

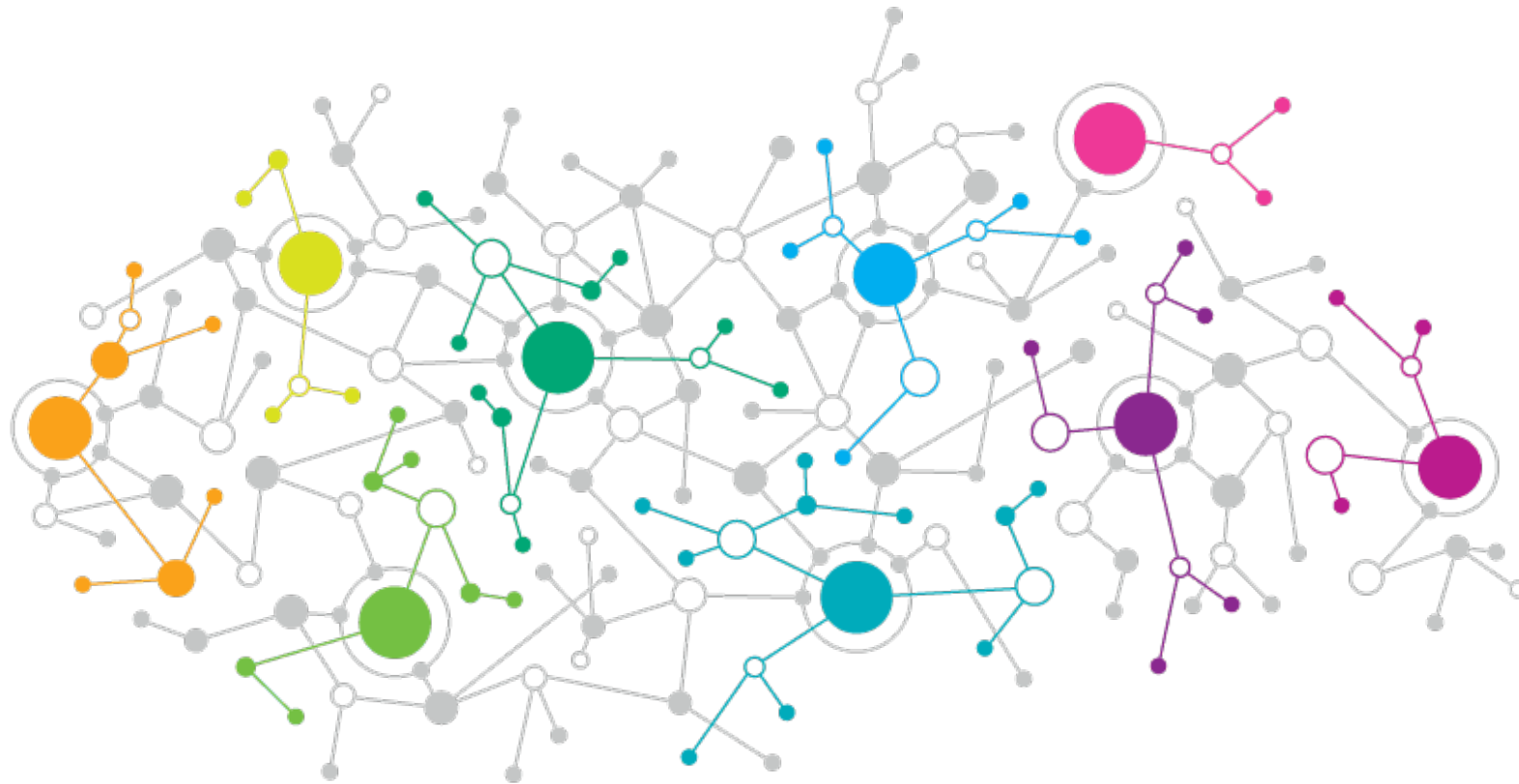
VISUALIZACIÓN DE DATOS

Oscar Rico



¿Qué es la visualización de datos?

Es la representación de datos en un formato gráfico



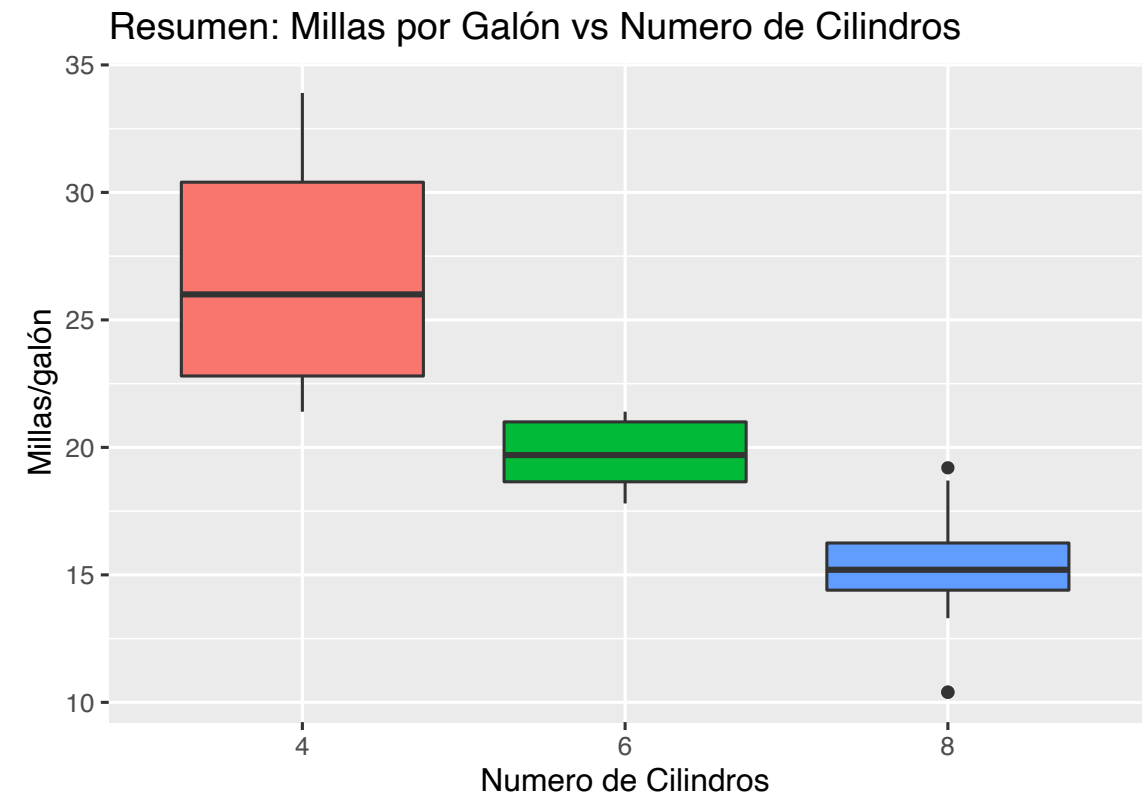
¿Porqué es necesaria la visualización de datos?

- Explorar patrones y tendencias
- Detectar patrones o anomalías
- Darle sentido a una gran cantidad de datos
- Comunicar y compartir información de forma eficiente

¿Porqué es necesaria la visualización de datos?

Una imagen dice más que mil palabras!

```
Console Terminal x Jobs x
~/Desktop/ ↗
> by(mtcars$mpg, mtcars$cyl, summary)
mtcars$cyl: 4
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 21.40  22.80   26.00   26.66  30.40   33.90
-----
mtcars$cyl: 6
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 17.80  18.65   19.70   19.74  21.00   21.40
-----
mtcars$cyl: 8
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 10.40  14.40   15.20   15.10  16.25   19.20
> |
```



Principios para gráficos analíticos

1. Mostrar comparaciones

1. La evidencia de una hipótesis siempre es relativa con otra
2. Siempre preguntarse ¿comparado con qué?

2. Mostrar causalidad/mecanismos/efectos

1. ¿Cuál es la base para la pregunta o hipótesis

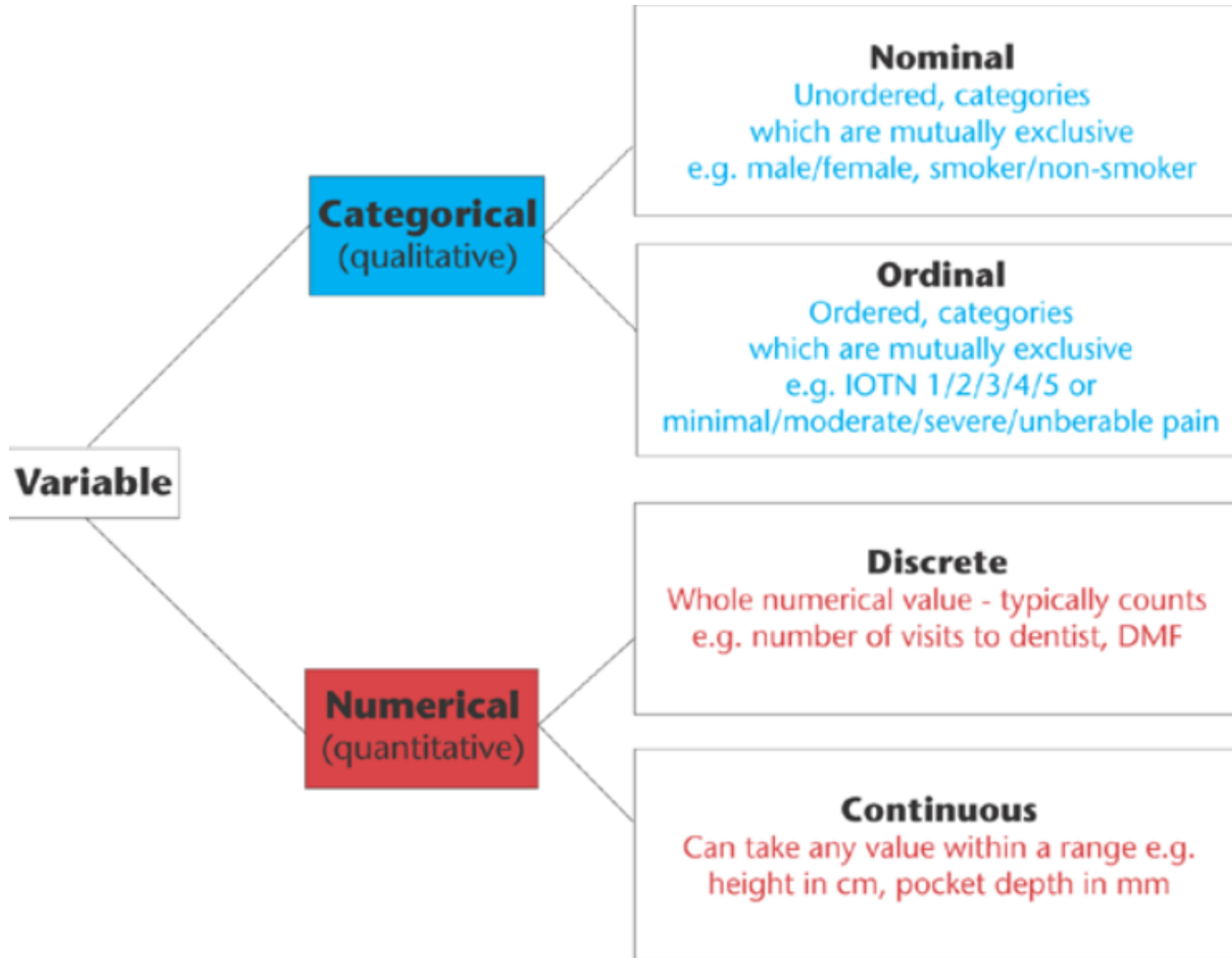
Principios para gráficos analíticos

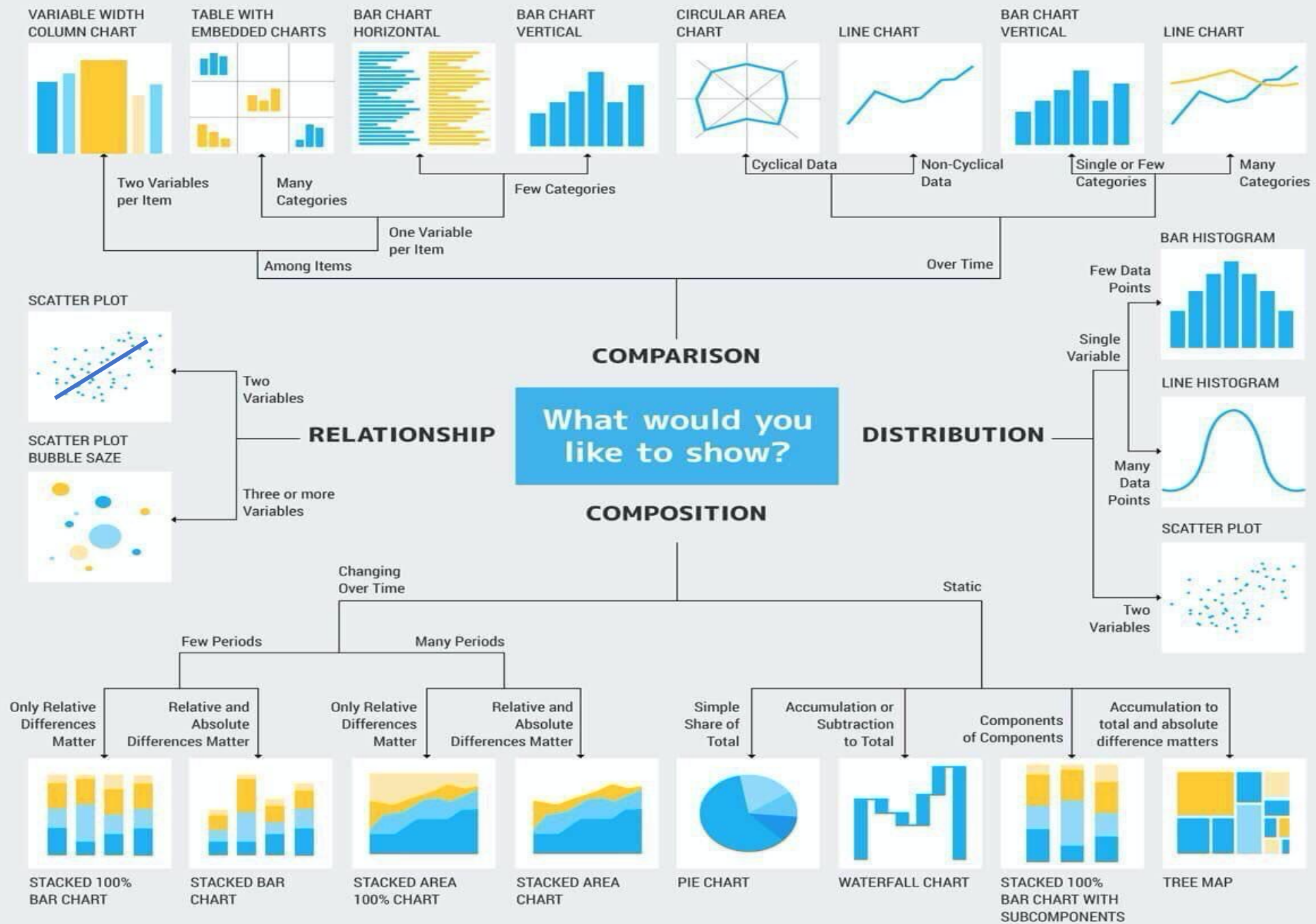
3. Mostrar datos multivariados

- Sistemas complejos

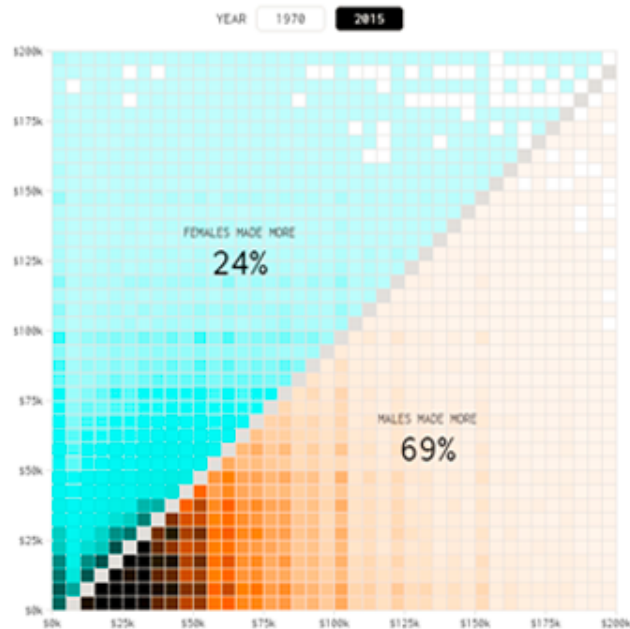
4. Mostrar una integración de evidencia

- Palabras completas, números, imágenes, diagramas
- No dejar que el programa o herramienta determine el análisis

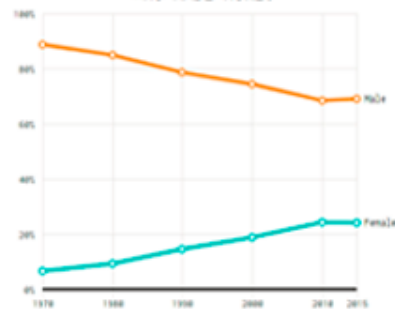




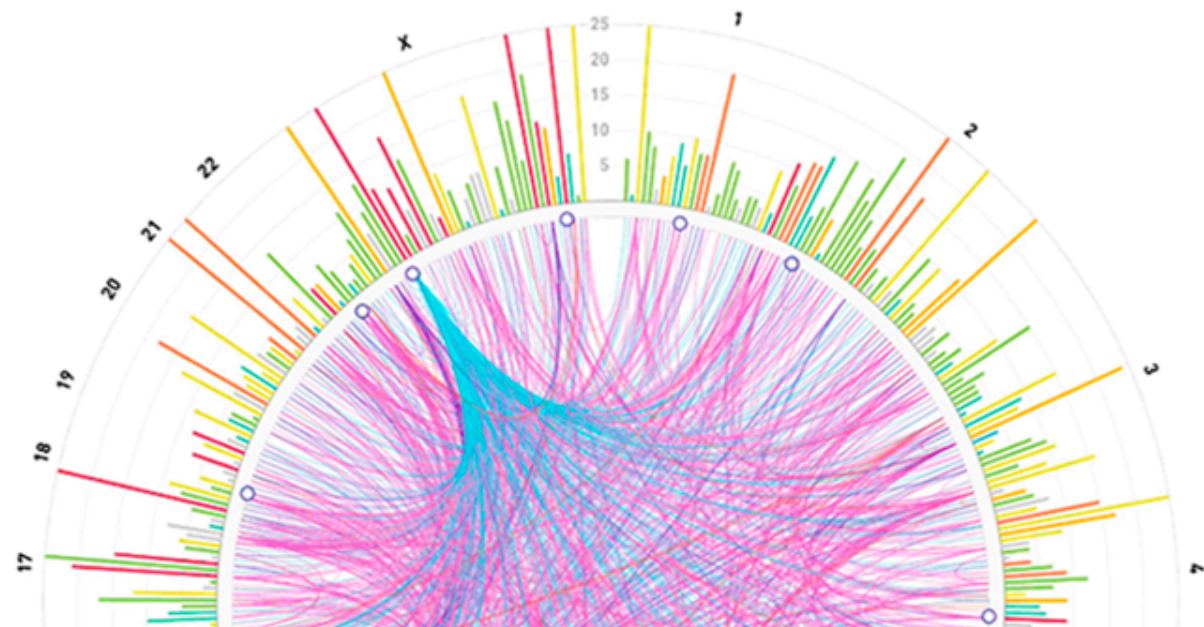
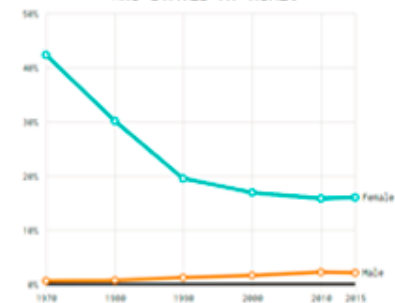
WHICH PARTNER EARNED MORE INCOME?



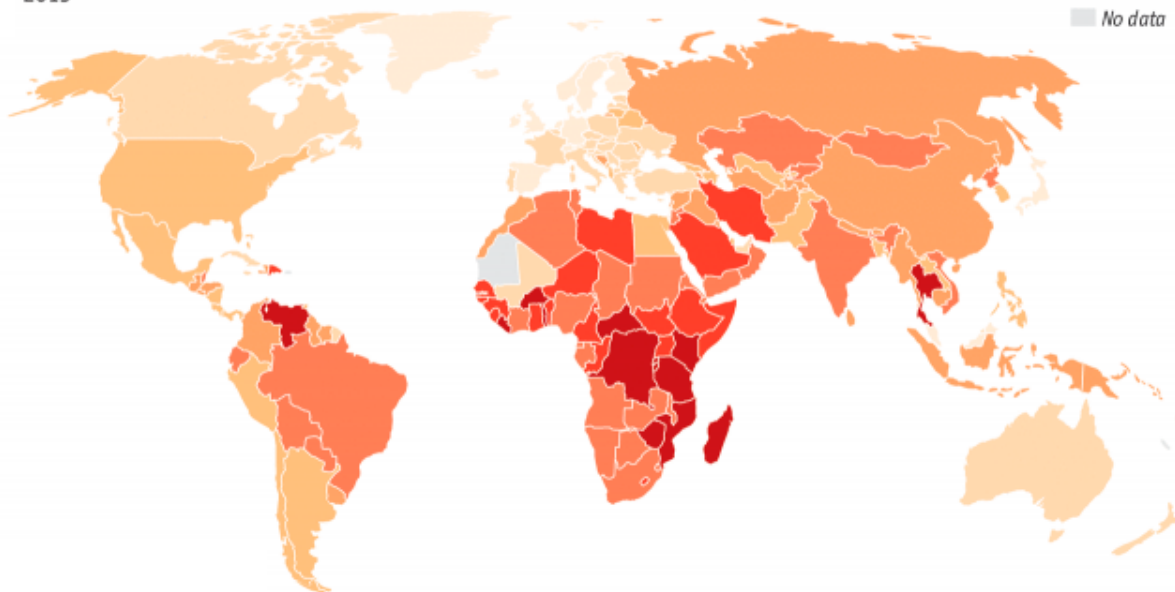
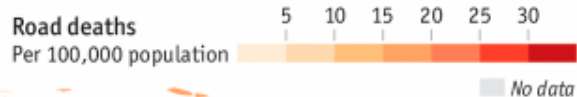
WHO MADE MORE?



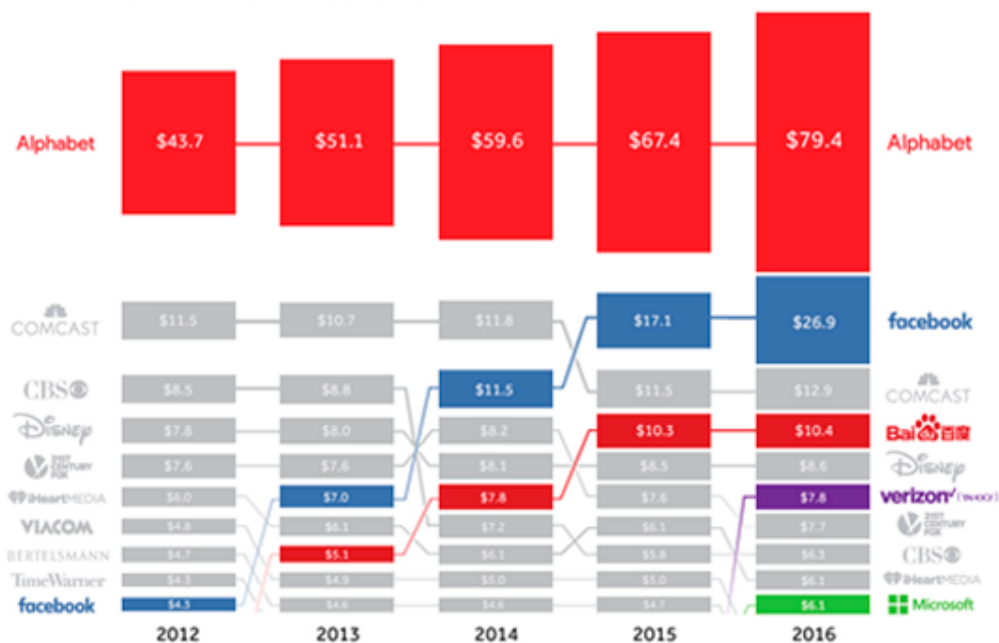
WHO STAYED AT HOME?



Dangerous driving 2015



THE TECH TAKEOVER OF ADVERTISING



Datos LIMPIOS y ORDENADOS

1. Cada columna debe ser una variable
2. Cada fila es una observación
3. Cada celda es un valor

Tidy data

Variables

Observations

usarrests.txt

reporters

Accueil Mise en page Tableaux

614

	A	B	C	D	E
1	State	Murder	Assault	UrbanPop	Rape
2	Alabama	13.2	236	58	21.2
3	Alaska	10	263	48	44.5
4	Arizona	8.1	294	80	31
5	Arkansas	8.8	190	50	19.5
6	California	9	276	91	40.6
7	Colorado	7.9	204	78	38.7
8	Connecticut	3.3	110	77	11.1
9	Delaware	5.9	238	72	15.8
10	Florida	15.4	335	80	31.9
11	Georgia	17.4	211	60	25.8
12	Hawaii	5.3	46	83	20.2
13	Idaho	2.6	120	54	14.2
14	Illinois	10.4	249	83	24
15	Indiana	7.2	113	65	21
16	Iowa	2.2	56	57	11.3
17	Kansas	6	115	66	18
18	Kentucky	0.7	100	63	16.3

usarrests.txt

Mode Normal Prêt

	treatmenta	treatmentb
John Smith	—	2
Jane Doe	16	11
Mary Johnson	3	1

Table 1: Typical presentation dataset.

	John Smith	Jane Doe	Mary Johnson
treatmenta	—	16	3
treatmentb	2	11	1

Table 2: The same data as in Table 1 but structured differently.

person	treatment	result
John Smith	a	—
Jane Doe	a	16
Mary Johnson	a	3
John Smith	b	2
Jane Doe	b	11
Mary Johnson	b	1

Table 3: The same data as in Table 1 but with variables in columns and observations in rows.

People still using excel for data analysis

