

Visualización de datos

Árboles y redes

Contenido

1

Contexto

2

Redes

3

Árboles



Contexto

Un poco de contexto...

La familia de lenguajes de codificación visual del diagrama de nodos utiliza el canal de **conexión**, donde las marcas representan los enlaces en lugar de los nodos.

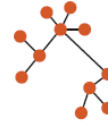
La segunda familia principal de modismos de codificación de redes son las **vistas matriciales** que muestran directamente las relaciones de adyacencia.

La estructura de árbol puede mostrarse con el canal de **contención** en el que las marcas de los enlaces adjuntos muestran las relaciones jerárquicas a través de la anidación.

Arrange Networks and Trees

Node-Link Diagrams Connection Marks

✓ NETWORKS ✓ TREES



Adjacency Matrix Derived Table

✓ NETWORKS ✓ TREES



Enclosure Containment Marks

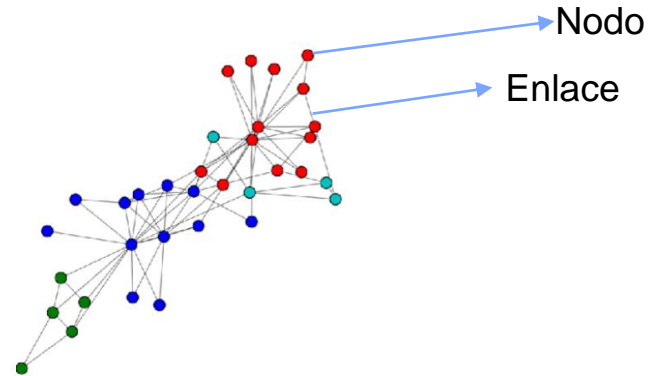
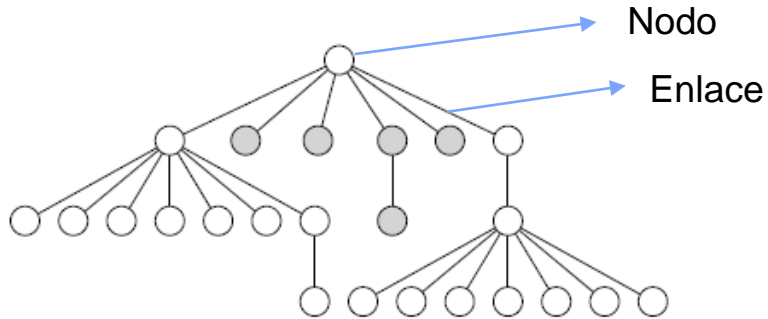
✗ NETWORKS ✓ TREES



Conexión: Marcas de enlace

El lenguaje de codificación visual más común para los datos de árboles y redes es el de los diagramas de **enlace de nodos**, donde los **nodos** se dibujan como marcas de **puntos** y los **enlaces** que los conectan se dibujan como marcas de **líneas**.

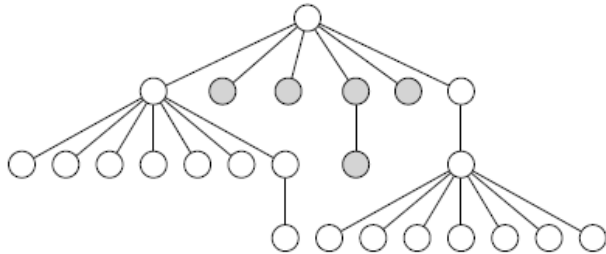
Este lenguaje utiliza marcas de **conexión** para indicar las **relaciones entre los elementos**.



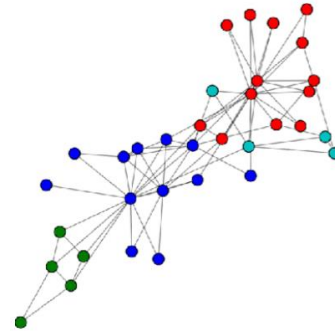
Conexión: Marcas de enlace

Utiliza el canal de posición espacial vertical para mostrar la profundidad en el árbol.

La posición espacial horizontal de un nodo no codifica directamente ningún atributo.



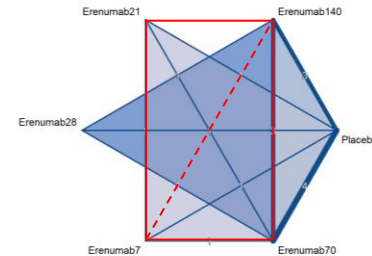
La posición espacial, tanto horizontal como vertical de los nodos, no codifica directamente ningún atributo. Sin embargo, si utiliza el espacio mediante un algoritmo.



Conexión: Marcas de enlace

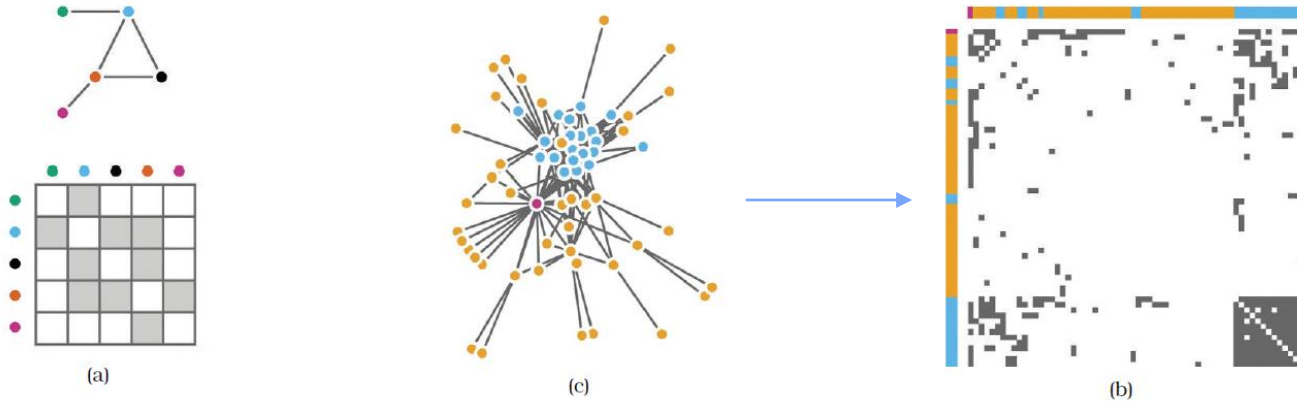


Los diagramas de enlaces de nodos son adecuados para tareas que implican la comprensión de la **topología de la red**: las conexiones directas e indirectas entre los nodos en términos del número de saltos entre ellos a través del conjunto de enlaces.

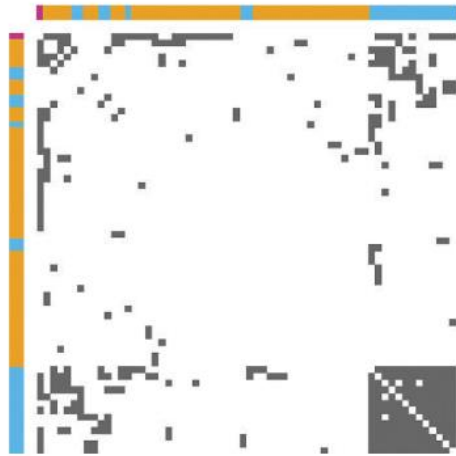


Matriz de adyacencia

Todos los **nodos** de la red se disponen a lo largo de los bordes **verticales** y **horizontales** de una región cuadrada y los **vínculos** entre dos nodos se indican coloreando una **marca de área** en la celda de la matriz que es la **intersección** entre su fila y su columna.



Matriz de adyacencia

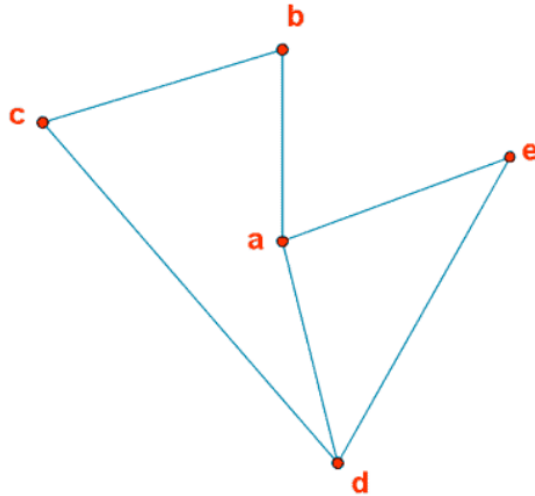


(b)

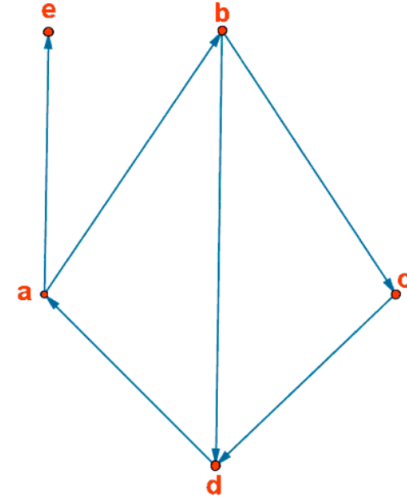
- Usa marcas de área para los enlaces usando como canales la posición espacial, tanto horizontal como vertical de los nodos para los nodos.
- La información adicional sobre otro atributo se codifica a menudo coloreando las celdas de la matriz, una posibilidad que se deja abierta por esta elección de diseño basada en el espacio.
- Para grafos no dirigidos, la matriz es simétrica.
- Para grafos dirigidos, la matriz no es simétrica.

Grafos: dirigidos y no dirigidos

Grafo no dirigido



Grafo dirigido



https://calculo.cc/temas/temas_algebra/matriz/teoria/matriz_grafo.html

Para saber más... <https://medium.com/techmacademy/graphs-adjacency-matrix-behind-the-scenes-visual-tour-6512493912c0>

Matriz de adyacencia



- Ver relaciones de conexión y dependencia
 - Identificar Clusters
- Las vistas matriciales de las redes pueden alcanzar una densidad de información muy alta, hasta un límite de mil nodos y un millón de aristas, al igual que los mapas de calor de los clústeres y todas las demás vistas matriciales que utilizan marcas de área pequeñas.

Contención: Marcas de jerarquía

Las marcas de contención son muy eficaces a la hora de mostrar información completa sobre la **estructura jerárquica**, a diferencia de las marcas de conexión que sólo muestran las relaciones de pareja entre dos elementos a la vez, en éstas marcas las relaciones jerárquicas se muestran con **contención** en lugar de **conexión**.

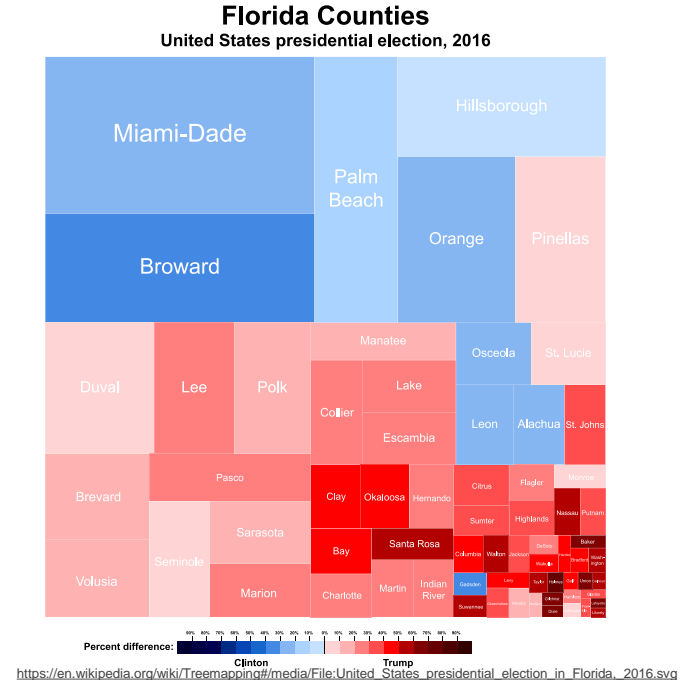


<https://www.amcharts.com/dataviz-tip-27-chart-types-treemap/treemap/>

Contención: Marcas de jerarquía

Todos los hijos de un nodo del árbol están encerrados dentro del **área** asignada a ese nodo, creando un diseño anidado. Además se puede agregar un canal de color.

Las marcas de contención no son tan eficaces como las marcas de conexión por pares de conexión por pares para las tareas centradas en **la estructura topológica**, como trazar caminos a través del árbol, pero brillan para las tareas que pertenecen a **comprender los valores de los atributos** en las hojas del árbol.



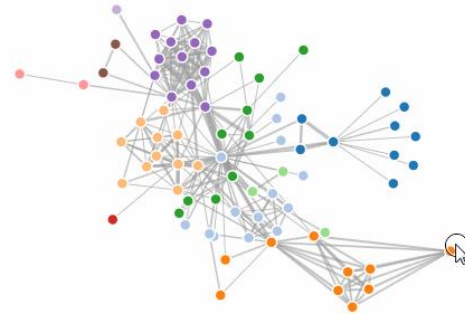
Redes

Conjunto de nodos y enlaces que expresan un tipo de relación. Donde cada nodo:

- Está etiquetado de forma única
- Contiene uno o más atributos categóricos que pueden tener más de un valor.

Red dirigida por la fuerza

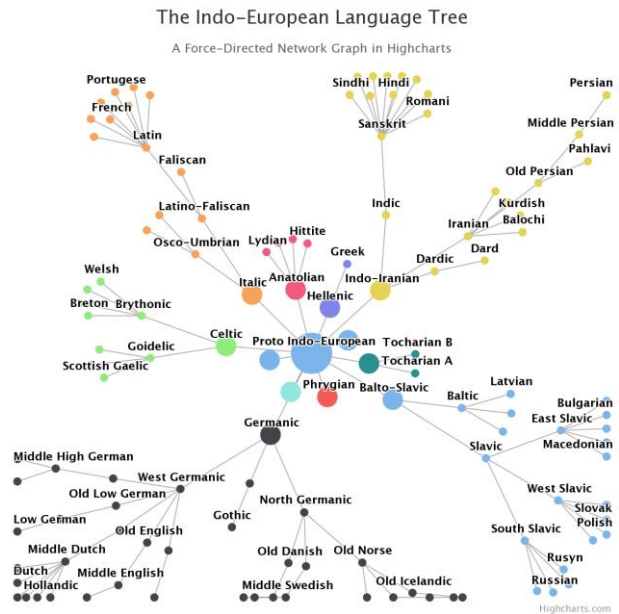
- Tipo de datos: Redes
- Codificación:
 - Marcas de puntos para nodos, marcas de líneas para enlaces.
 - Posición espacial: sin significado codificado directamente (se deja libre para minimizar los cruces)
- Tareas: Explorar la topología, localizar caminos.



<http://jsfiddle.net/draditya91/RRUzZ/>

Red dirigida por la fuerza

- ¿Semántica de proximidad?
 - A veces tiene sentido
 - A veces arbitraria, artefacto del algoritmo de diseño
 - Tensión con la longitud (las aristas largas son más destacadas visualmente que las cortas)
- Consideraciones:
 - Docenas/cientos de nodos, cientos de enlaces
 - Densidad de nodos/enlaces:
 $\# \text{Enlaces} < 4 \# \text{Nodos}$

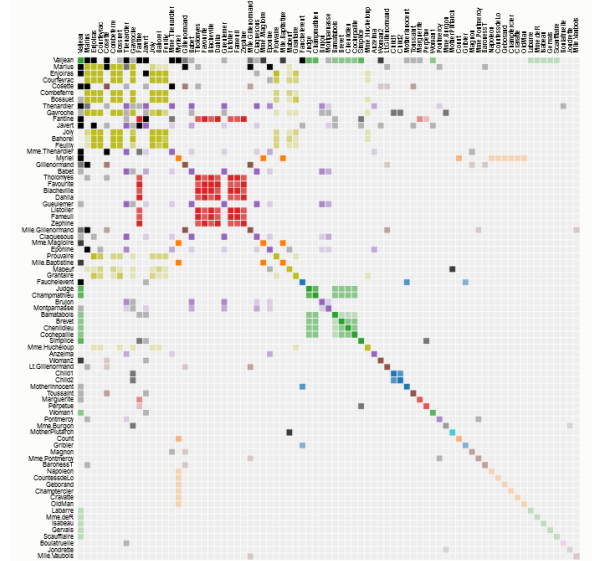


<https://www.highcharts.com/demo/network-graph>

Matriz de adyacencia

- Datos: red (transformada en los mismos datos/codificación que el mapa de calor)
- Datos derivados: tabla de la red
 - 1 atributo cuantitativo
 - 2 atributos categóricos: lista de nodos x 2
- Codificación visual: la celda muestra la presencia/ausencia de un enlace
- Escalabilidad: 1K nodos, 1M enlaces

Les Misérables Co-occurrence



Source: [The Stanford GraphBase](https://stanford-graphbase.org/).

<https://bost.ocks.org/mike/miserables/>

Conexión frente a matriz



Pros: Comprensión de la estructura topológica de la red, como el **trazado de rutas** y la **búsqueda de vecindarios topológicos** locales a un pequeño número de saltos de un nodo objetivo, y también pueden ser muy eficaces para tareas como la **visión general** o la **búsqueda de subestructuras** similares.

Contras: Una alta densidad de nodos y enlaces, se vuelven imposibles de leer debido a la **oclusión** de las aristas que se cruzan entre sí y que cruzan por debajo de los nodos.

Pros: Uno de los puntos fuertes de las vistas matriciales es la **escalabilidad perceptiva** para redes grandes y densas

Otro punto fuerte de las vistas matriciales es su **previsibilidad**, **estabilidad** y su capacidad de **reordenación**.

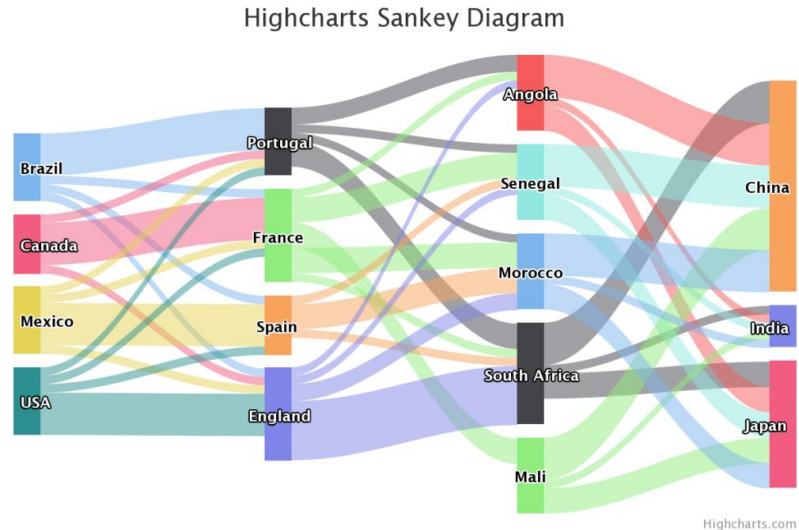
Contras: la mayoría de los usuarios suelen necesitar formación para interpretar las vistas matriciales, debido a la falta de **familiaridad** para las redes. Muestran los enlaces de forma más **indirecta** que las conexiones directas de los diagramas de nodos.

- [illegible]

<https://observablehq.com/@john-guerra/mouseover-chord-diagram>

Sankey diagram

- Datos: Redes (pequeño número de nodos)
- Canales: Posición horizontal, color
- Tareas: Resumir las conexiones. Identificar el grado más alto
- Consideraciones: Normalmente es bueno para origen -> destino



<https://dataforvisualization.com/charts/sankey-diagram/>

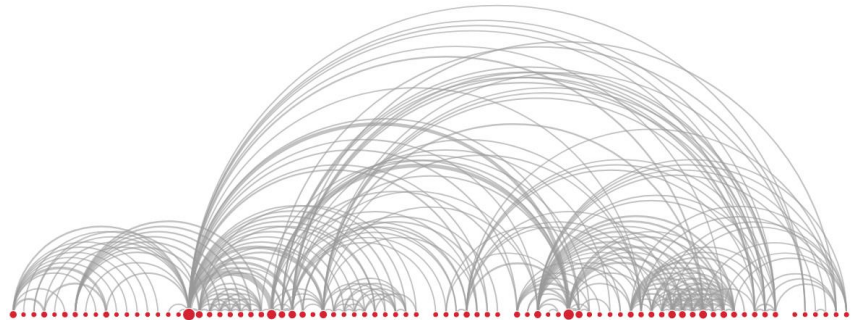
100

- [illegible]

<https://observablehq.com/@d3/hierarchical-edge-bundling>

Arc diagram

- Datos: Redes (pocos nodos)
- Tareas: Resumir las conexiones comunes.
- Consideraciones:
 - El orden de los nodos importa
 - Mejor con datos muy agrupados



<https://blockbuilder.org/rpgove/53bb49d6ed762139f33bdaea1f3a9e1c>

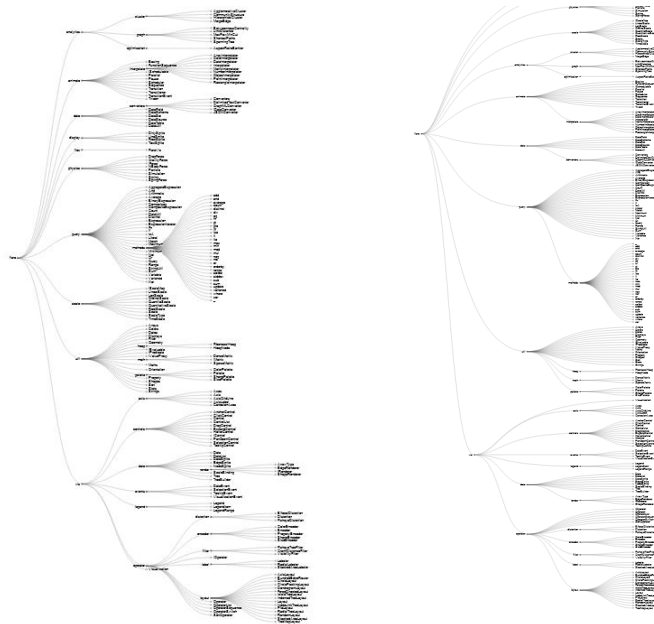
Ejemplo: <https://observablehq.com/@d3/arc-diagram>

Árboles

- Conjunto de nodos y enlaces que expresan la relación padre-hijo. Donde cada nodo:
 - Está etiquetado de forma única en el árbol.
 - Contiene una o más variables numéricas con valores en el tiempo.
 - Contiene uno o más atributos categóricos que pueden tener más de un valor.

Tidy Tree - Dendrogram

- Datos: Árbol
- Codificación:
 - Marcas de conexión de enlaces
 - Marcas de puntos para los nodos
 - Orientación de la posición
 - Vertical para enlaces
 - Horizontal para nodos
- Tareas: comprensión de la topología, seguimiento de trayectorias
- Escalabilidad: Cientos de enlaces y nodos



Ejemplo collapsible: <https://observablehq.com/@d3/collapsible-tree>

<https://bl.ocks.org/mbostock/e9ba78a2c1070980d1b530800ce7fa2b>

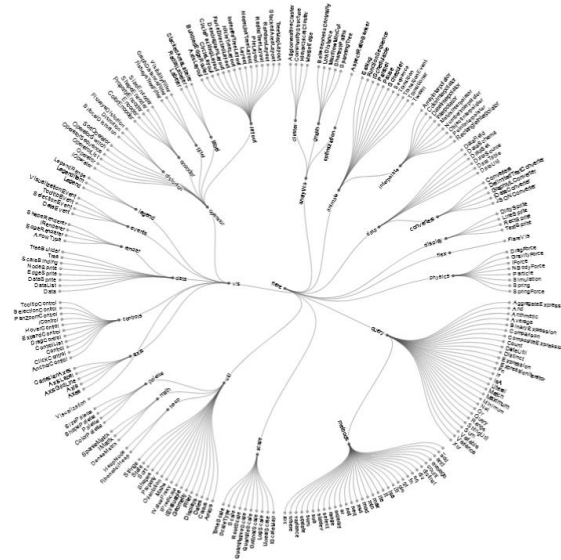
100

- [illegible]

<https://observablehq.com/@d3/radial-tidy-tree>

Dendogramas de conglomerados radiales

- Tareas: Resumir las conexiones comunes.
- Consideraciones:
 - Más difícil de leer
 - Mejor uso del espacio



<https://observablehq.com/@d3/radial-dendrogram>

Treemap

- Datos: Árbol, 1 atributo cuantitativo en los nodos de las hojas.
- Codificación:
 - Marcas de contención del área para la estructura jerárquica
 - El tamaño codifica el atributo cuantitativo
- Tarea: Consultar el atributo en los nodos de la hoja
- Escalabilidad: 1M de nodos hoja



<https://finviz.com/map.ashx?t=sec>

Gracias

¿Preguntas?

Bibliografía

- Munzner Tamara, Visualization Analysis and Design, Department of Computer Science University of British Columbia , Capitulo 9 (2014)
- <https://www.twosixlabs.com/6-ways-visualize-graphs/>
- Basado en el Curso de visualización de datos en D3 por Jhon Alexis Guerra https://johnguerra.co/lectures/visualAnalytics_fall2019/