
IIC2026

Visualización de Información

— Hernán F. Valdivieso López —
(2022 - 2 / Clase 07)

Temas de la clase - Visualización de datos tabulares

1. Organización de datos tabulares.
 - a. Expresión de valores cuantitativos.
 - b. Definición de regiones categóricas.
 - c. Orientación de ejes.

Organización de datos tabulares

Datos tabulares

- Una tabla, un excel. Cada fila es un **ítem** y cada columna es un **atributo**.

ID	Nombre	Edad	BMI*	Talla polera	Anime preferido
1	Santiago	20	22.3	L	One Piece
2	Felipe	15	19.1	M	SNK
3	Francisca	25	15	M	HxH
...

Datos tabulares

- Una tabla, un excel. Cada fila es un **ítem** y cada columna es un **atributo**.
- **Atributo llave:** valor único que nos permite identificar al ítem.

ID 	Nombre	Edad	BMI*	Talla polera	Anime preferido
1	Santiago	20	22.3	L	One Piece
2	Felipe	15	19.1	M	SNK
3	Francisca	25	15	M	HxH
...

Datos tabulares

- Una tabla, un excel. Cada fila es un **ítem** y cada columna es un **atributo**.
- **Atributo valor**: valores que pueden estar repetidos entre ítems.

ID	Nombre	Edad	BMI*	Talla polera	Anime preferido
1	Santiago	20	22.3	L	One Piece
2	Felipe	15	19.1	M	SNK
3	Francisca	25	15	M	HxH
...

Datos tabulares

- Una tabla, un excel. Cada fila es un **ítem** y cada columna es un **atributo**.
- **Atributo valor categórico**: pueden ser texto o número, pero representan categorías no ordenables.

ID	Nombre	Edad	BMI*	Talla polera	Anime preferido
1	Santiago	20	22.3	L	One Piece
2	Felipe	15	19.1	M	SNK
3	Francisca	25	15	M	HxH
...

Datos tabulares

- Una tabla, un excel. Cada fila es un **ítem** y cada columna es un **atributo**.
- **Atributo valor ordenado**: pueden ser texto o número, pero representan valores ordenables. Dentro tendremos cuantitativos y ordinales (categorías ordenables).

ID	Nombre	Edad	BMI*	Talla polera	Anime preferido
1	Santiago	20	22.3	L	One Piece
2	Felipe	15	19.1	M	SNK
3	Francisca	25	15	M	HxH
...

Organización de datos tabulares

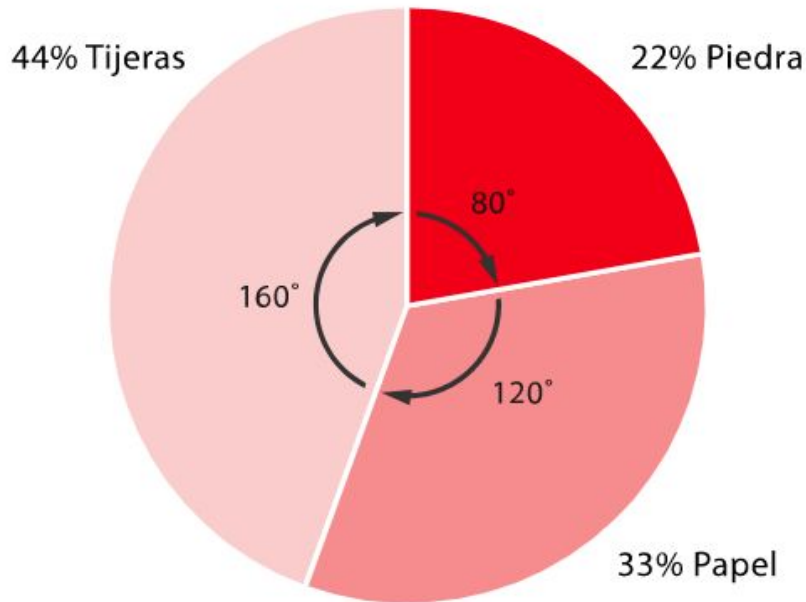
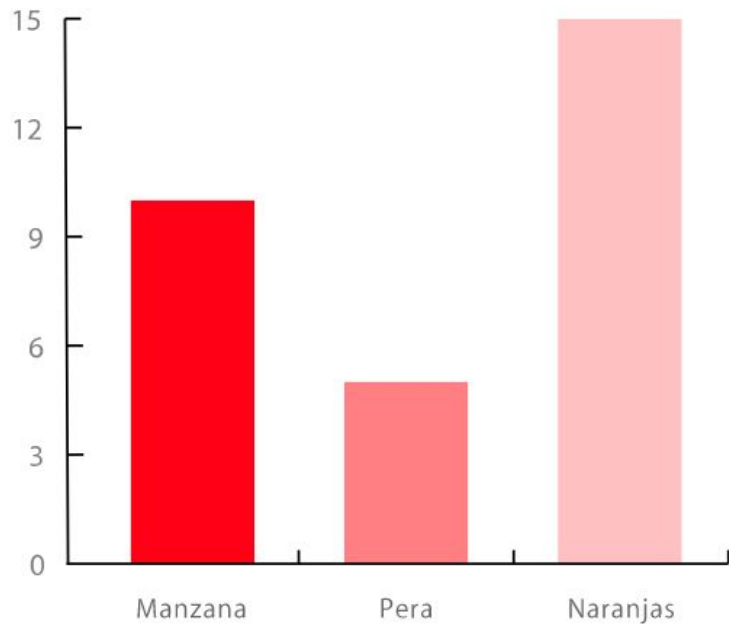
- La organización de datos tabulares corresponde a las decisiones de diseño sobre el **uso del espacio** para codificar datos.
- Son las decisiones de codificación visual más importantes a realizar en una visualización, porque crea el mapa mental del usuario de cómo entenderá un dataset.
- Veremos estas decisiones aplicadas a datos tabulares porque no tienen una representación intrínseca.
 - Datos de red: podemos hacer un grafo o árbol.
 - Datos espaciales: podemos hacer un mapa.
 - Datos tabulares: ¿Hacemos un gráfico de torta? ¿un histograma? ¿Un gráfico de dispersión? ... dependerá mucho de la tarea visual y de los tipos de atributos.

Organización de datos tabulares

- Las decisiones a tomar dependen de diferentes factores:
 - Tarea visual (identificar tendencias, comparar distribuciones, etc).
 - Tipo del atributo valor (categórico, ordenado).
 - La cantidad de ítems en el *dataset*.
 - **Incluso, la cantidad de atributos llaves que se desean visualizar.**

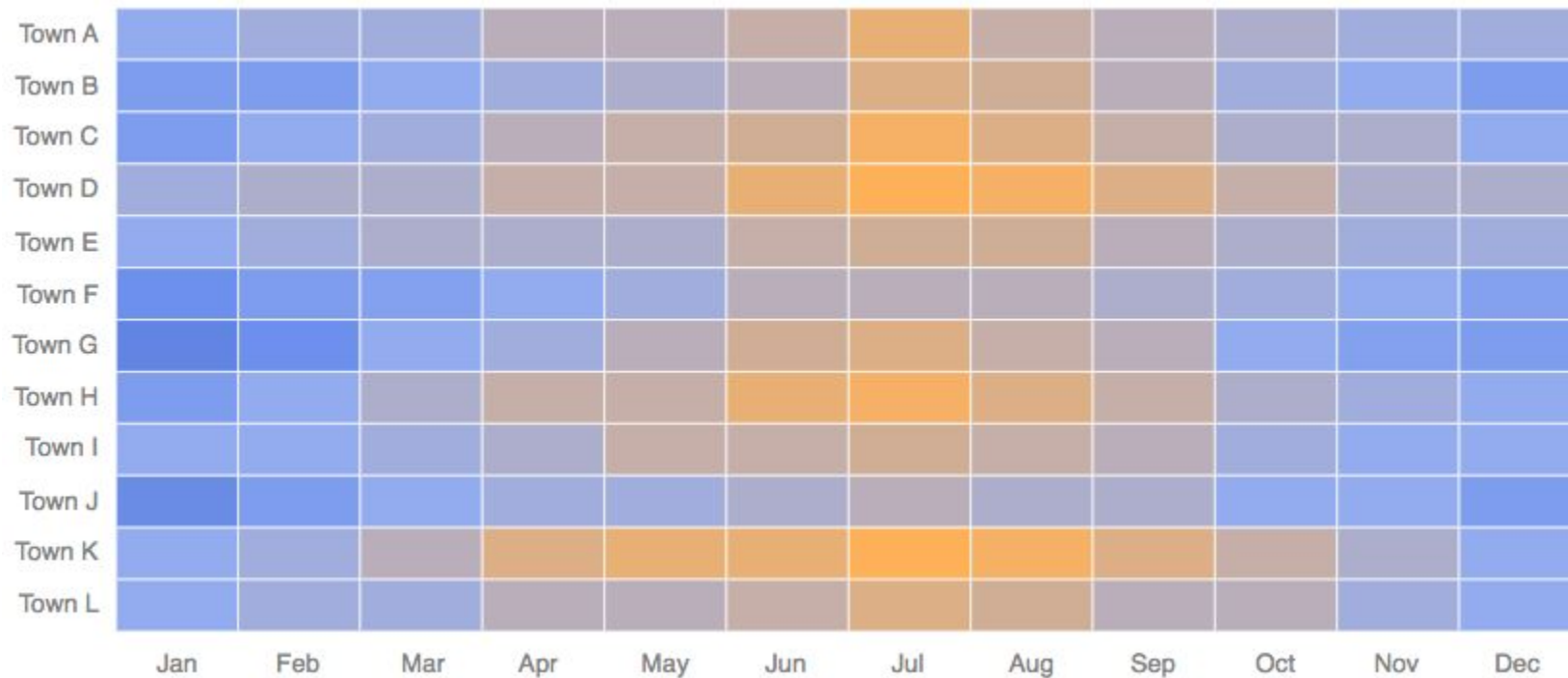
Organización de datos tabulares

- Gráfico de barra o de torta: un atributo llave y un atributo valor.



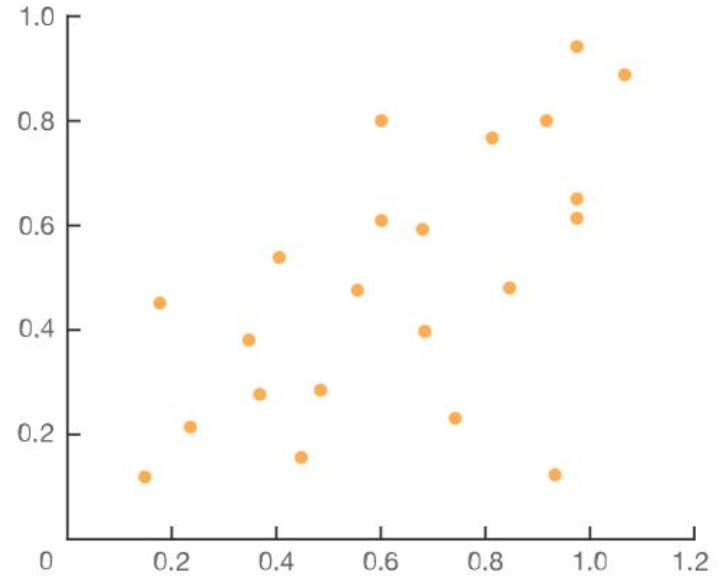
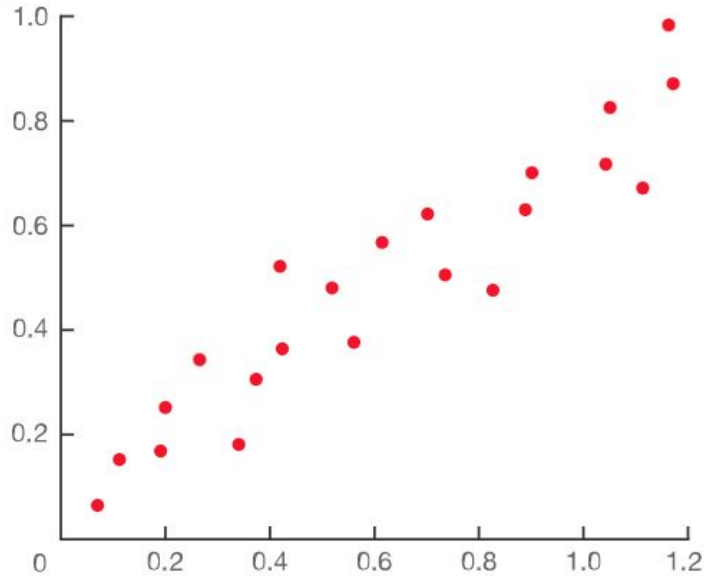
Organización de datos tabulares

- Mapa de calor: Cuando tenemos dos atributos llaves y un atributo valor.



Organización de datos tabulares

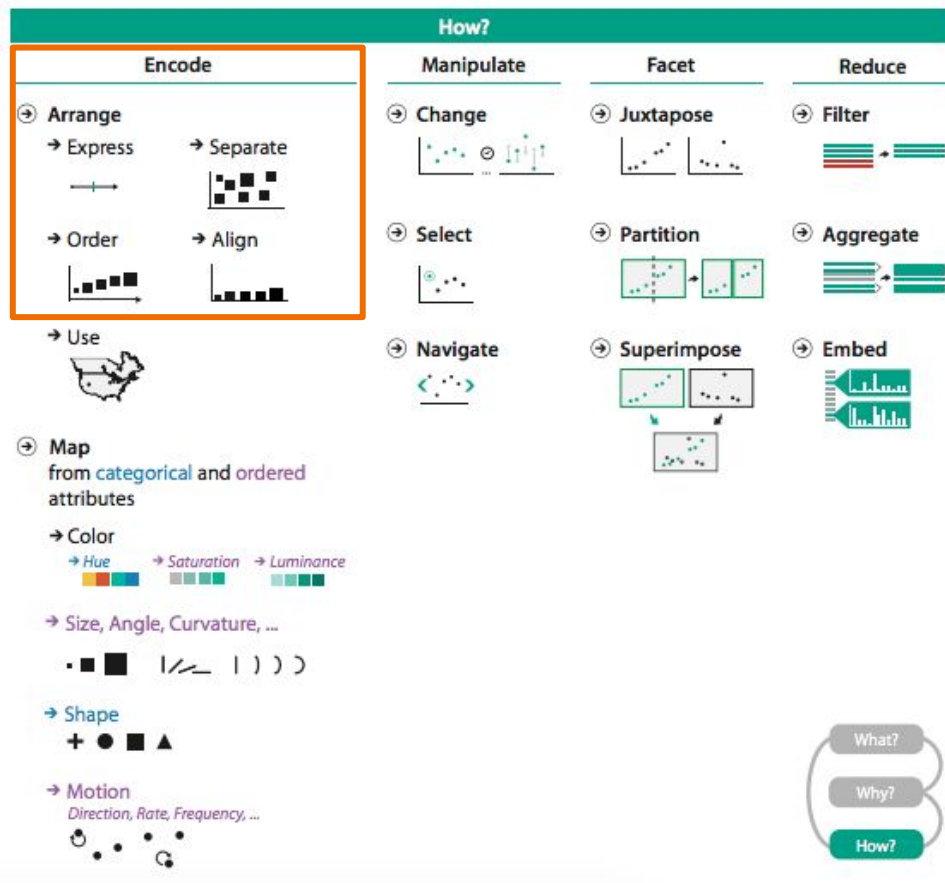
- Gráfico de dispersión: Cuando no tenemos atributos llaves y tenemos dos atributos valor cuantitativos.



Organización de datos tabulares

- En general, las decisiones para el uso de espacio en datos tabulares recaen en 3 sub-decisiones:
 - Expresión de valores cuantitativos.
 - Definición de regiones categóricas.
 - Orientación de ejes.

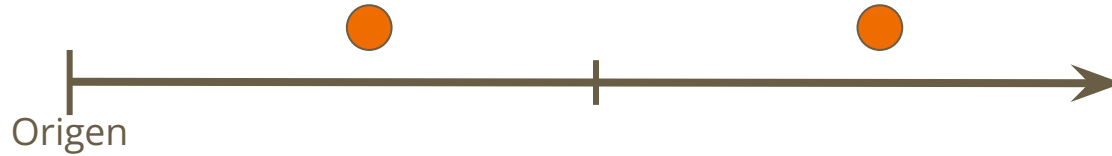
Organización de datos tabulares



Expresión de valores cuantitativos

Expresión de valores cuantitativos

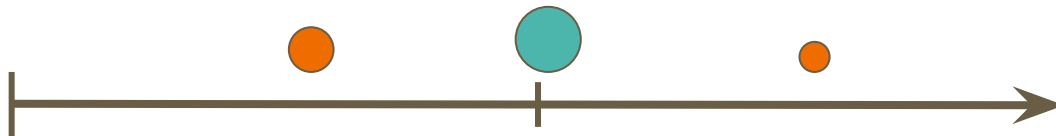
- Conocido en el modelo de Munzner con el término de *express*.
- Expresar el valor de un dato cuantitativo mediante la posición en un eje.



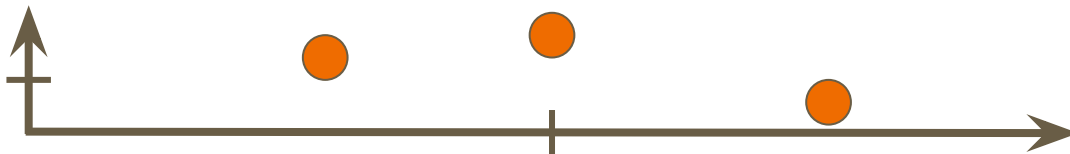
- Generalmente se interpreta que más lejos del origen es un valor más alto.

Expresión de valores cuantitativos

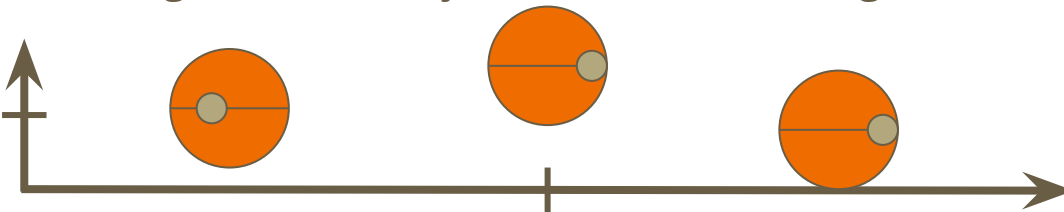
- Más atributos pueden codificarse al mismo tiempo utilizando:
 - Propiedades visuales no espaciales como el **color** o **tamaño**.



- Añadir más ejes.



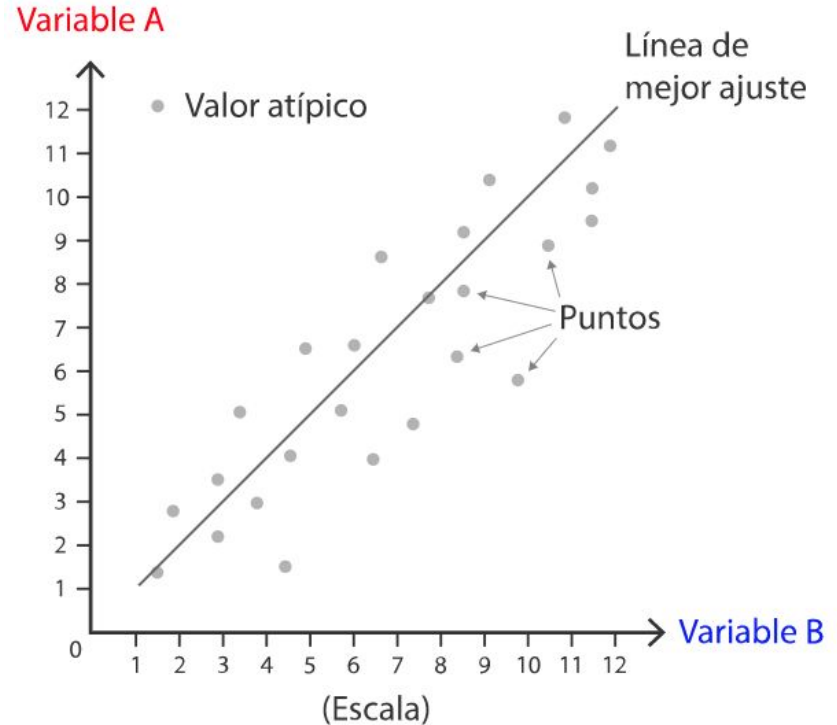
- Utilizar "glifos" (un conjunto de diferentes figuras).



Expresión de valores cuantitativos

Gráfico de dispersión o *scatterplot*

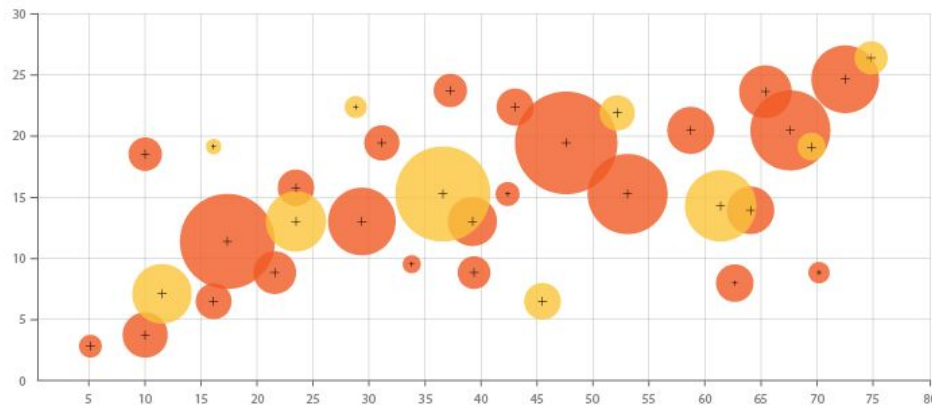
- Codifican dos atributos mediante posición horizontal y vertical de forma simultánea. Se utilizan puntos para representar la información,
- Podemos utilizar el color para codificar un tercer atributo.



Expresión de valores cuantitativos

Gráfico de dispersión o *scatterplot*

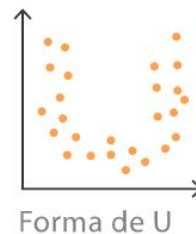
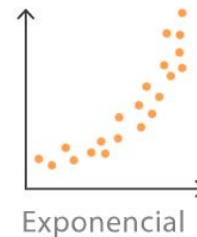
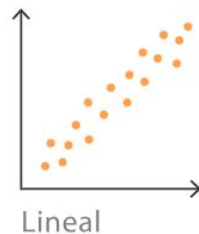
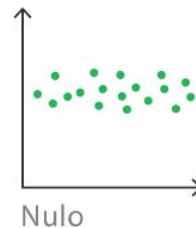
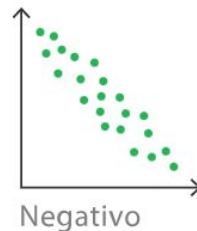
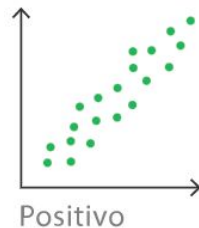
- Codifican dos atributos mediante posición horizontal y vertical de forma simultánea. Se utilizan puntos para representar la información,
- Podemos utilizar el color para codificar un tercer atributo.
- Podemos incluir el tamaño para codificar un cuarto atributo. Aunque ahí estamos presente ante un **gráfico de burbuja o *bubblechart***.



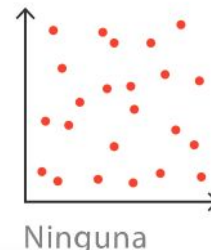
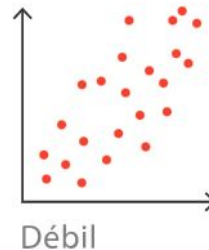
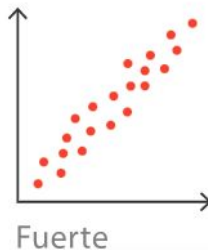
Expresión de valores cuantitativos

Gráfico de dispersión o *scatterplot*

- Ofrecer un resumen de los datos.
- Caracterizar distribuciones.
- Encontrar datos atípicos.
- Encontrar valores extremos.
- Identificar correlación entre dos atributos.



Fuerza de Correlación



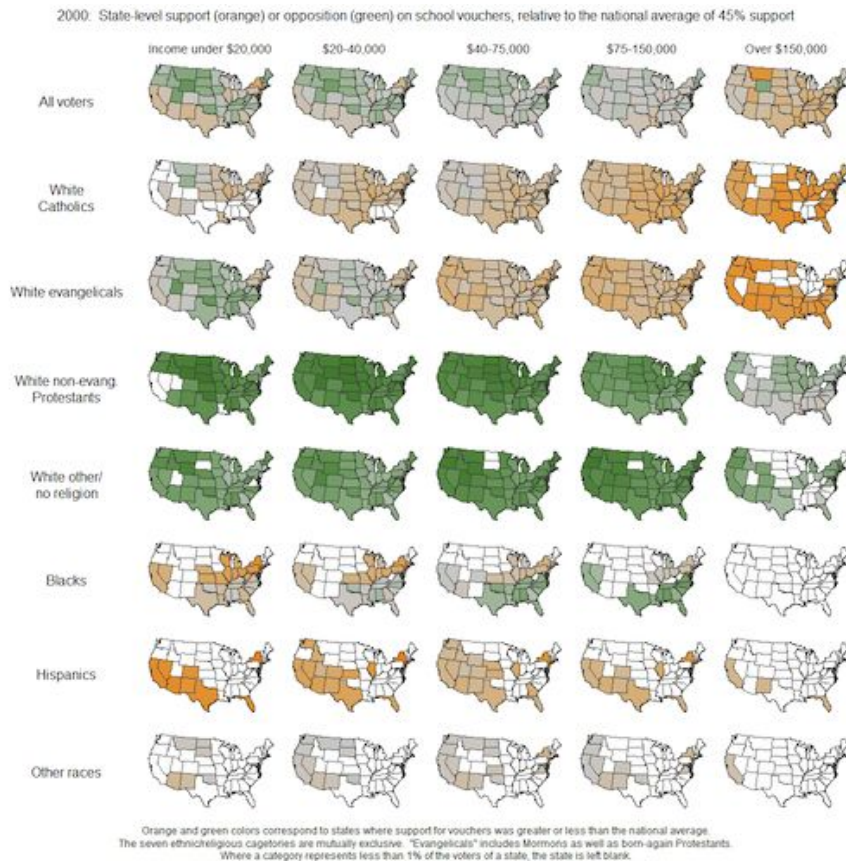
Definición de regiones categóricas

Definición de regiones categóricas

- Acabamos de ver que la posición en un eje tiene semántica ordenada y los atributos categóricos no tiene.
- 🤔 **¿Cómo se utiliza el espacio para datos categóricos?**
 - Si generamos áreas distinguibles, con límites y contiguas, las posiciones si se logran captar la naturaleza de los datos categóricos 😊.
 - Dibujar todos los ítems con el mismo valor para un atributo categórico en la misma región utiliza proximidad espacial para representar su similaridad.



Definición de regiones categóricas



Definición de regiones categóricas

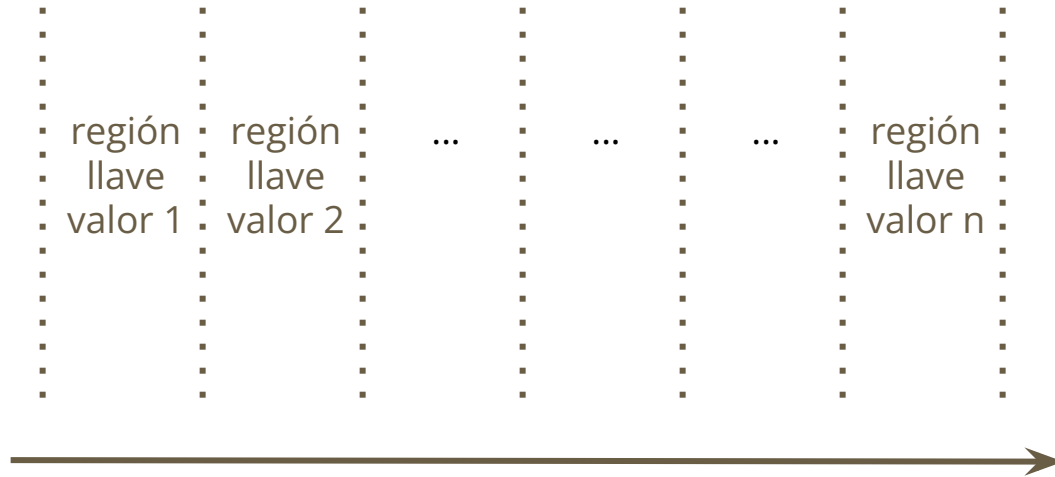
Dentro de las decisiones para definir una región categórica se encuentran:

- Separar regiones (cuantas filas y columnas) (debe darse de alguna).
- Alinear regiones (¿hay un eje común?) (opcional).
- Ordenar regiones (debe darse de alguna).

Definición de regiones categóricas

Cuando tenemos 1 atributo llave o categórico

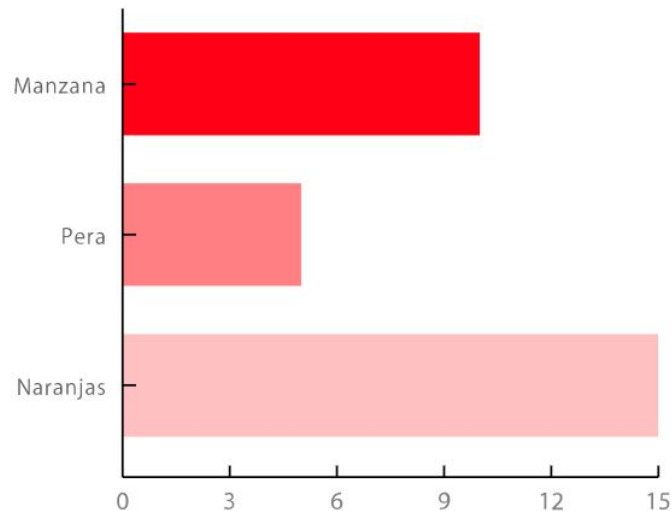
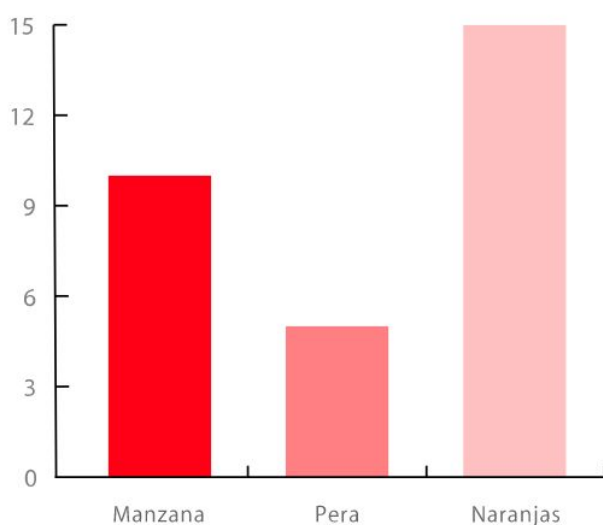
- Generalmente se organizan como una lista unidimensional vertical u horizontal. En la dirección que avanza se dividen distintas regiones, y en su dirección perpendicular se expande la región de la llave.



Definición de regiones categóricas

Gráfico de barra o *bar chart*.

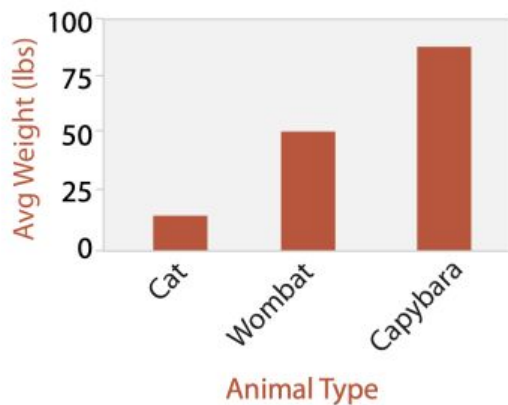
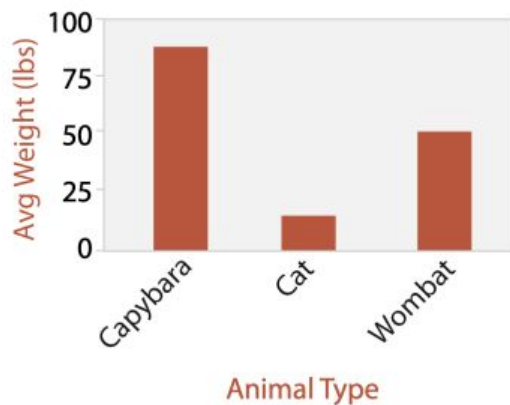
- En cada región se incluye una marca de línea, la cual se alinean bajo un eje y su largo codifica al atributo numérico.



Definición de regiones categóricas

Gráfico de barra o *bar chart*.

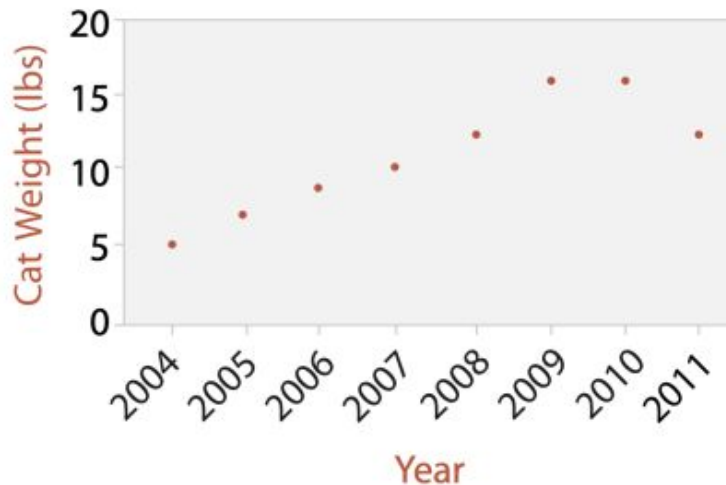
- 🤔 ¿Y el orden de barras? → Depende del tipo de atributo.
 - Ordenados: usamos su orden intrínseco.
 - Categórico: (a) arbitrario, (b) orden alfabético, lo cual tiene la ventaja de hacer más fácil la búsqueda de ítems, o (c) según el atributo cuantitativo codificado.



Definición de regiones categóricas

Gráfico de puntos o *dot chart*.

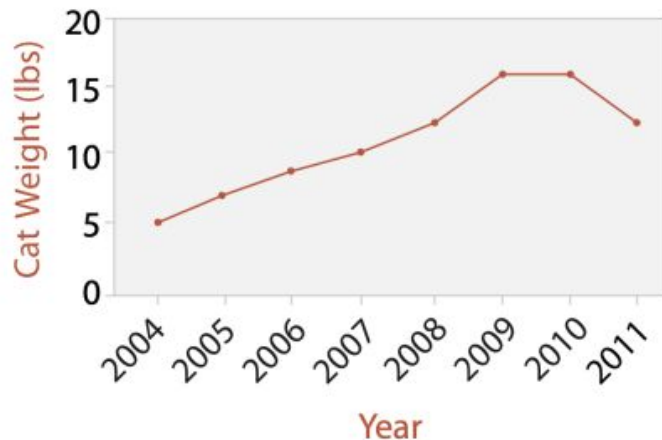
- En cada región se incluye un punto, la cual se alinean bajo un eje y su posición codifica al atributo numérico.



Definición de regiones categóricas

Gráfico de línea o *line chart*.

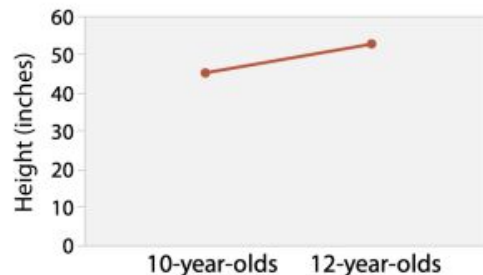
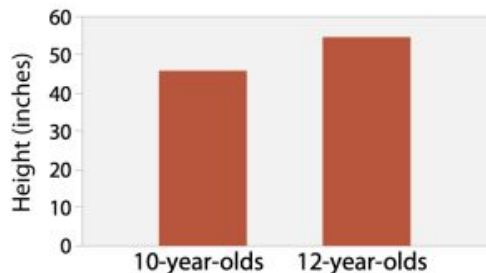
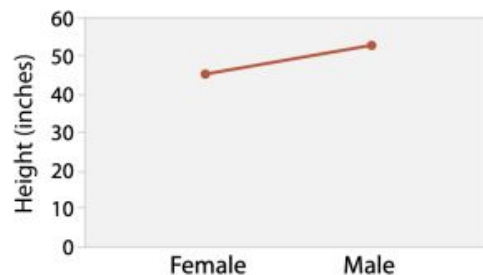
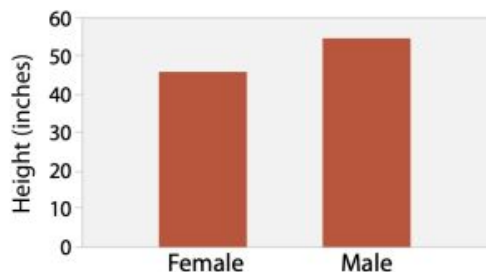
- Gráfico de punto donde se incluye una línea que une los puntos. Se enfatiza la tendencia de los puntos y permite una comparación entre valores vecinos.
- El uso de línea provoca una interpretación de orden en los puntos. Por este motivo, un gráfico de línea **debe usar una llave ordenada**.



Definición de regiones categóricas

Gráfico de línea o *line chart*.

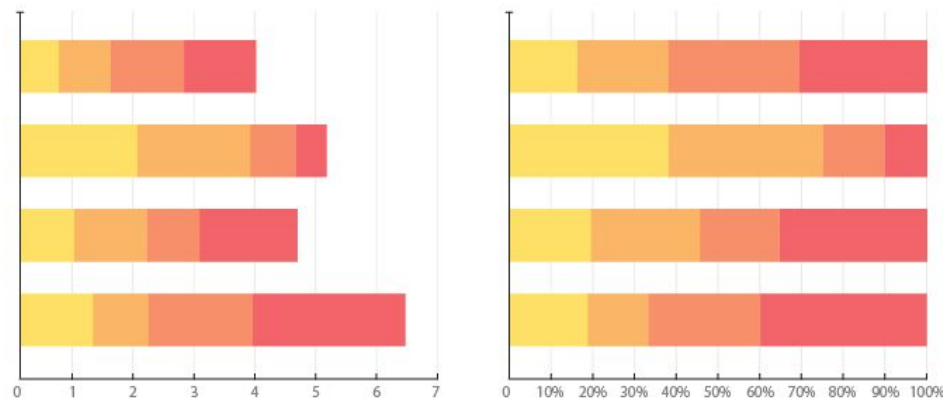
- El uso de línea provoca una interpretación de orden en los puntos. Por este motivo, un gráfico de línea **debe usar una llave ordenada**.



Definición de regiones categóricas

🤔 ¿Si tengo más atributos que funcionan, en conjunto, como llaves?

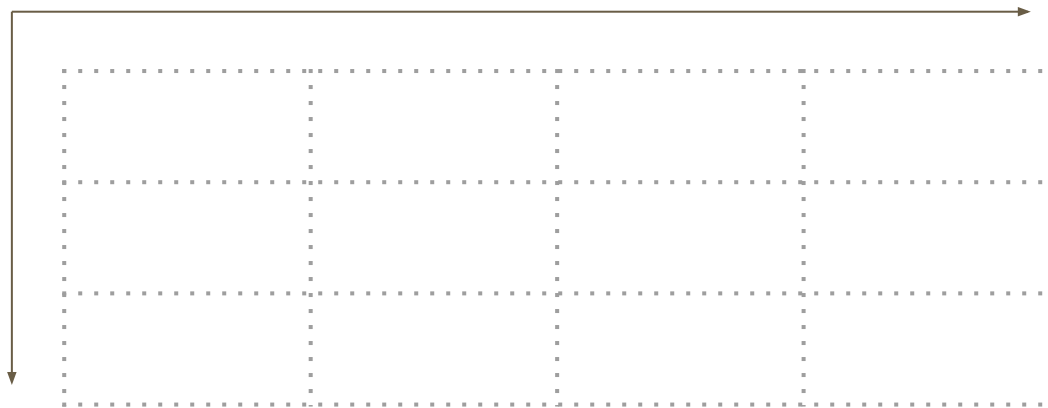
- Gráfico de barras apiladas o *stacked bar chart*. Cada región contiene una barra grande compuesta por sub-barras.
 - El largo completo sigue codificando un valor, pero las sub barras permiten ver sub-componentes de ese valor.
 - Se pueden alinear las barras a ambos extremos para ver la proporción de cada sub-componente.



Definición de regiones categóricas

🤔 ¿Si hay más atributos que funcionan, en conjunto, como llaves?

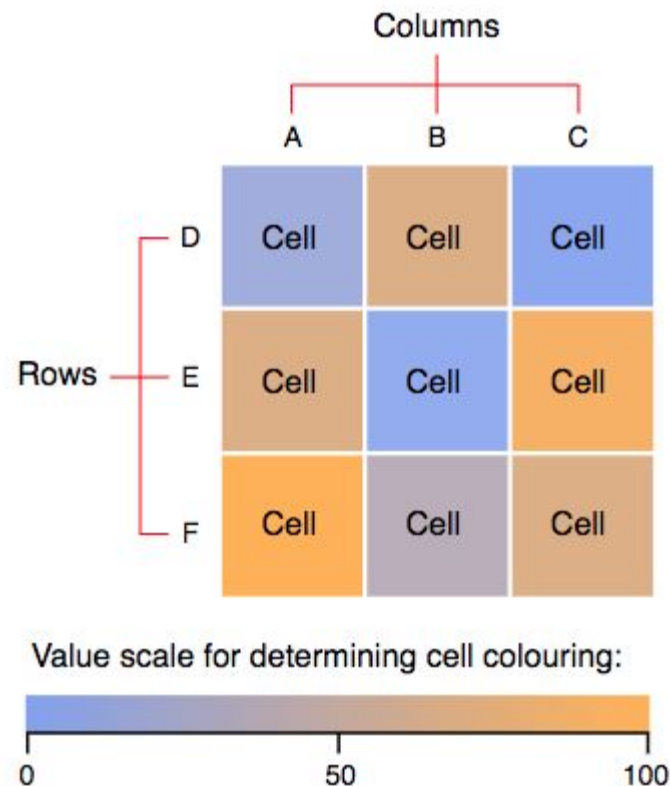
- Otra opción es recurrir a la **separación matricial**.
- Uno de los atributos se despliega en las filas y otro en las columnas, de forma que las celdas que resultan como combinaciones de valores son las regiones para mostrar o codificar información.



Definición de regiones categóricas

Mapa de calor o *heatmap*

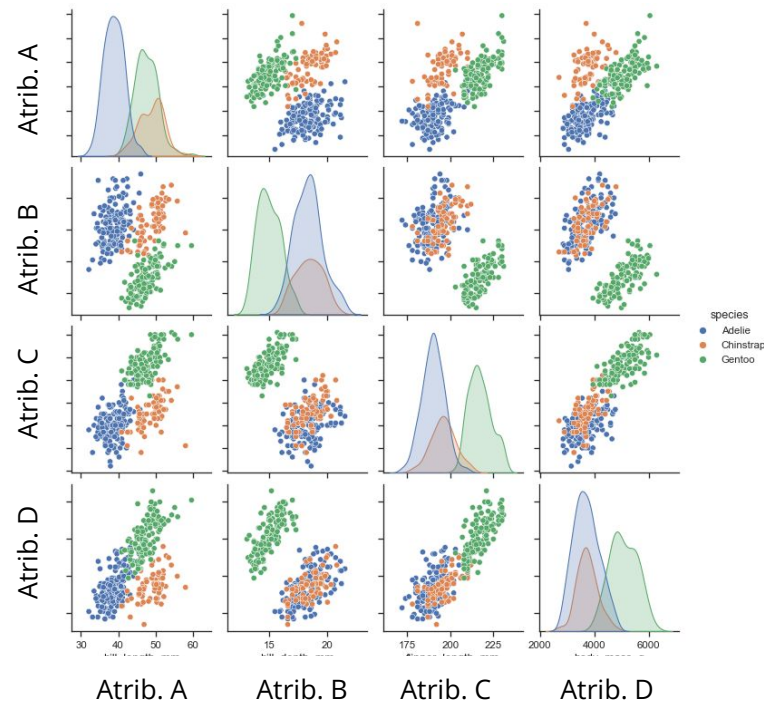
- Cada fila es un valor distinto de un atributo llave.
- Cada columna es un valor distinto del otro atributo llave.
- Cada celda es el atributo valor correspondiente a las llaves (fila, columna). Su color codifica el valor del atributo.



Definición de regiones categóricas

Matriz de gráficos de dispersión o *Scatterplot matrix*

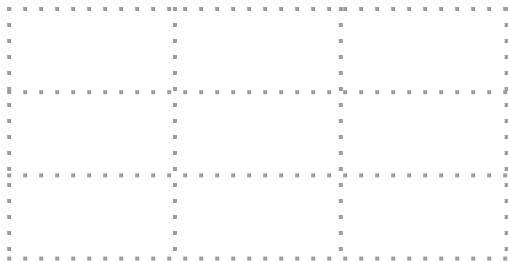
- La matriz se forma por combinaciones de atributos del *dataset*.
- Cada celda presenta un gráfico de dispersión de todos los datos donde los ejes están definidos según posición de la celda.
- La diagonal se suele omitir porque sería la combinación identidad (atributo A VS atributo A). A veces se suele incluir un histograma en la diagonal.



Definición de regiones categóricas

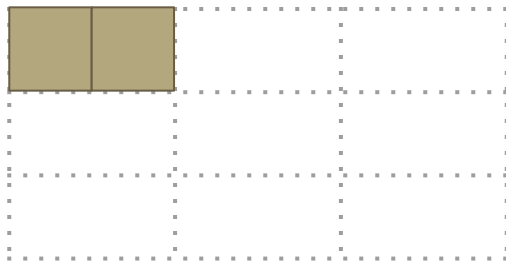
🤔 ¿Existen más formas de definir regiones categóricas?

- Si, una que abordaremos más adelante: Subdivisiones recursivas.
- Cada región definida inicialmente se vuelve a dividir en sub-regiones y así sucesivamente.



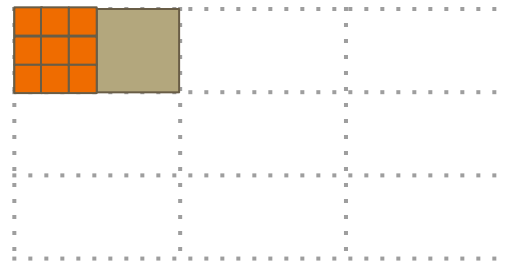
División inicial

Cada región muestra los resultados de encuestas por país.



Segunda división

Para cada país se segmentan los resultados entre fumadores y no fumadores.



Tercera división

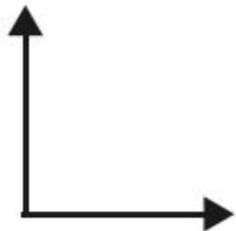
Para cada grupo de personas se segmentan los resultados por rango de edad.

Orientación de ejes

Orientación de ejes

- Al inicio de la clase mencionamos que uno puede **expresar** valores a través de un eje, pero la pregunta que nos faltó definir es: **¿Cómo se van a orientar los ejes?**
 - Rectilíneo (*rectilinear*)
 - Paralelo (*parallel*)
 - Radial (*radial*)

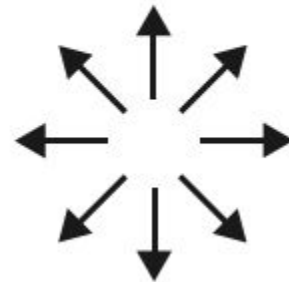
→ Rectilinear



→ Parallel

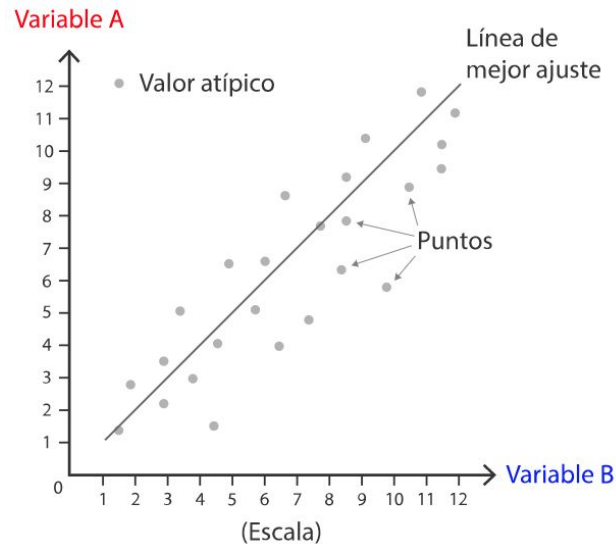
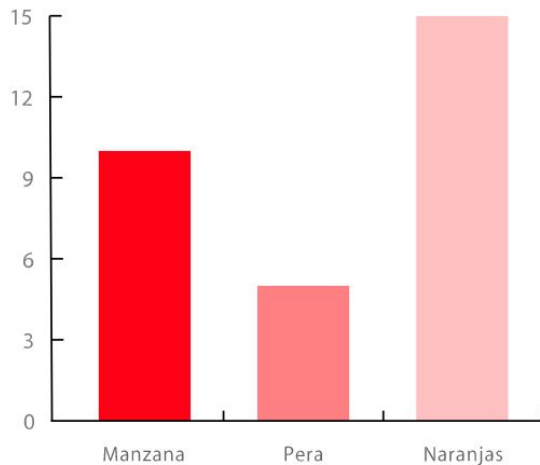


→ Radial



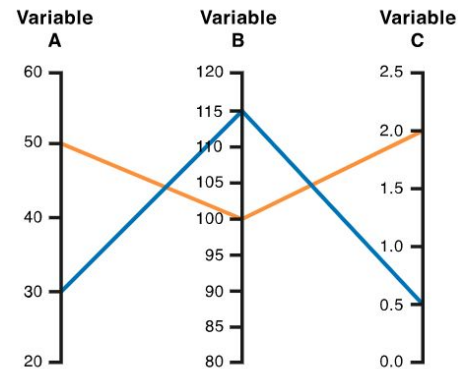
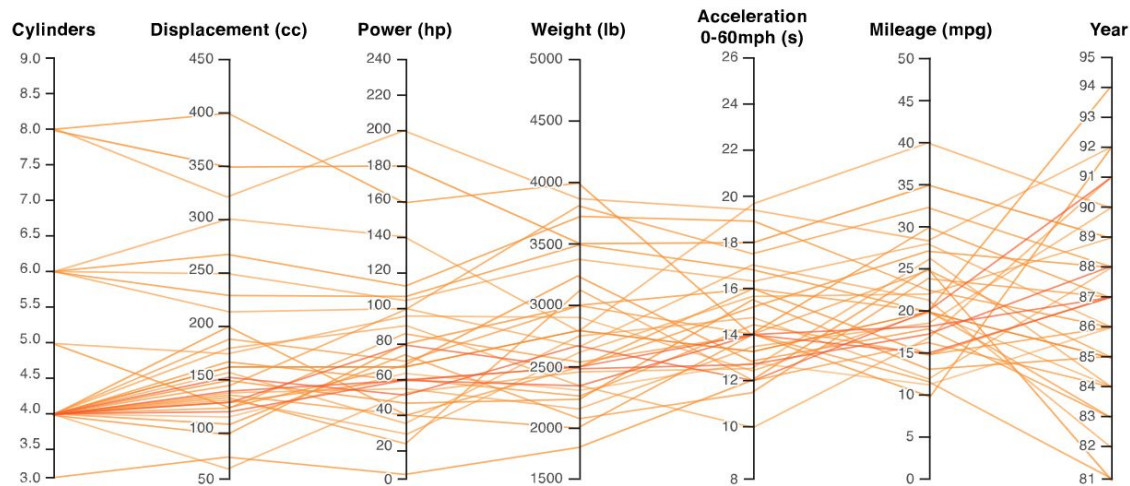
Orientación de ejes - Rectilíneo

- Es el más frecuente. Generalmente con 2 ejes o con 3 cuando nos pasamos a visualizaciones 3D.
- Nos restringe la cantidad de atributos a visualizar por cantidad de ejes.



Orientación de ejes - Paralelo

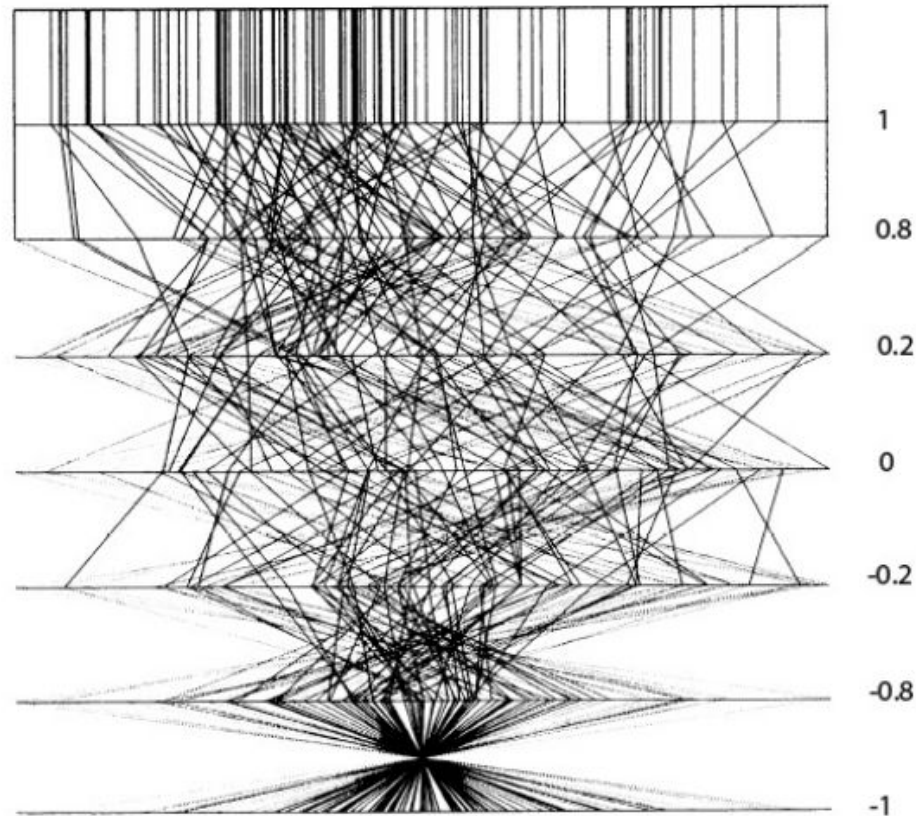
- Podemos incluir la cantidad de ejes que queramos.
- Gráfico de coordenadas paralelas (*parallel coordinates*).** Útil para resumir todos los atributos cuantitativos, para mostrar el rango de atributos individuales, seleccionar un rango de ítems, detección de datos atípicos y relaciones entre pares de atributos.



Data			
	Variable A	Variable B	Variable C
Item 1	50	100	2.0
Item 2	30	115	0.5

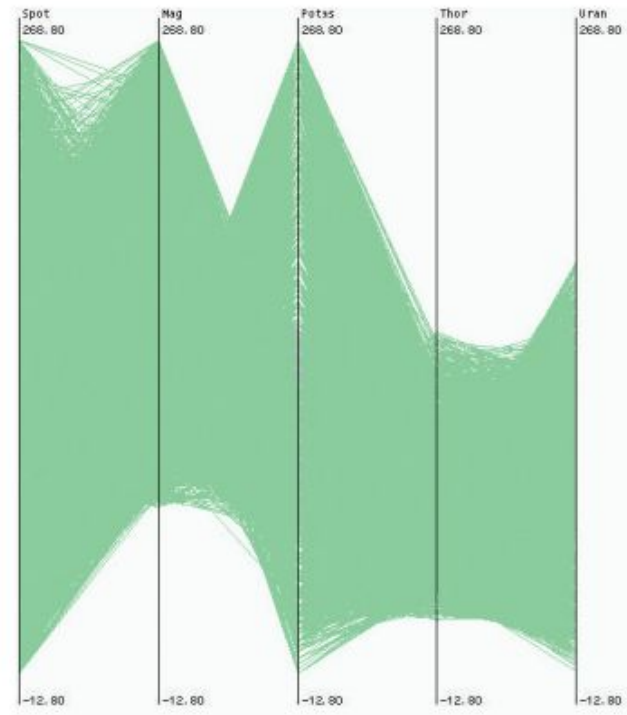
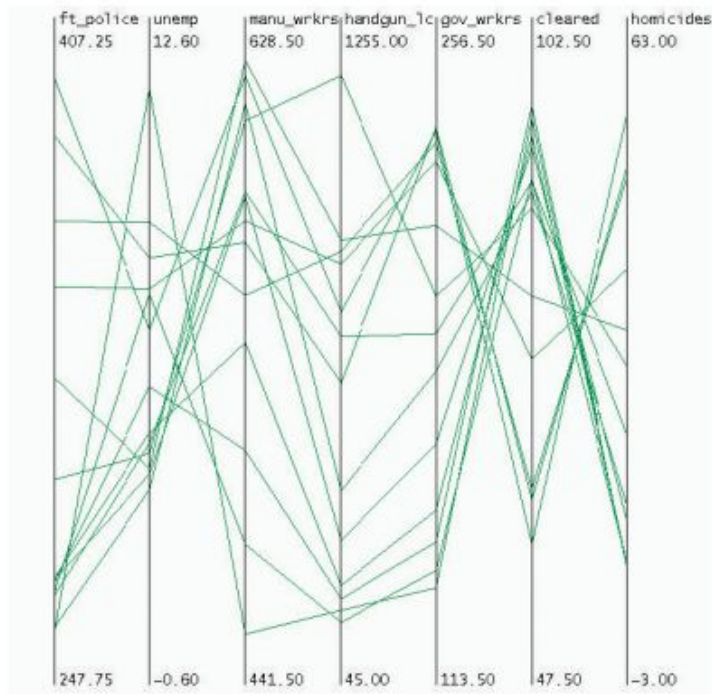
Orientación de ejes - Paralelo

- **Correlación positiva:** conexiones paralelas entre sus ejes.
- **Correlación negativa:** líneas que se cruzan e intersectan en una región concentrada.
- **Poca correlación:** mezcla de líneas en múltiples ángulos.



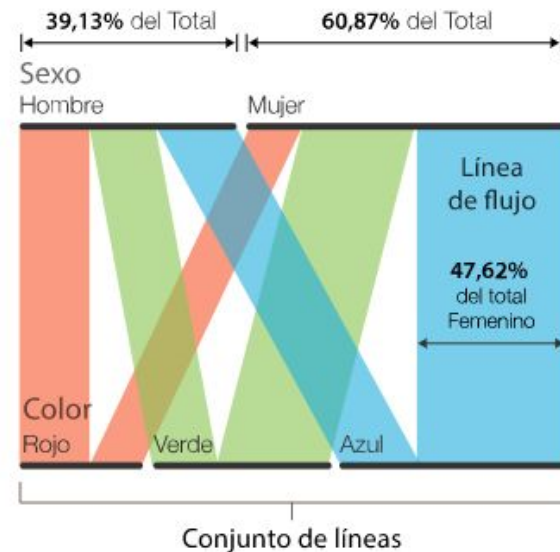
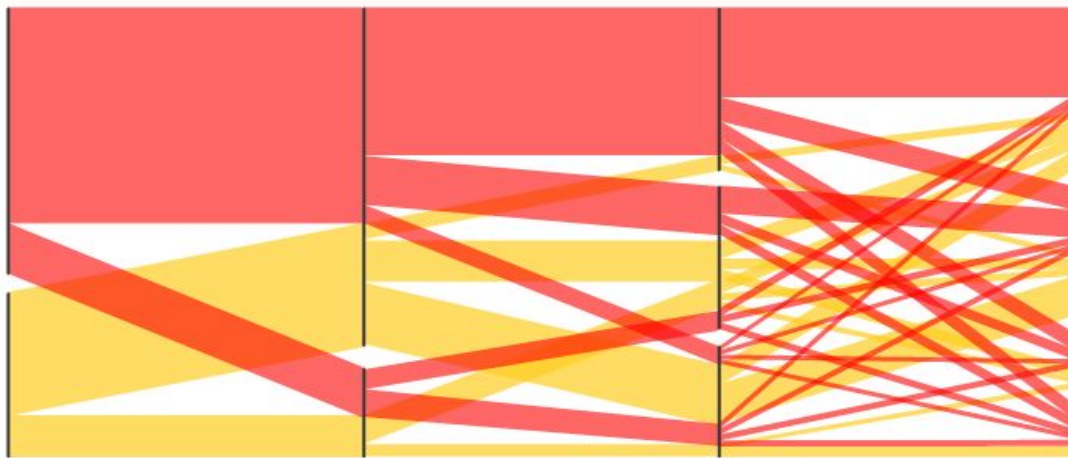
Orientación de ejes - Paralelo

- 👁️ con la cantidad de ítems. A medida que llegamos a los miles de líneas, se vuelve ineficiente su uso para muchas de sus tareas visuales.



Orientación de ejes - Paralelo

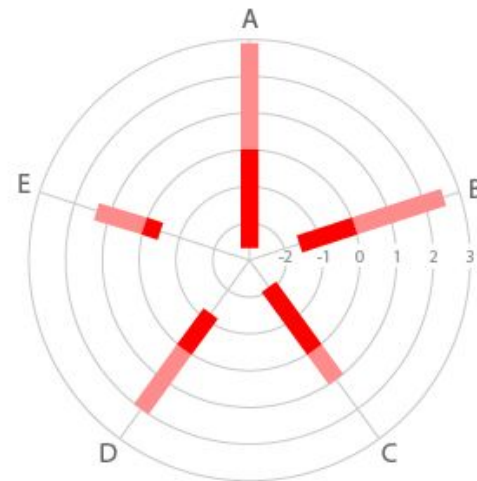
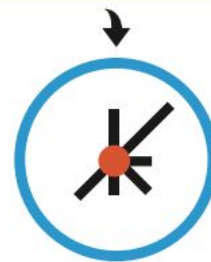
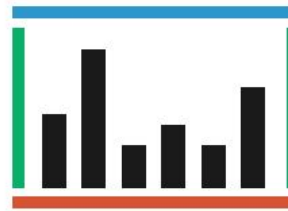
- Otro ejemplo es el **gráfico de conjuntos paralelos** (*parallel sets*). Útil para resumir la proporción de los atributos categóricos y comparar la proporción entre pares de atributos.



Sexo	Color	Cantidad	% del sexo TOTAL
Hombre	Rojo	35	32,41
	Verde	33	30,56
	Azul	40	37,04
Mujer	Rojo	28	16,67
	Verde	60	35,71
	Azul	80	47,62

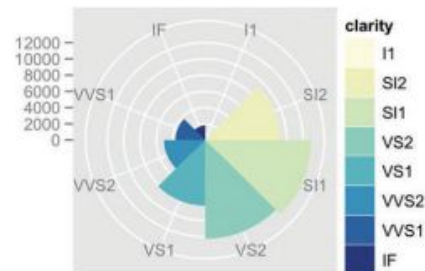
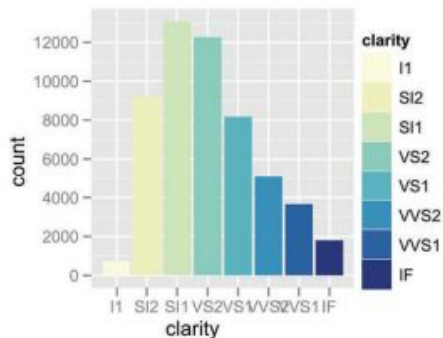
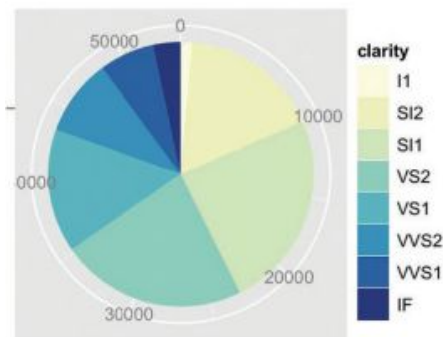
Orientación de ejes - Radial

- Los ítems se distribuyen a lo largo de un círculo, donde se usa el canal de ángulo como distinguir los ítems y se puede apoyar en otros canales como el color: largo y posición.
- Utilizados frecuentemente:
 - Para datos cíclicos (meses por ejemplo).
 - Por temas estéticos.
 - Por falta de espacio para uso de ejes rectilíneos.



Orientación de ejes - Radial

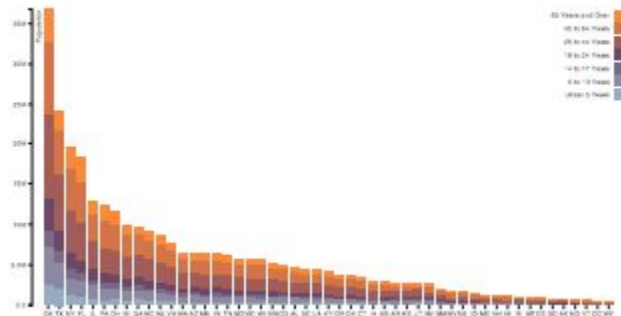
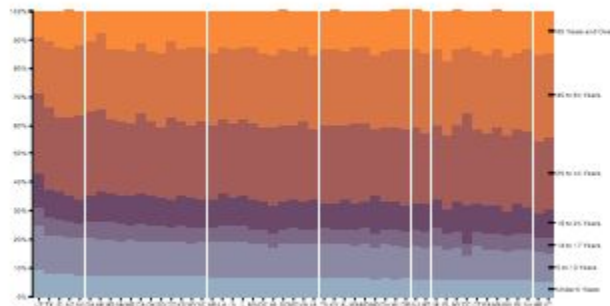
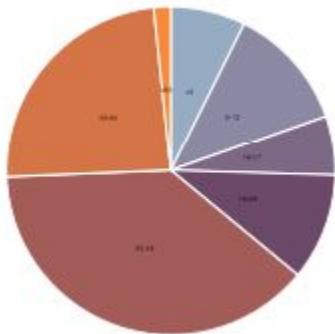
- Otro caso conocidos son:
 - **El gráfico de torta o *pie chart*** - Utiliza esa área circular para codificar un único atributo. El ángulo (o tamaño del área circular) representa el valor.
 - **El gráfico área polar o *polar area chart*** - Dedicar el mismo ángulo para cada atributo, pero varía el largo de la región para representar el valor. Este gráfico busca arreglar en cierta medida el problema de las proporciones que tiene el gráfico de torta convencional.



Orientación de ejes - Radial

Mini reflexión

- Los gráficos de torta y polares se suelen ocupar para mostrar la relación de una parte versus el total, pero **no son los únicos gráficos** capaces de hacer esto.
- Si se normalizan los datos con respecto a un total, es posible replicar esta tarea en gráficos de barras apiladas.



Revisión de Contenidos (RC)

- Se publicará entre **hoy o mañana** un control sobre esta clase.
- Misma modalidad que el control anterior:
 - N preguntas
 - N-1 preguntas correctas para obtener el punto RC
 - Intentos infinitos y verán el puntaje obtenido en cada intento.
- 2 semanas desde su publicación.

Próimos eventos

Próxima clase

- Utilidades D3 II (Eventos, transiciones y data join personalizado).
- Traer *notebook* 

Ayudantía de mañana

- Construir una visualización paso a paso (usar *data joins*, escalas y ejes).

IIC2026

Visualización de Información

— Hernán F. Valdivieso López —
(2022 - 2 / Clase 07)
