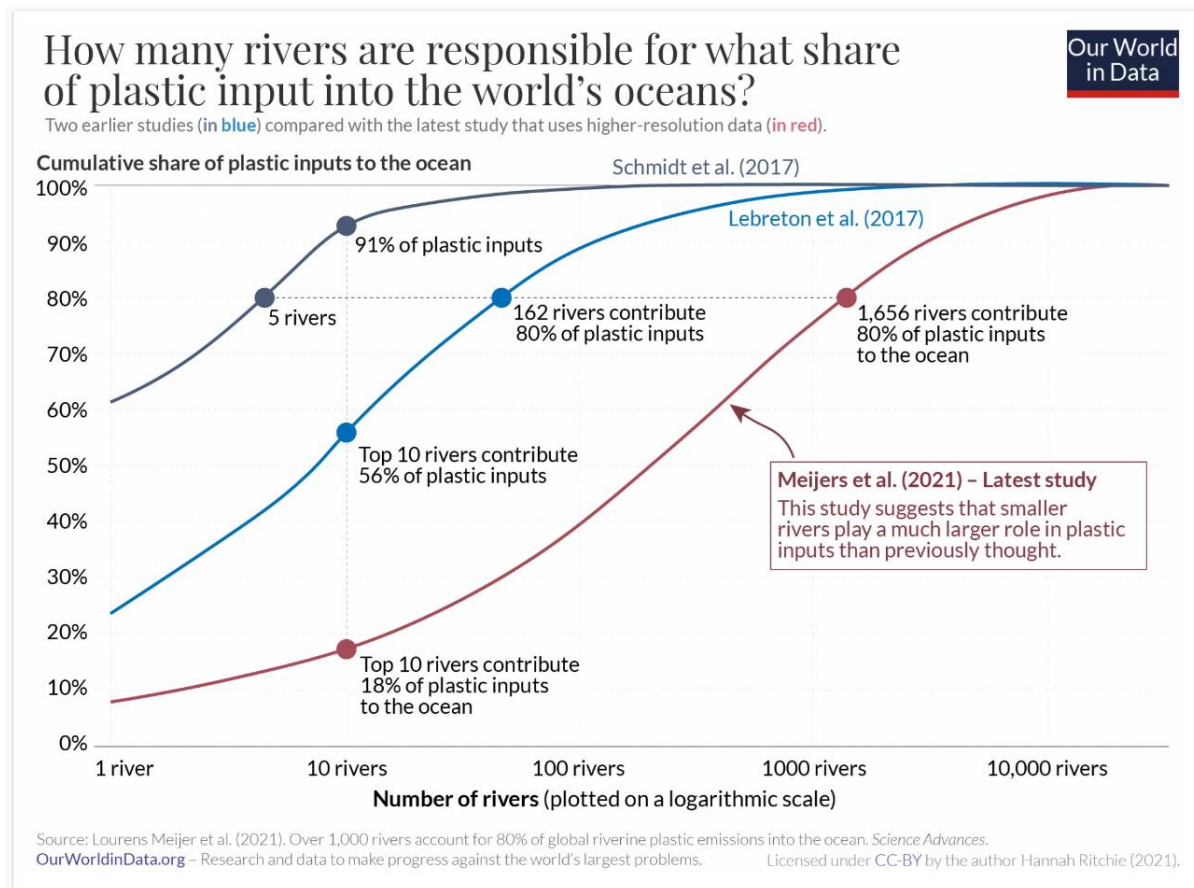
Agregando más variables

- Cada nueva variable que agregamos necesita un nuevo encoding.
- En seaborn la estructura es ligeramente distinta
 - [catplot](#)
- Otras librerías como Plotly/d3 permiten gráficos más complejos e interactivos



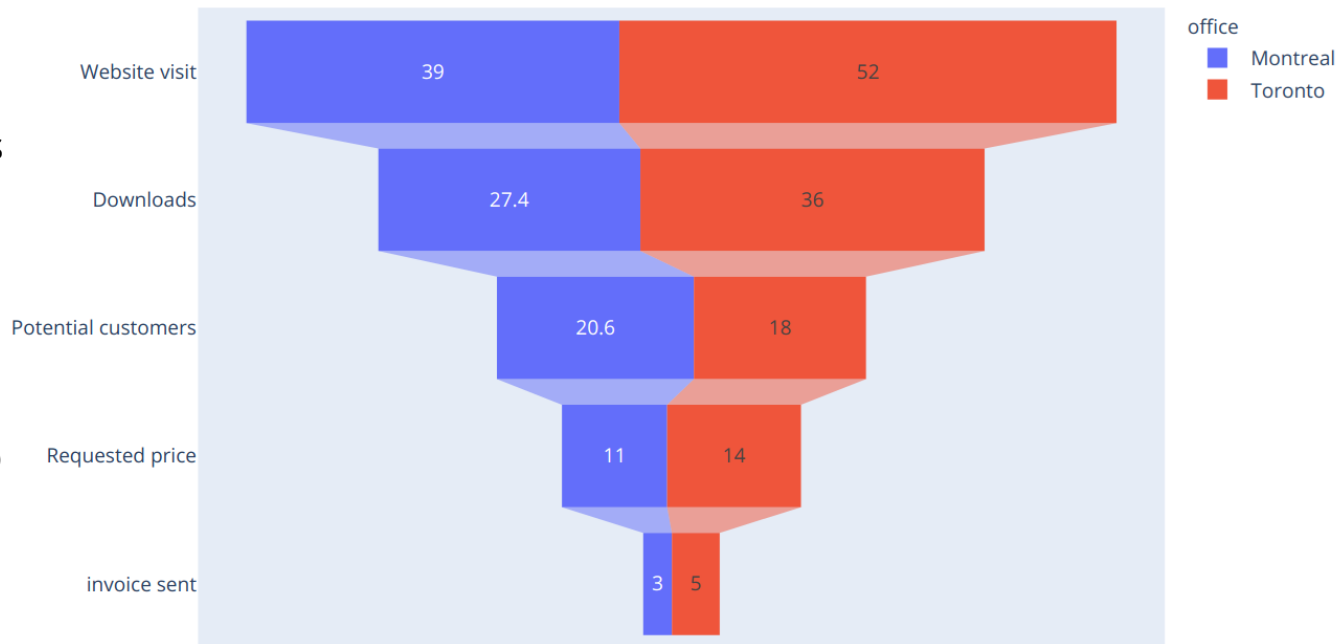
Gráficos más difíciles de leer

- Visualizaciones menos intuitivas pero más informativas
- Transformaciones complejas a los datos



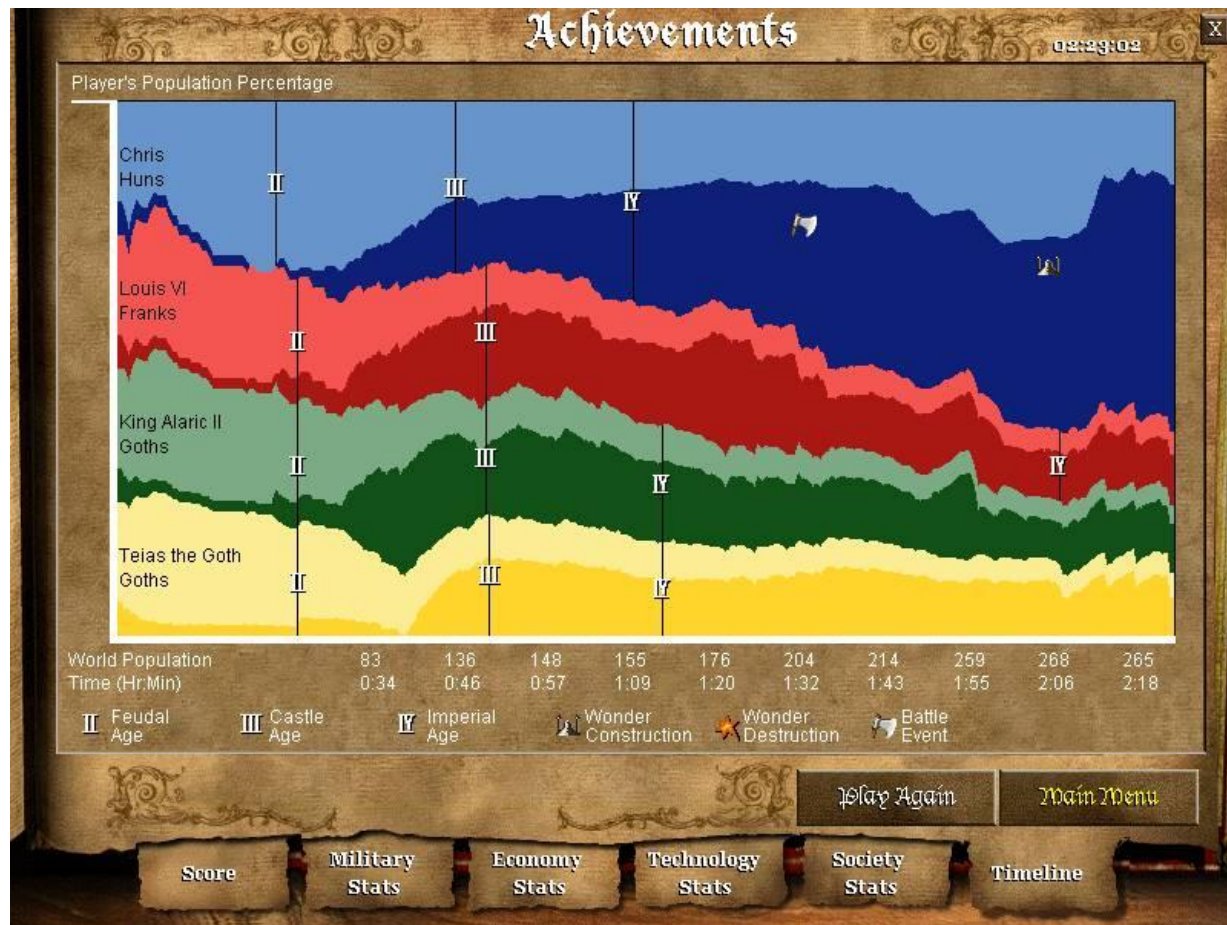
Gráficos más difíciles de leer

- Visualizaciones menos intuitivas pero más informativas
- Transformaciones complejas a los datos
- Tipos de gráficos poco comunes



Gráficos más difíciles de leer

- Visualizaciones menos intuitivas pero más informativas
- Transformaciones complejas a los datos
- Tipos de gráficos poco comunes
- Sólo para expertos de dominio



Buenas prácticas de visualización

para maximizar la efectividad de la comunicación

Básicos a tener en cuenta

- Tamaño del texto
- Visibilidad de elementos contra el fondo
- Método de presentación
 - Proyector: tiene baja resolución. Usar colores brillantes (ni oscuros ni pasteles).
 - Papel: ¿color o blanco y negro? No hay posibilidad de zoom.
 - Pantalla digital: distintas resoluciones y dispositivos. Distintos métodos de interacción.

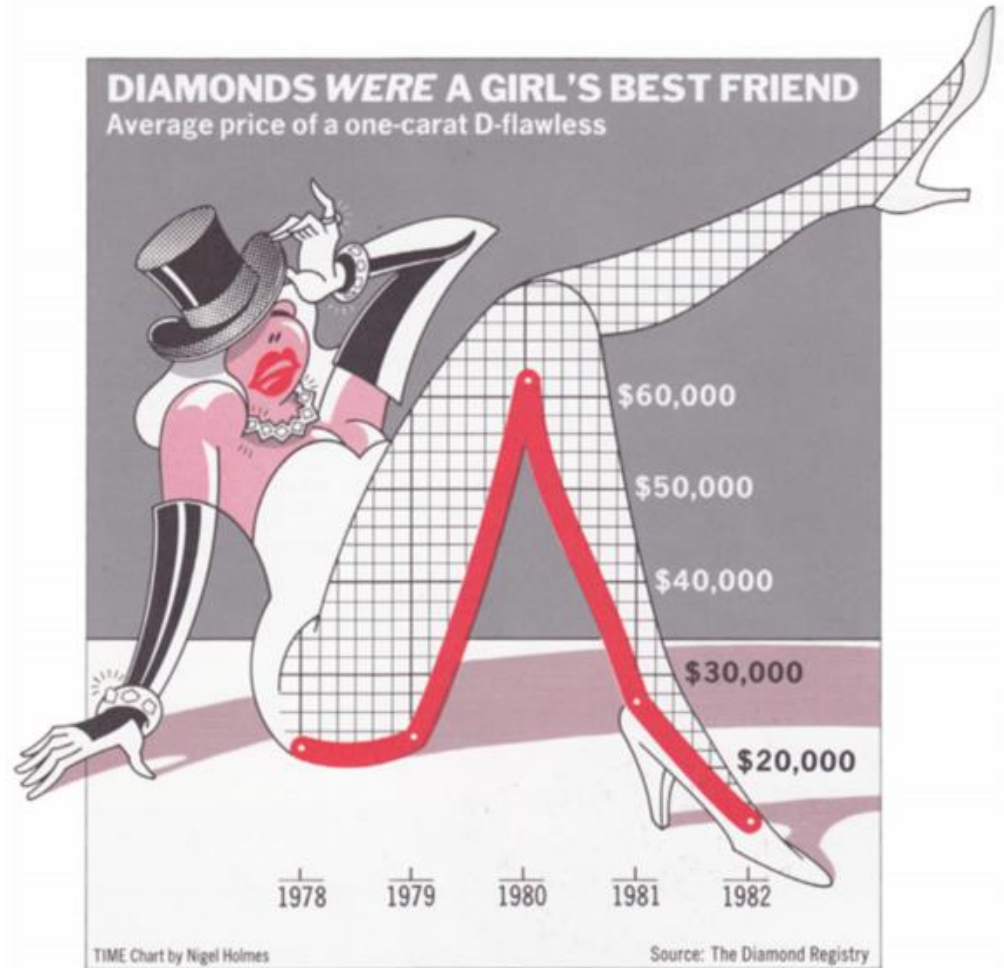
Chartjunk - Less is more

Edward Tufte: maximizar proporción tinta-datos

1. La prioridad es mostrar los datos
2. Maximizar la proporción datos a tinta (toda la “tinta” se usa para datos)
3. Eliminar la tinta no destinada a datos
4. Eliminar la tinta redundante
5. Revisar y editar

Ejemplo

¡Solamente una pequeña
cantidad de elementos
visuales son relevantes!

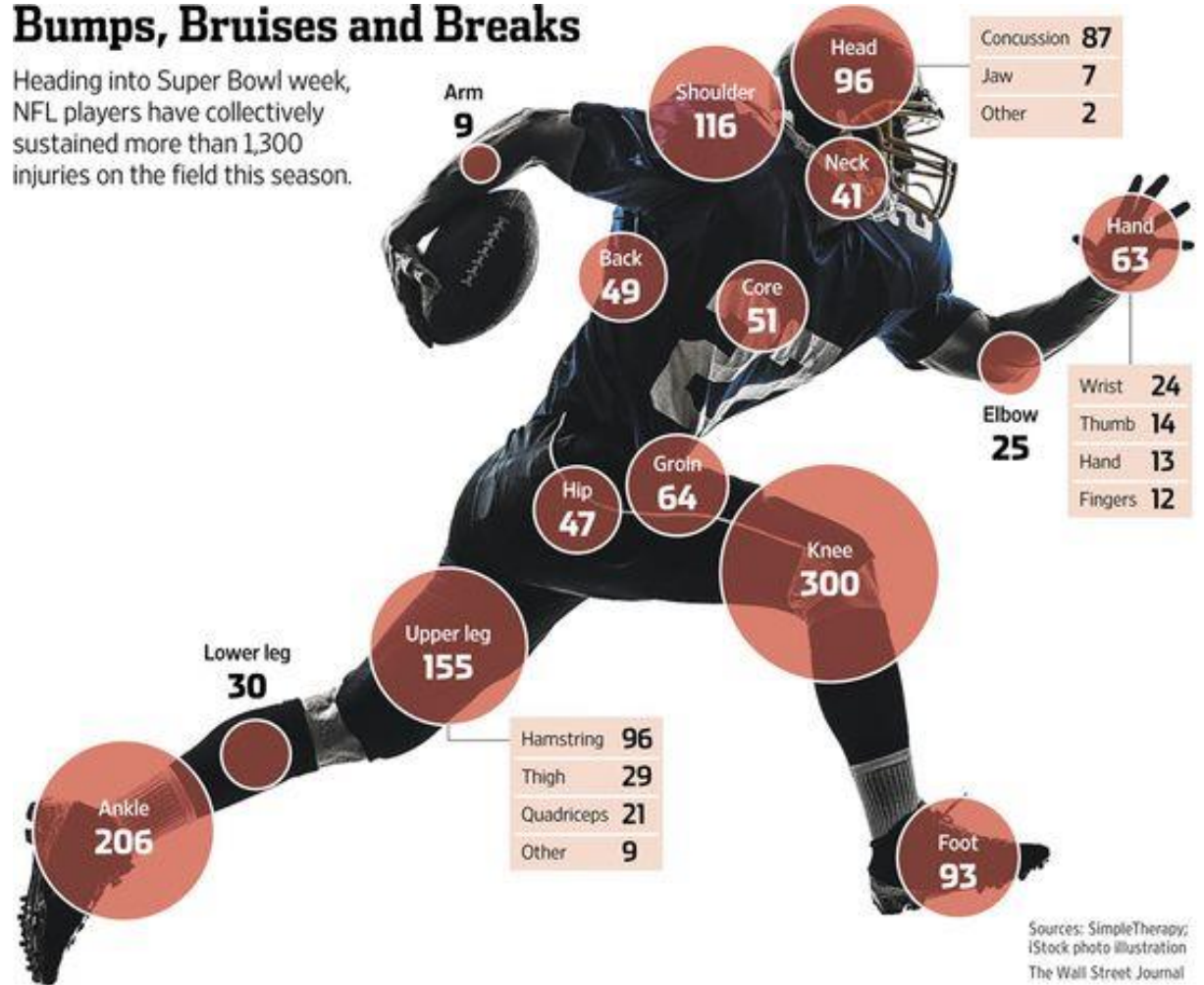


Ejemplo

Todos los elementos
visuales son relevantes con
respecto a la información
transmitida

Bumps, Bruises and Breaks

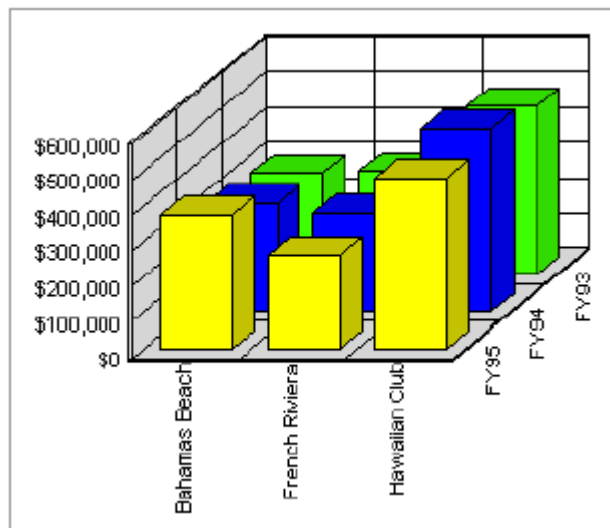
Heading into Super Bowl week, NFL players have collectively sustained more than 1,300 injuries on the field this season.



Sources: SimpleTherapy;
iStock photo illustration
The Wall Street Journal

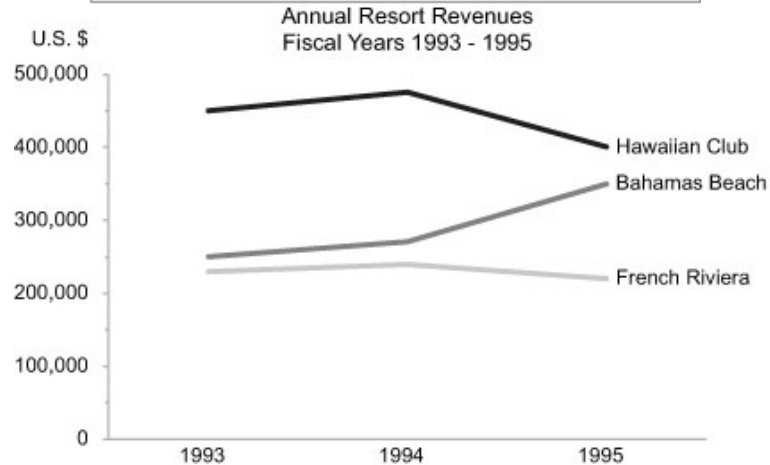
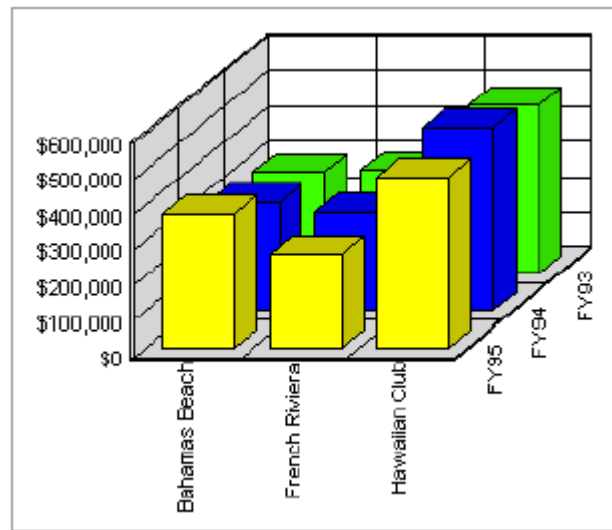
Ejemplo

- Las barras son imposibles de leer
- Las grillas no sirven para nada
- Las etiquetas verticales son difíciles de leer
- El eje z, que representa años, va al revés.



Ejemplo

- Las barras son imposibles de leer
- Las grillas no sirven para nada
- Las etiquetas verticales son difíciles de leer
- El eje z, que representa años, va al revés.



Primero, que quede bien en
blanco y negro

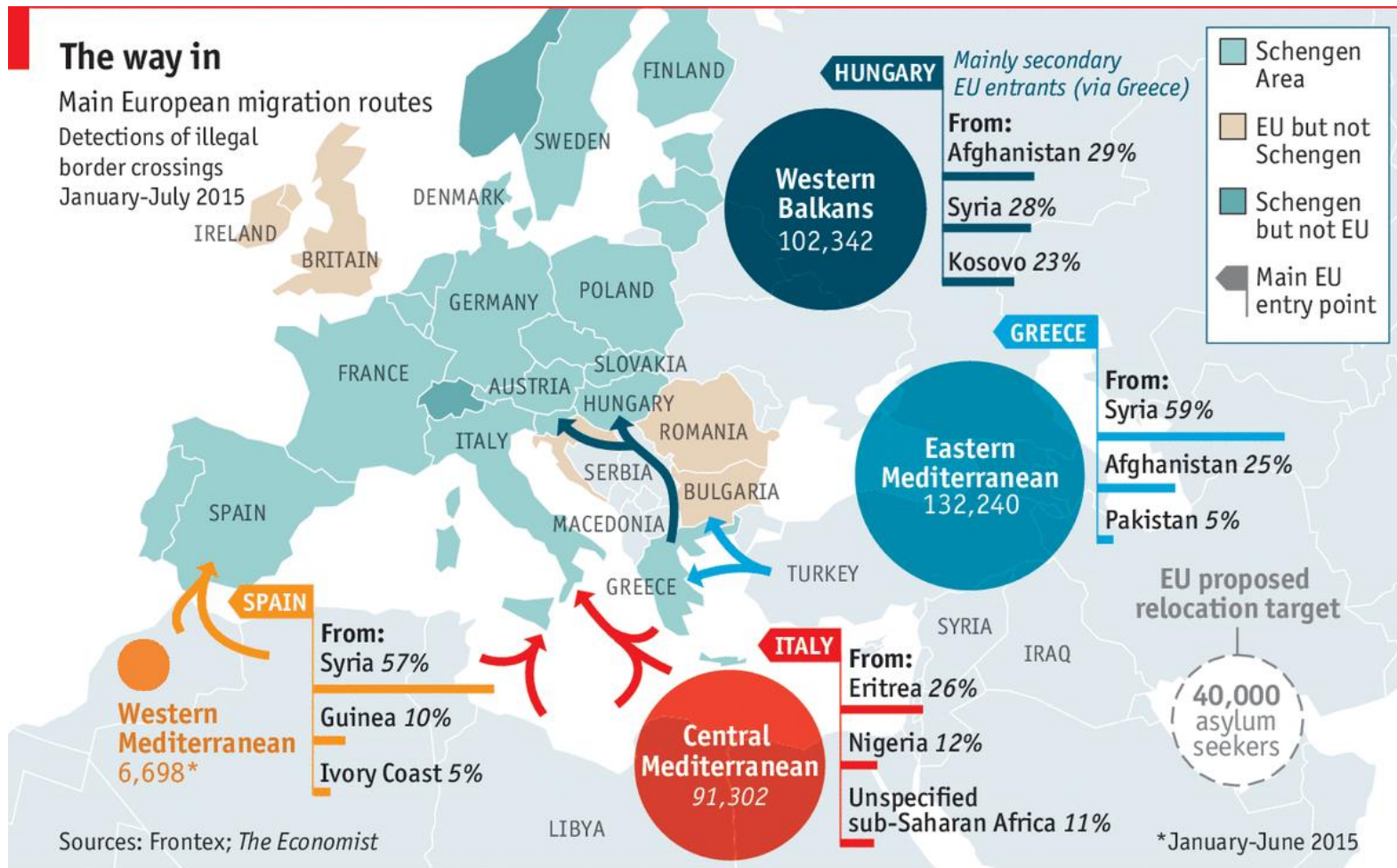
Propiedades de los encodings con color

- No tienen orden
- Nos basamos en convenciones que pueden no reproducirse:
 - Rojo-azul, rosa-celeste para masculino y femenino
 - Celeste para el océano, verde para la pradera, marrón para la montaña
- No todos los percibimos iguales (daltonismo)
- Evocan emociones inconscientes (o subconscientes)

The way in

Main European migration routes

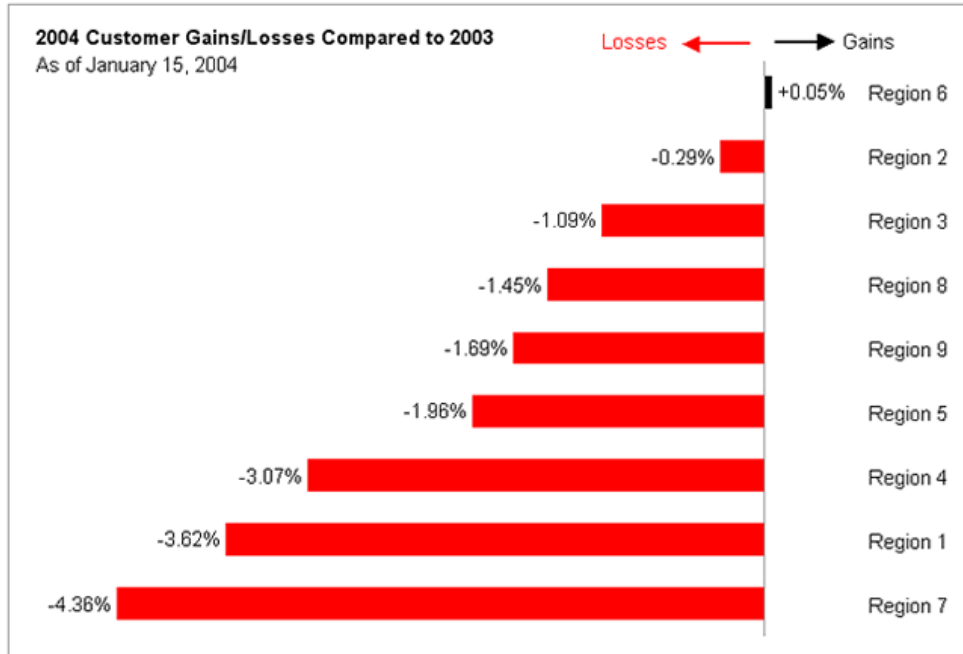
Detections of illegal
border crossings
January-July 2015



Paletas de colores y daltonismo

- Daltonismo:
 - [Coblis](#): pueden subir una imagen y comprobar cómo la vería una persona daltónica.
 - Usar paletas colorblind safe que vienen en las librerías de visualización
- Estética:
 - [Coolors](#): Fácil de usar para conseguir combinaciones rápidas.
 - [Paletton](#): Más completa pero requiere un poco más de conocimiento de colores.

Refuerzo a través de encodings dobles



En este ejemplo, el encoding de color refuerza las pérdidas sin desbalancear el gráfico.

¿Jura decir(se) la verdad, toda la
verdad y nada más que la verdad?

¿Preguntas?