

Resumen Paper

Visual Encodings: Visual Encoding es traducir la información en un elemento visual para que sea más fácil de comprender.

Graphical Perception: Es la capacidad humana de visualmente interpretar gráficos

Time-Series Data

Index Charts

Un index chart es una línea interactiva que muestra cambios en el porcentaje para un conjunto de información que evoluciona con el tiempo en un punto específico.

Entonces dependiendo dónde posiciones el cursos, el gráfico va cambiando. Este ejemplo resulta bastante claro.

Stream Graph

Stream Graph

Also known as a ThemeRiver. This type of visualisation is a variation of a Stacked Area Graph, but instead of plotting values against a fixed, straight axis, a Stream Graph has values displaced around a varying central baseline.

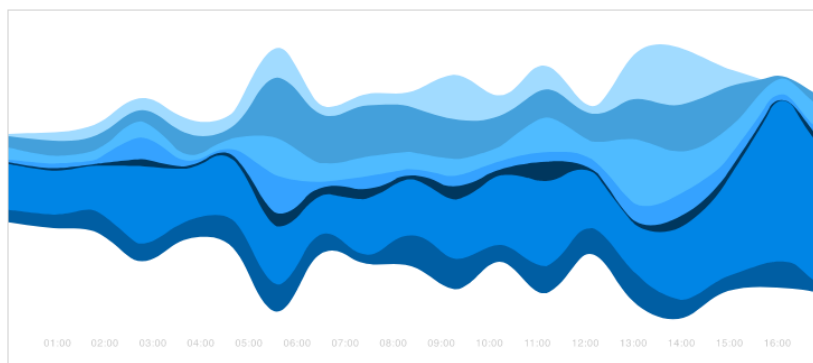
 https://datavizcatalogue.com/methods/stream_graph.html



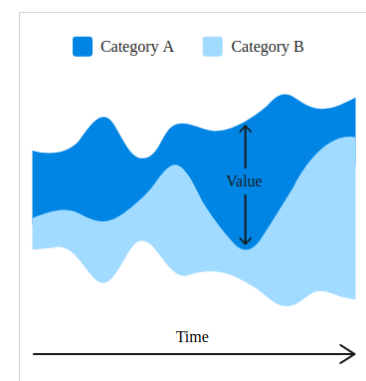
Es una variación del Stacked Area Graph pero en vez de graficar la información contra un eje fijo un stream graph describe los valores aumentando o disminuyendo el grosor del gráfico. En este tipo de gráfico el grosor de cada *stream* o elemento a estudiar es proporcional al valor de esa categoría.

Son ideales para datasets de alto volumen para descubrir *trends* y patrones a través del tiempo

Stream Graph



Anatomy





Tiene algunas limitaciones como que no soporta número negativo y es inútil ante información que no puede ser sumada.

Small Multiples

Small multiple

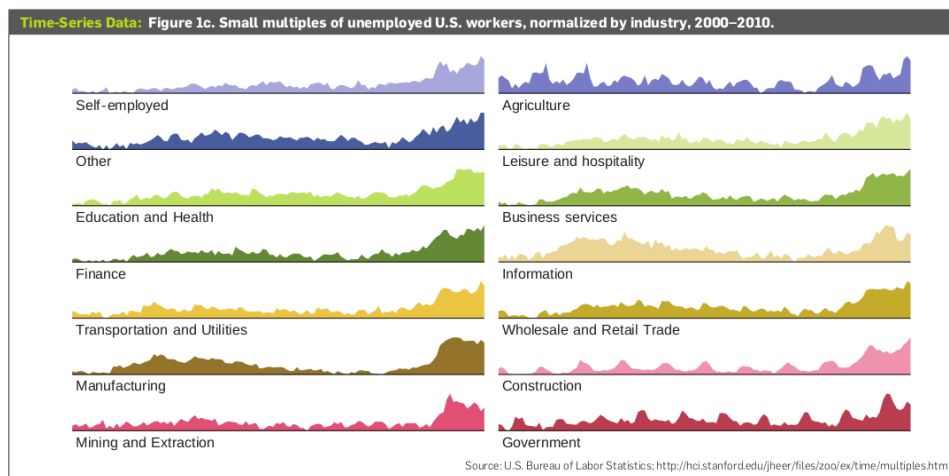
A small multiple (sometimes called trellis chart, lattice chart, grid chart, or panel chart) is a series of similar graphs or charts using the same scale and axes, allowing them to be easily compared. It uses multiple views to show different partitions of a dataset. The term

W https://en.wikipedia.org/wiki/Small_multiple



Es una serie de gráficos similares utilizando la misma escala y los mismos ejes de forma que puedan ser fácilmente comparables. Utiliza distintas vistas para mostrar diferentes particiones del dataset.

Se pueden utilizar small multiples para casi cualquier tipo de visualización.



Horizon Graphs

Usamos este tipo de gráfico cuando queremos comparar más de una serie de tiempo a la vez. Es una técnica para incrementar la densidad de la información en función del tiempo pero manteniendo la resolución.

Es como una combinación de un heat map y un gráfico en función del tiempo.

Acá hay una explicación piola que explica cómo crearlo y de esta manera se pueden entender un poco mejor.

Statistical Distributions

exploratory data analysis: Obtener información de como la información está distribuida para ver cómo se transforma la información y modelando decisiones.

Algunas técnicas conocidas son el histograma y el box-and-whisker plot

Stem-and-Leaf plots

Sirve para estudiar grupos de números. Típicamente agrupa los números según su primer dígito significativo y luego agrupa en orden los valores según el segundo dígito significativo.

Northbound	Hour	Southbound
	5	05 48
45 20 03	6	02 23 35 57
55 49 32 20 13 01	7	00 07 16 20 26 30 37 46 52 59
58 53 49 44 38 32 25 19 13 08 02	8	01 08 12 17 21 29 31 35 39 44 49 53 58
59 57 54 50 47 44 39 35 31 28 24 21 18 14 09 05 00	9	03 10 18 27 32 37 45 51 58
52 48 44 39 34 29 23 18 12 05	10	00 07 14 21 30 39 48 57
53 47 41 37 32 27 22 15 07	11	06 11 19 27 34 41 50 59
55 49 35 29 23 16 08 01	12	02 15 30 45 57
56 48 44 39 32 27 21 14 05	13	03 10 18 23 29 37 45 56
50 45 35 30 25 20 15 05	14	00 09 18 27 39 48 57
52 43 32 24 12 03	15	01 17 29 41 55
58 44 31 26 15 06	16	10 25 38 50
56 40 30 22 11	17	00 20 34 53
55 41 32 23 14 01	18	05 14 21 29 37 45 56
58 49 42 36 28 22 16 09	19	02 09 14 19 23 27 32 36 40 44 48 53 57
57 51 46 39 33 28 23 17 13 08 02	20	09 17 26 34 40 49 55
52 43 30 21 15 06	21	10 20 30 40 50
45 30 16 03	22	15 35 55
50 30 10		

Anatomy

STEMS	LEAVES
0	5 8
1	2 3 5 7
2	0 0 0 5 8 8 9
3	0 0 1 3 3 3 6 6 7 7 7 7 7 8 8 8 8 9 9
4	1 3 5 5 5 6 7 7 8 8 8 8 9 9
5	0 0 0 1 1 1 1 2 6 8
6	0 0 1 1 2 4 4 4 4 4 8 8 9
7	0 5 5 5 5 7
8	3 4 4 5 6 6 6 7 8 9
9	0 1 2 2 2 2 5 5 6 8 9 9
10	2 2 2 5 7

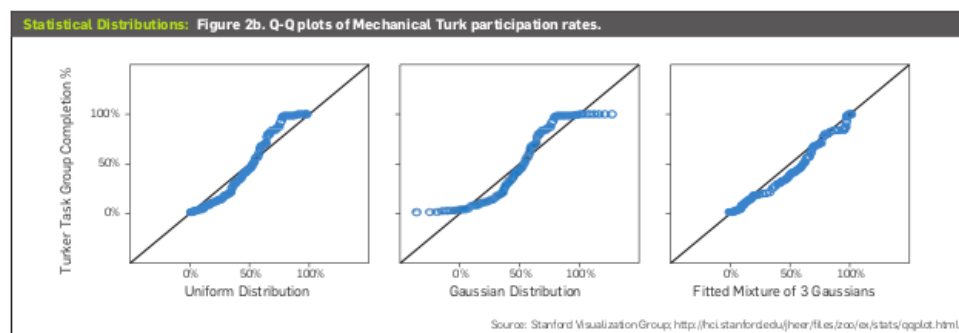
Raw Data in Row:
102, 102, 102, 105, 107

Son muy útiles para darle a la persona que lee un rápido vistazo de la distribución de la información como para también reconocer *outliers*.

- ⚠ Una de sus mayores debilidades es que están limitados por el tamaño del dataset que pueden manejar. Si son pocos datos, entonces el gráfico es inútil, pero si se tienen demasiados datos, se transforma en una bola de números incomprensibles.

Q-Q Plots

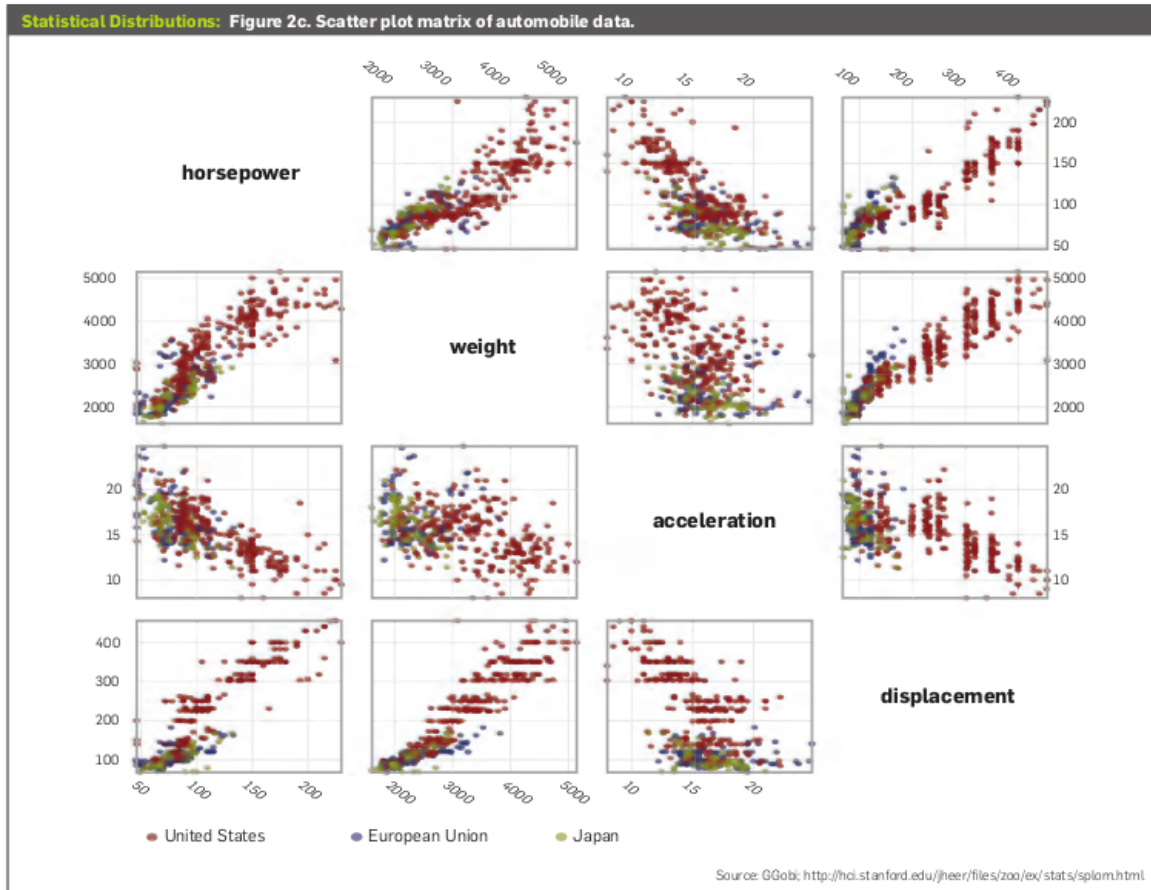
Compara dos distribuciones probabilísticas graficando sus quantiles uno contra otro. Si son similares, los valores graficados se encontrarán cerca de la diagonal central. Si las distribuciones están relacionadas linealmente, los valores también van a trazar una línea sólo que variando la pendiente y altura.



⚠ Una limitación de este tipo de gráfico es que requiere que quien lo lee posea conocimientos de estadística.

SPOM (Scatter Plot Matrix)

Muy útil para representar información con múltiples variables. Utilizamos múltiples pequeños scatter plots que muestran un set de relaciones de a pares entre las variables.



Cada par está representado dos veces, de forma transpuesta entre sí. Si vamos al [gráfico interactivo](http://hci.stanford.edu/jheer/files/zoo/ex/stats/splom.html) utiliza una técnica llamada **brushing-and-linking** en la cual haciendo un cuadrado sobre un gráfico te muestra ese mismo cuadrado en todos los demás

Parallel Coordinates

https://datavizcatalogue.com/methods/parallel_coordinates.html

https://datavizcatalogue.com/methods/parallel_coordinates.html

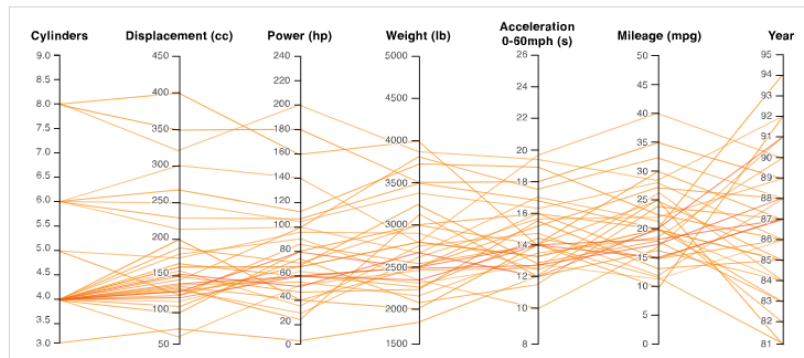
Es un tipo de gráfico utilizado para graficar información numérica multivariable.

Cada variable tiene su propio eje y los ejes son puestos en paralelos entre sí. Cada eje puede tener una escala diferente o tener todos la misma escala normalizada. Los valores se grafican como líneas que van de eje en eje.

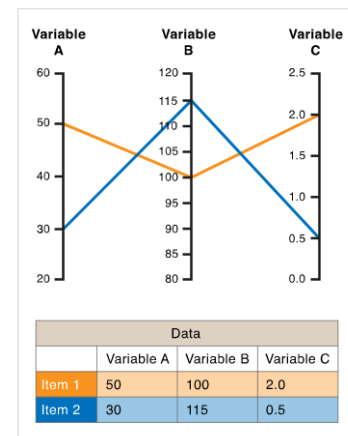
El orden de los ejes es importante y puede impactar en la forma en la que quien ve el gráfico entiende la información. Una razón para esto es que relaciones adyacente son más fáciles de notar que relaciones más lejanas, entonces reordenar los ejes puede llevarnos a descubrir patrones y correlaciones entre las variables.

⚠ Una desventaja que tienen es que si se tiene un dataset muy grande se puede poner muy denso y difícil de leer. Una forma de solucionar este problema es utilizando una técnica llamada **brushing** que al acercarse el mouse a una línea o a un cluster de líneas las pone en foco, haciendo casi invisibles a las demás líneas.

Parallel Coordinates Plot



Anatomy



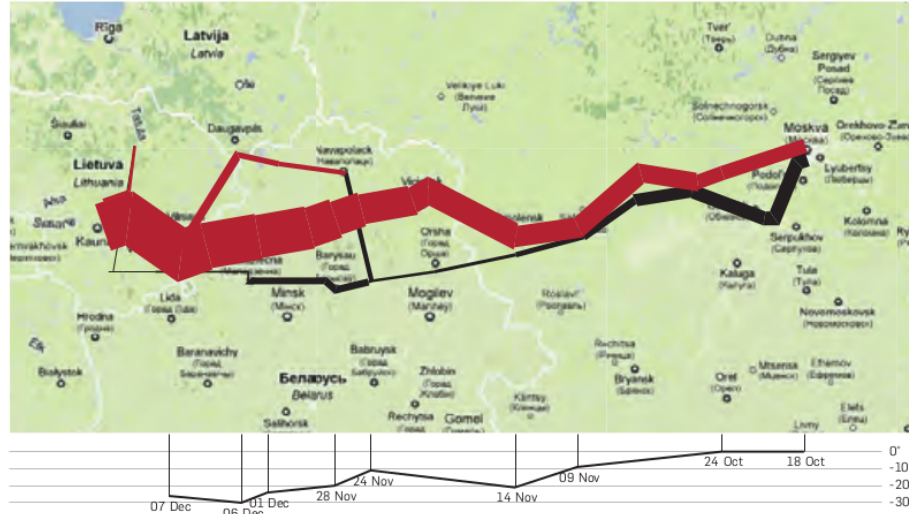
Maps

Flow Maps

Obtenemos un flow map agregándole líneas gruesas por encima de un mapa geográfico. Este tipo de mapa nos muestra el movimiento de una cantidad en el espacio e indirectamente también en el tiempo.

Este tipo de mapas suele tener una gran cantidad de información multivariable: direcciones, grueso de la línea, color, etc.

Maps: Figure 3a. Flow map of Napoleon's March on Moscow, based on the work of Charles Minard.



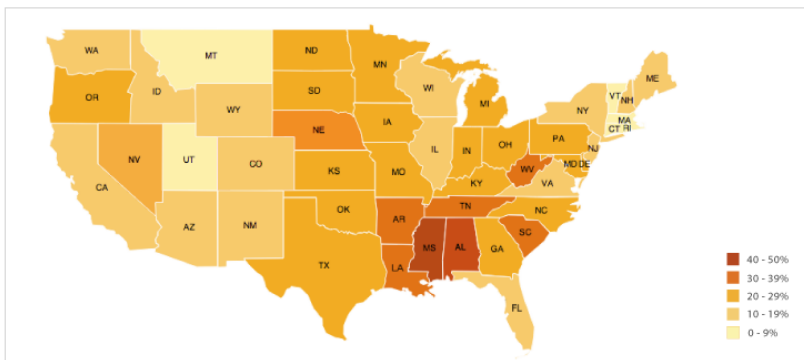
<http://hci.stanford.edu/jheer/files/zoo/ex/maps/napoleon.html>

En este ejemplo tenemos la línea roja que representa la idea de Napoleón con el grosor siendo la cantidad de soldados y en negro vemos la retirada. Junto con la línea negra vemos abajo una escala con las temperaturas en cada posición.

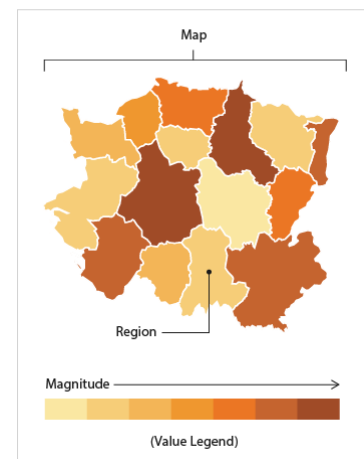
Choropleth Maps

Este tipo de mapa muestra áreas geográficas o regiones coloreadas o con distintos patrones en relación a una variable de la información. Esto nos provee una forma de visualizar valores a través del área geográfica, lo cual nos puede mostrar la variación de la información a través de la posición.

Choropleth Map



Anatomy

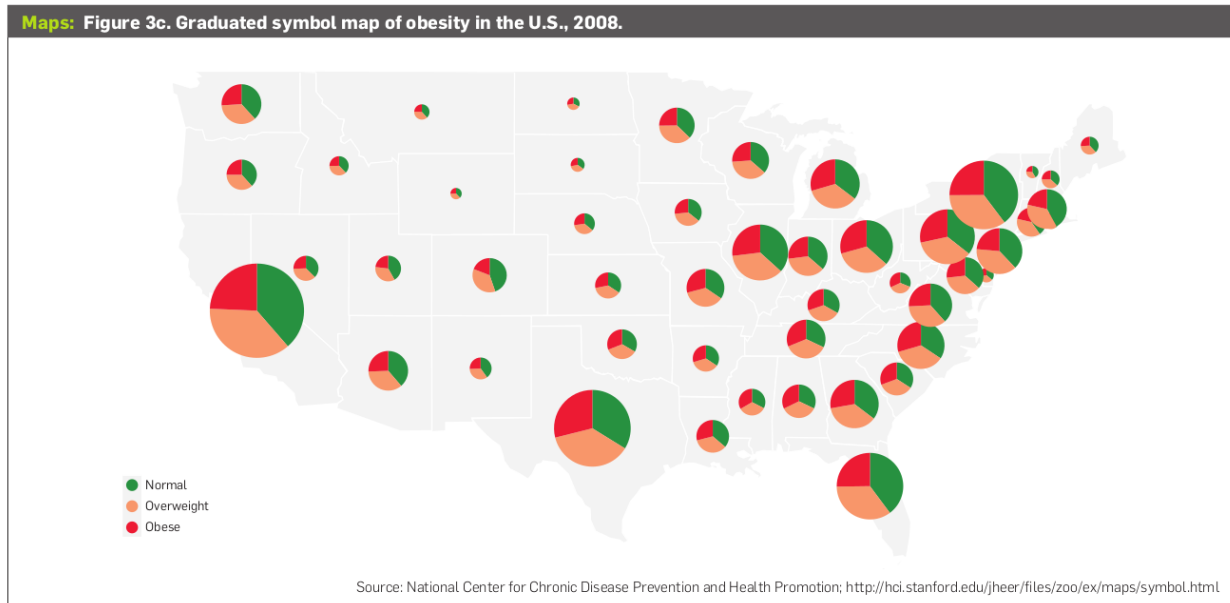


⚠ Una desventaja de usar colores es que no se puede leer o comparar valores exactos utilizando el mapa. Además, otro problema es que las áreas más grande suelen tener más énfasis que las pequeñas.

⊘ Un error común es utilizar valores crudos (como población) en vez de utilizar valores normalizados.

Graduates Symbol Maps

Es una alternativa al Choropleth map, este tipo de mapa pone símbolos sobre un mapa para representar información de cada área/región. Entonces nos permite obtener más información que simplemente el color



Este mapa por ejemplo con el tamaño del círculo muestra el total de la población de dicho estado y en el pie chart muestra las proporciones de gente con un BMI rating específico.

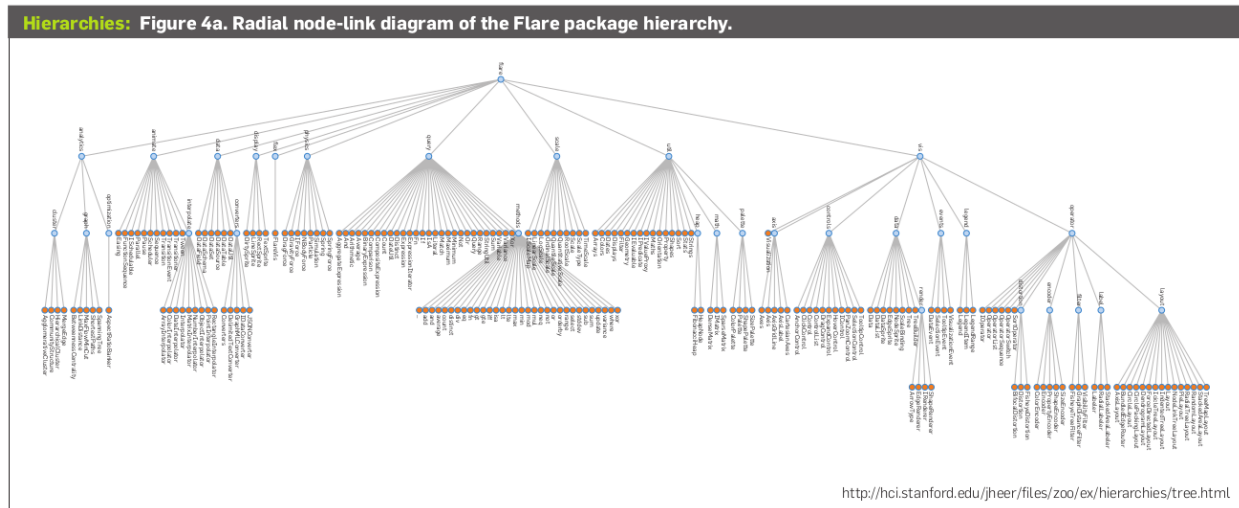
Cartograms

Un cartogram distorsiona la forma de las regiones geográficas para que el área directamente represente a la información. Hay muchos tipos, los cuales se pueden ver [aquí](#).

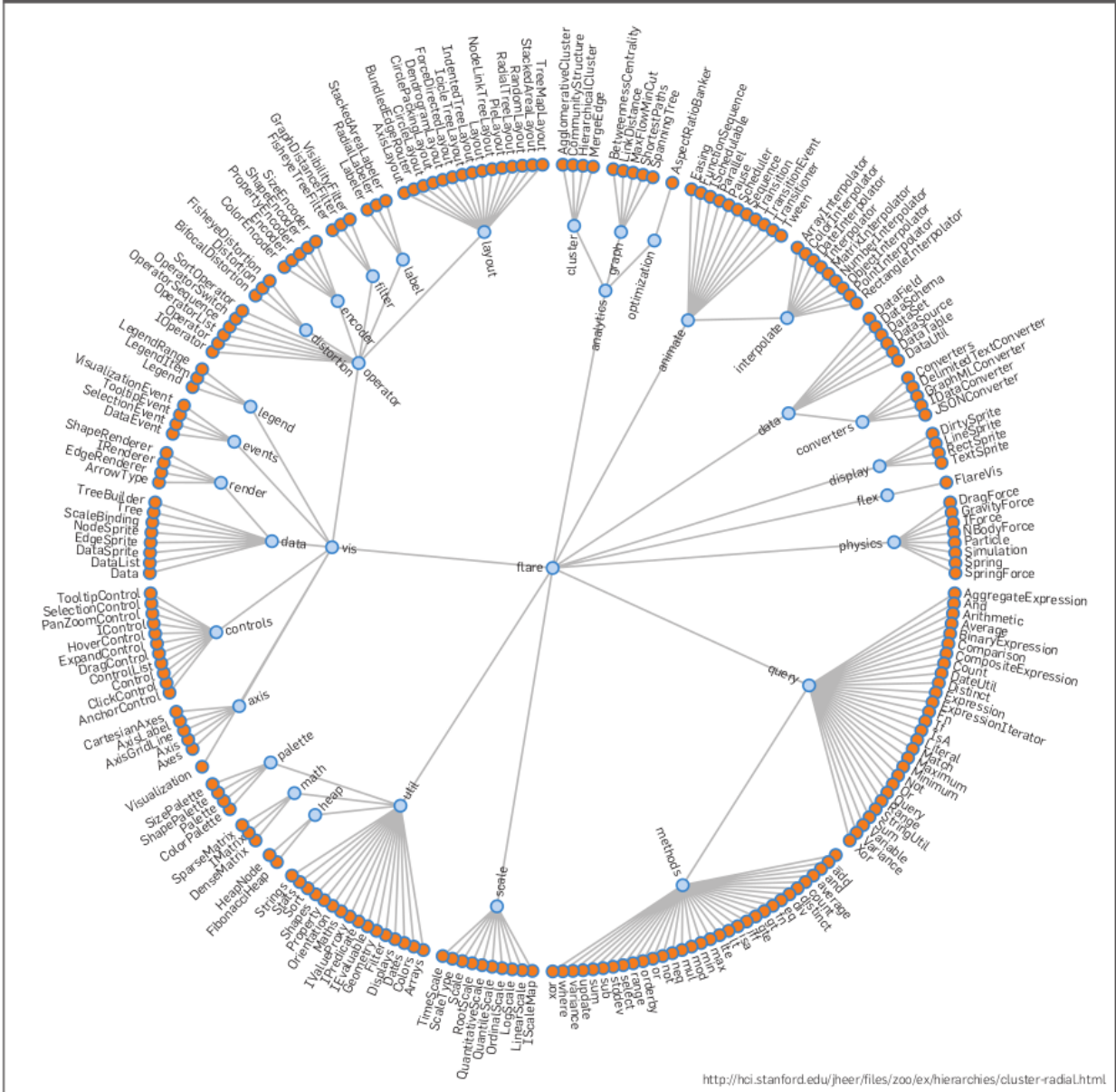
Hierarchies

Node-link diagram

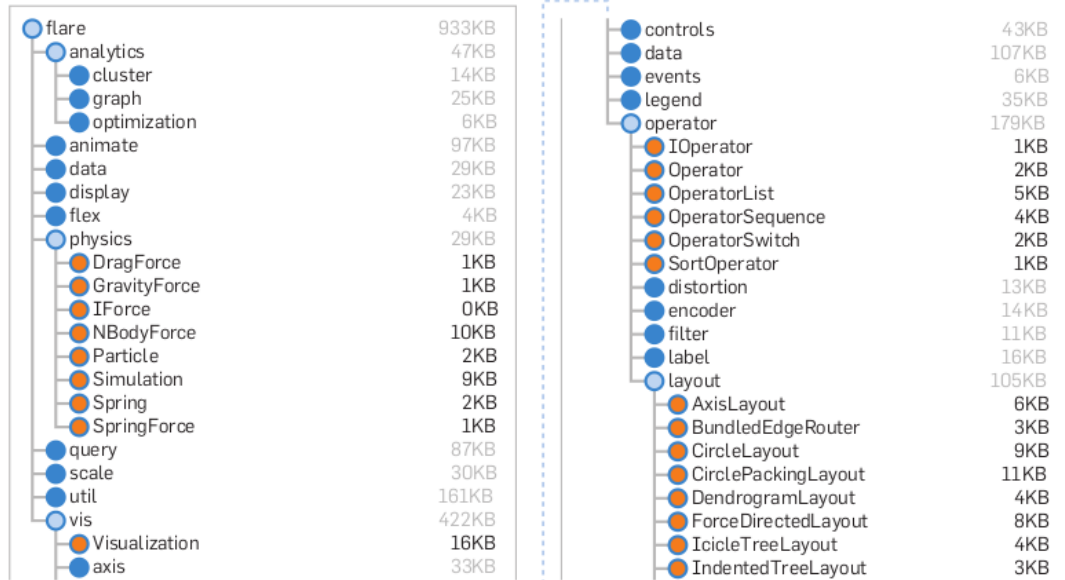
Visión clásica con grafos que puede tener distintos estilos:



Hierarchies: Figure 4b. Cartesian node-link diagram of the Flare package hierarchy.



Hierarchies: Figure 4c. Indented tree layout of the Flare package hierarchy.



<http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/indent.html>

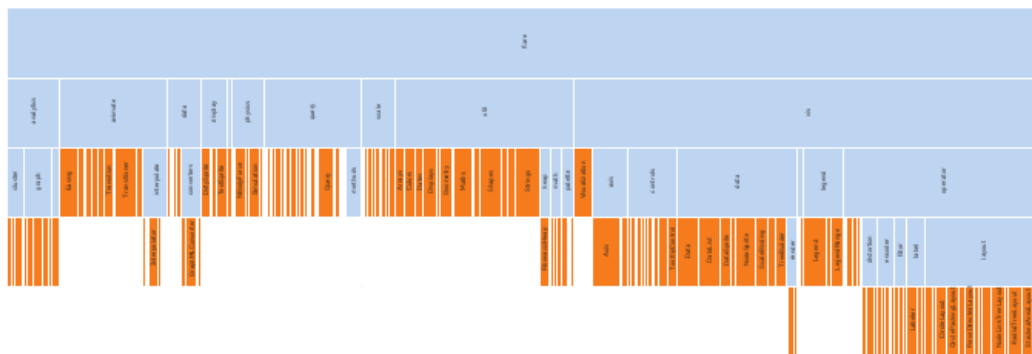
Adjacency Diagrams

Es una vvariente del node-link diagram que se caracteriza por la importancia del espacio ocupado de los campos.

En vez de dibujar un padre e hijo en una jerarquía, los nodos se dibujan como áreas sólidas y su posición relativa a los bloques adyacentes muestran la posición en la jerarquía.

Podemos ver por ejemplo el **icicle layout** que además de mostrar jerarquía, usa el concepto de tamaño del bloque para mostrar el tamaño del mismo. Entonces, en este caso como estamos hablando de archivos de software, mayor área implica mayor tamaño de archivo:

Hierarchies: Figure 4d. Icicle tree layout of the Flare package hierarchy.



<http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/icicle.html>

Este tipo de diagrama también toma importancia el espacio que ocupan y para demostrar jerarquía en vez de utilizar qué bloques son adjacientes usa qué bloques contienen a qué bloque. Entonces un bloque padre

Pet Shop			
Birds		Mammals	
Zebra Finches	Canaries	Dogs	Fish
			Goldfish
			Coy Carp
			Reptiles
			Geckos

```

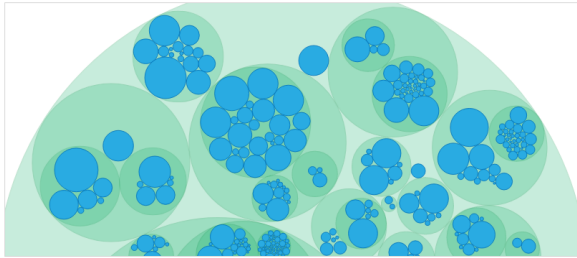
graph TD
    A["A  
200"] --> B["B  
80"]
    A --> C["C  
120"]

```

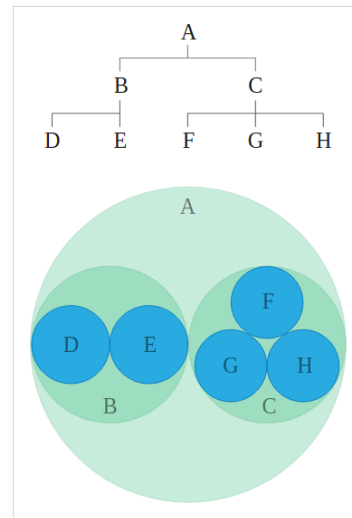
⚠ La desventaja de este tipo de gráfico es que no muestra los niveles de jerarquía tan claramente como otros gráficos de jerarquía

Podemos utilizar círculos en vez de rectángulos (llamado **circle packing**)

Circle Packing



Anatomy



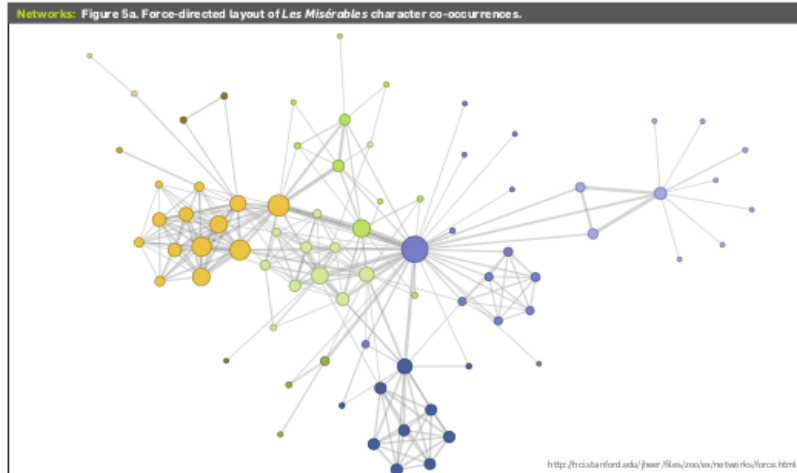
Networks

Además de querer organizar la información, también podemos querer explorar las relaciones entre los datos. Por ejemplo, en una red social quién sigue a quién.

Para eso (para sorpresa de nadie) utilizamos grafos.

Force-directed layouts (?)

Pareciera ser la forma normal de mostrar grafos, no termino de entender bien si hay algún orden específico además del color layout

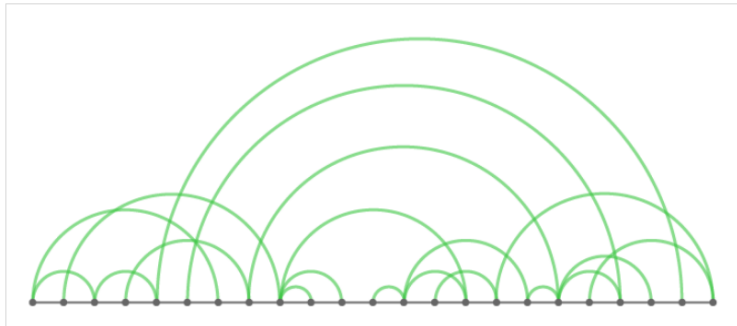


Arc Diagrams

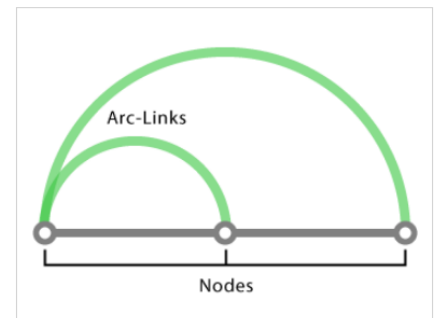
Los nodos se ponen a lo largo de un único eje y los arcos muestran las conexiones entre los nodos.

El grosor de cada arco denota la frecuencia de la relación entre origen y destino. Estos diagramas son útiles para buscar co-ocurrencias en la información.

Arc Diagram



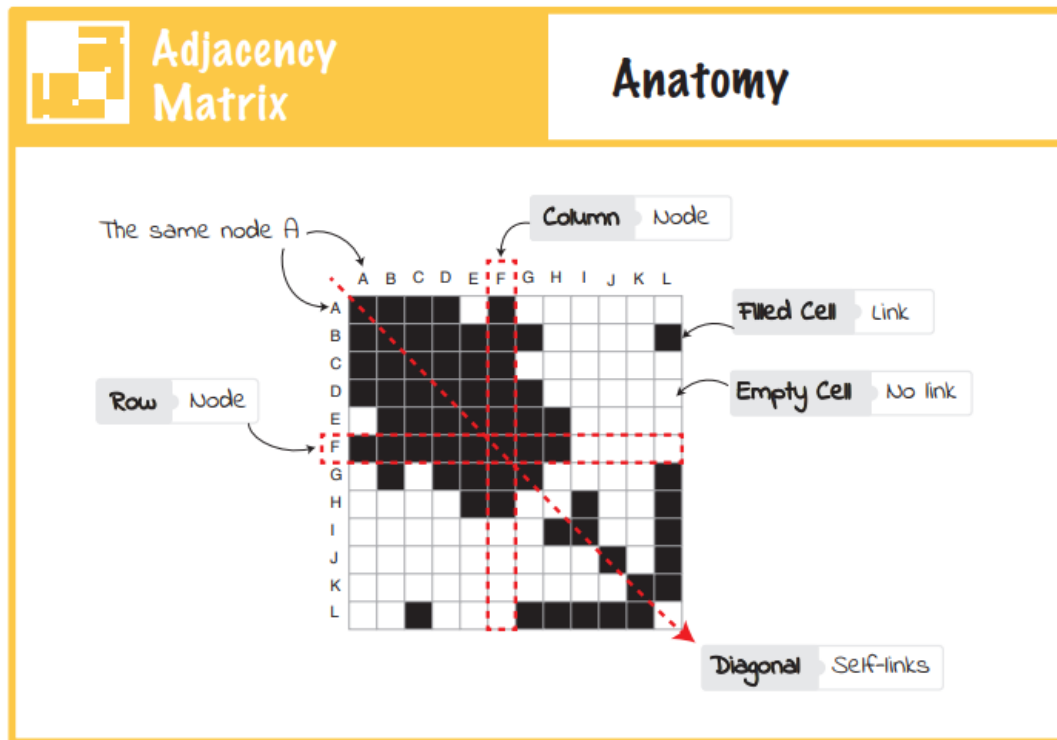
Anatomy



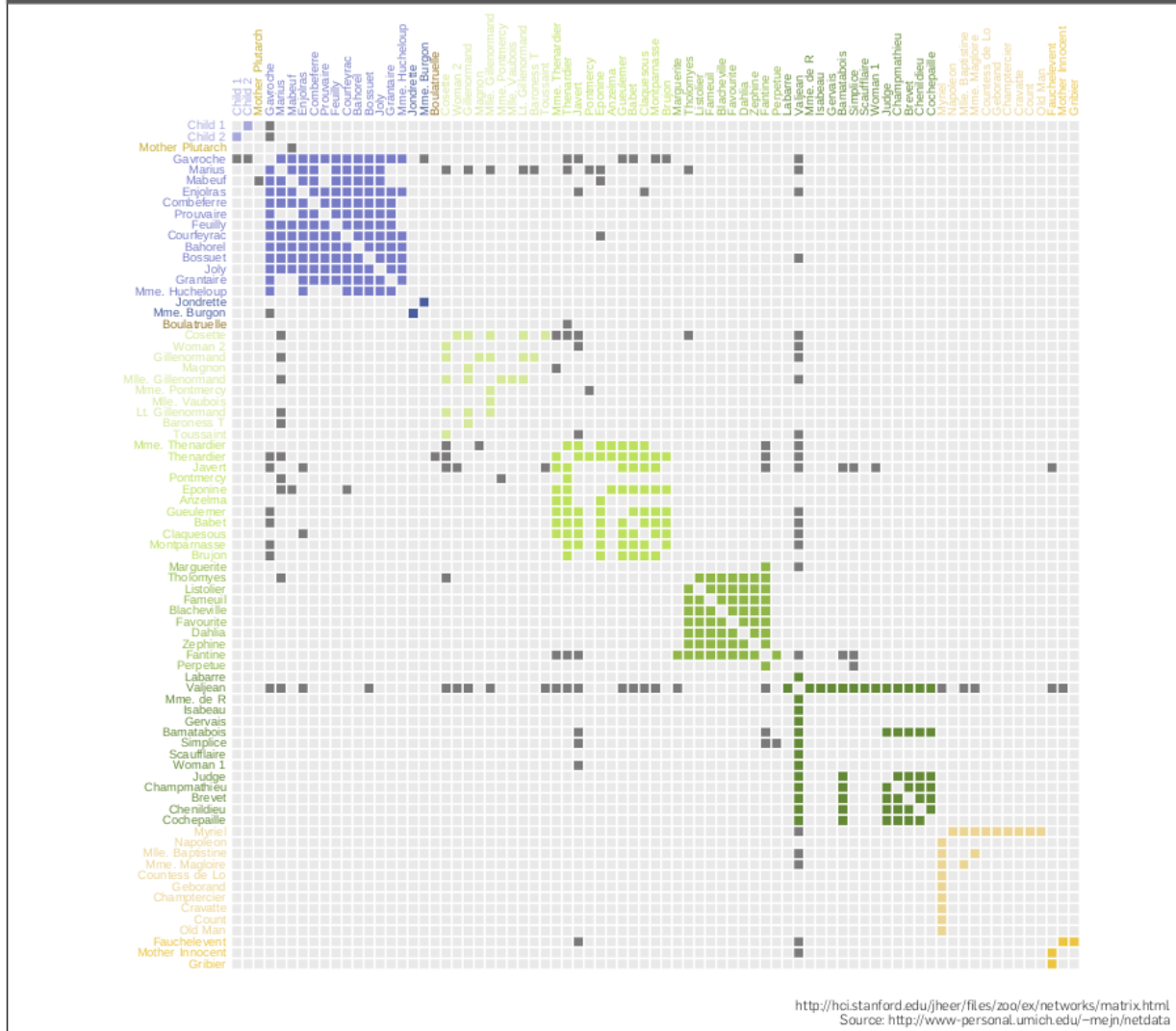
A La desventaja de los diagramas de arco es que no muestran estructura y conexiones entre nodos tan bien como la representación 2D y si hay muchas aristas puede ser difícil de leer.

Matrix View

Utilizamos la matriz de adyacencia de un grafo para representar la información. Es muy útil si las filas están bien ordenadas de modo que se vean bien los pequeños "clusters" de relación y es muy útil si se tiene grafos muy grandes y altamente conectados.



Networks: Figure 5c. Matrix view of *Les Misérables* character co-occurrences.



Useful reads:

- <https://www.researchgate.net/publication/338657160> Cheat Sheets for Data Visualization Techniques
- <https://datavizcatalogue.com/>