

Visualización de Información y Analítica Visual

Daniela Flores (<u>diflores@uc.cl</u>)
Hernán Valdivieso (<u>hvaldivieso@ing.puc.cl</u>)

Hernán Valdivieso

- Ingeniero Civil en Computación. Magíster en Ciencias de la Ingeniería (Visualización de Información e Inteligencia Artificial), Pontificia Universidad Católica de Chile.
- **2020 ahora:** Profesor Instructor diplomado UC: diplomado de Big Data y Ciencia de Datos, diplomado en Machine Learning Aplicado y diplomado en Inteligencia de Negocio.
- **2022 ahora:** Profesor Instructor pregrado UC: Visualización de información y Programación Avanzada.
- Página web: https://hernan4444.github.io/

Daniela Flores

- Magíster en Ciencias de la Ingeniería (Visualización de Información e Inteligencia Artificial), Pontificia Universidad Católica de Chile.
- **2015 ahora**: ayudante de cursos de pregrado y diplomado: Introducción a la Programación, Exploratorio de Computación, Programación como Herramienta para la Ingeniería, Visualización de Información, Minería de Datos.
- **2021 ahora:** *Software Engineer* en Fintoc, equipo de Infraestructura. Fintoc es una empresa que permite pagar en línea sin las comisiones de las tarjetas de débito y crédito.

Formalidades

Objetivos del curso

- Conocer conceptos fundamentales de abstracción de datos, percepción y procesamiento humano de la información.
- Entender y aplicar técnicas de reducción de la dimensionalidad, técnicas de visualización de datos espaciales, temporales y tabulares en el contexto de *Machine* learning.

Competencias

- Aplicar un proceso de diseño para crear visualizaciones efectivas.
- Utilizar principios de percepción y cognición humana en la visualización.
- **Diseñar** y **evaluar** de forma justificada la representación visual de un conjunto de datos determinado.
- Aplicar distintos métodos de visualización para un rango variado de datasets.
- Describir problemas éticos y de privacidad en el manejo de datos.

Material

- El *backbone* de este curso está basado en el trabajo de Tamara Munzner, a partir de su libro Visualization: Analysis & Design.
- Sin embargo, también toma material de...
 - o Interactive Data Visualization de Matthew O. Ward, Georges Grinstein, Daniel Keim.
 - The Visual Display of Quantitative Information de Edward Tufte.
 - Information Visualization: Perception for Design de Colin Ware.
 - Visualize This de Nathan Yau.
 - Algunos papers de visualización de información.

Nueve clases - Contenidos

- 1. Introducción y conceptos iniciales.
- 2. Abstracción de datos y tareas visuales
- 3. Decisiones de diseño y visualizaciones interactivas.
- 4. Principios de diseño y gráficos tabulares con Matplotib/Seaborn.
- 5. Teoría del color y visualización espaciales y de texto.
- 6. Privacidad de datos en la visualización y reducción de dimensión.
- 7. Streamlit Librería en Python para crear aplicaciones web con visualizaciones.
- 8. Validación de visualización, Grafos e Inteligencia Artificial explicable (XAI)
- 9. Examen

Nueve clases - Evaluaciones

- Publicar <u>taller evaluado 1</u> sobre Altair.
- 2. Resolver dudas sobre el <u>taller evaluado 1</u> sobre Altair.
- 3. Publicar <u>control evaluado</u> sobre contenidos de clase 2 y 3.
- Publicar <u>taller NO evaluado</u> sobre uso de Matplotlib/seaborn y crítica de visualización.
- 5. Publicar <u>taller evaluado 2</u> sobre contenidos de clase 5
- 6. Publicar <u>taller evaluado 3</u> sobre contenidos de clase 6
- 7. Publicar <u>taller NO evaluado</u> sobre uso de Streamlit.
- 8. Resolver dudas del <u>taller NO evaluado</u> sobre uso de Streamlit.
- 9. **Examen**

Evaluaciones

- En total serán **3 talleres evaluados, 1 control de alternativas**, 2 talleres sin nota y **un examen**,
- Los talleres se publican un jueves y se entregan el sábado de la semana siguiente a las 23:59.
- Para el control tendrán 4 días para responder en la plataforma.
- El examen será la última clase y tendrán 4 días para entregar. **Tendremos una** clase completa para trabajar en él.
- La calificación final del curso se calcula como el promedio de las 4 mejores notas.

Evaluaciones - Importante

- Todo se entrega en el <u>sitio del curso</u>. No se aceptarán entregas por correo.
- Todas las entregas de código son en IPYNB. En caso de entregar en otro formato, tendrán 10 décimas de descuento en la evaluación.
- Las entregas que sean en parejas deben incluir los nombres de todos los integrantes en la portada. Si el nombre no está, entonces **no realizó la actividad**.
- No se aceptan entregas posteriores a las 24 horas de atraso.
- No se aceptan más de una respuesta por actividad. En caso de entregar más de una respuesta, se revisará la primera.

Comunicación

- En clases.
- Dudas administrativas vía correo electrónico a ambos profesores <u>diflores@uc.cl</u> y <u>hvaldivieso@ing.puc.cl</u>. El asunto debe partir con "[MIA]".
 - [MIA] Cálculo nota final
 - o [MIA] Envío certificado médico
- Dudas de materia o evaluaciones, mediante el foro del curso.

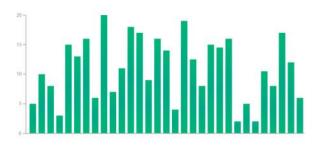
Clase 1: Introducción y conceptos iniciales

Contenidos

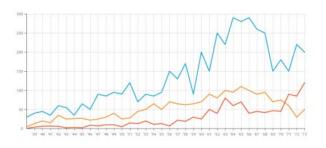
- Introducción
- Percepción, memoria y visión
- Marcas y canales
- Eficiencia de canales

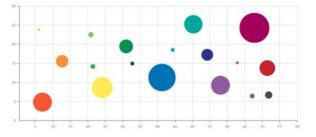
Introducción

¿Qué es la visualización?









¿Qué es la visualización?

Según los investigadores del área:

- "Transformación de lo simbólico a lo geométrico". [McCormick et al. 1987]
- "[...] encontrar la **memoria artificial** que mejor apoya a nuestros medios naturales de **percepción**." [Bertin 1967]
- "El uso de representaciones visuales de datos, generados por computador, interactivos, para amplificar nuestra cognición." [Card, Mackinlay, & Shneiderman 1999]

¿Qué es la visualización de información?

Según Robert Spence, 2014

Actividad cognitiva, facilitada por representaciones visuales externas donde personas construyen una representación mental interna del mundo.

¿Qué es la visualización de información?

Según Munzner, 2014:

Sistemas de visualización computarizado que brindan una **representación visual de los datasets** (conjunto de datos) que están diseñados para ayudar a las personas a **realizar tareas más eficazmente**

¿Qué es la visualización de información?

Se identifican 3 temas claves:

- Representación de un concepto abstracto, como datos.
- Uso de imágenes visuales.
- Ayudar a personas: a entender, a convencer, a realizar acciones eficientemente, etc
 . . .

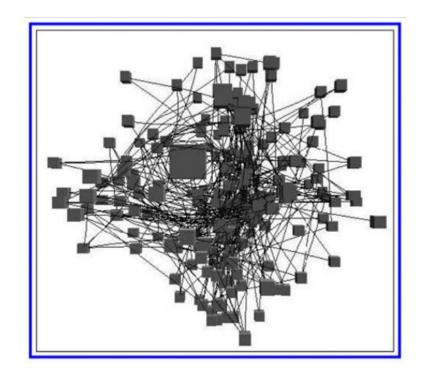


¿Hacer gráficos?



¿Hacer gráficos?





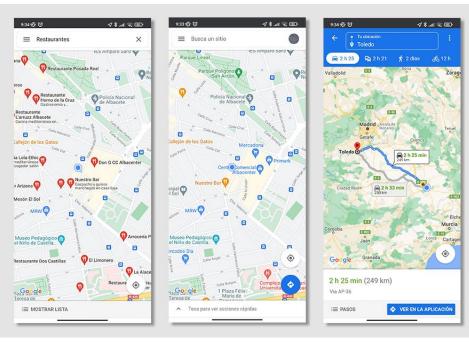
No solamente es hacer gráficos, hay todo un proceso detrás:

- 🤔 ¿Por dónde parto?
- 😕 ¿Hacia dónde voy?
- Hay mejores opciones que otras?
- 🤔 ¿En 3D se verá mejor?
- Bara debería enfocar en la efectividad?
- 😕 ¿En qué me enfoco al validar una visualización?

¿Qué tienen de bueno las visualizaciones?

Para empezar, nosotros generamos muchos datos...





¿Qué tienen de bueno las visualizaciones?

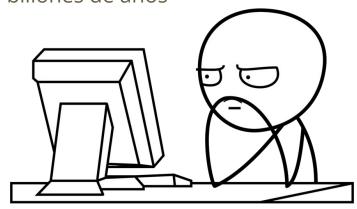
Para empezar, nosotros generamos muchos datos... ¿Cuantos datos?

¿Qué tiene de bueno las visualizaciones?

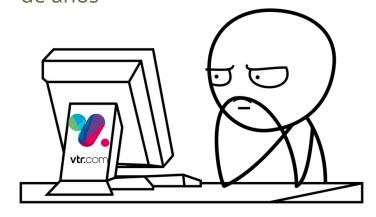
Para empezar, nosotros generamos muchos datos... ¿Cuántos datos?

Para el 2025 se estiman 175 zetabytes ...

Descargando a 25MB/s toma 1,8 billones de años



Con VTR (9MB/s) toma 9 billones de años



La visualización es un poderoso aliado

- Comprender las relaciones entre conjuntos de datos
- Entender algo sobre los datos
- Resaltar información importante
- Plantear un argumento convincente
- A nadie le gusta leer "datos crudos"
- Podemos encontrar outliers
- Podemos entretener o inspirar

Casos de usos

Puedes encontrar uso de visualizaciones para...

- Explorar o entender un *dataset*
- Explicar resultados de un modelo de ML.
- Calibrar modelos de ML.
- Comparar y seleccionar modelos de ML.
- Enseñar conceptos de ML.

Lectura recomendada

Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers

Fred Hohman, *Member, IEEE*, Minsuk Kahng, *Member, IEEE*, Robert Pienta, *Member, IEEE*, and Duen Horng Chau, *Member, IEEE*

Ejemplo de casos - Chihuahua o muffin



Ejemplo de casos - John Snow

En 1854, durante una epidemia de Cólera en Londres, el Dr. John Snow usa un análisis espacial para **apoyar su hipótesis.**

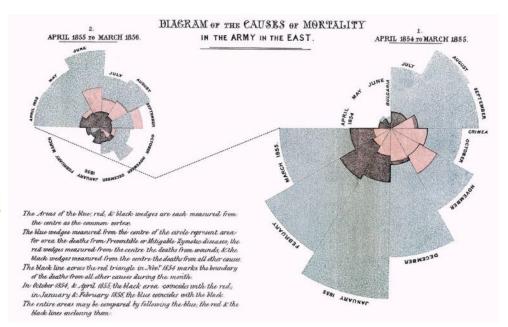
Encontró que las muertes eran principalmente por una bomba de agua contaminada.



Ejemplo de casos - Florence Nightingale

En 1858, Florence Nightingale, dibujó este gráfico para ilustrar las **causas de la mortalidad** de los soldados en el hospital militar que dirigía durante la guerra de Crimea.

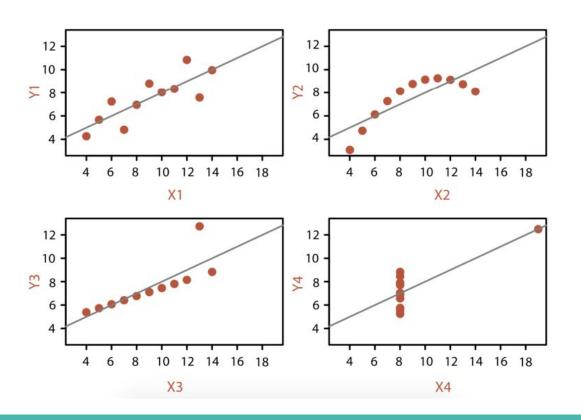
Se dio cuenta de que el 80% de los muertos eran víctimas de los deficientes tratamientos sanitarios.



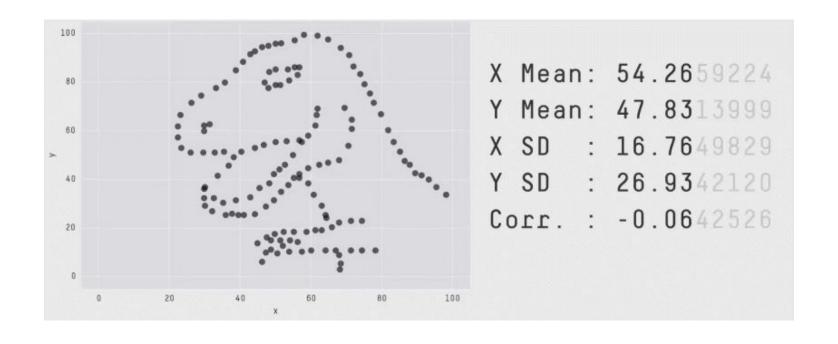
Ejemplo de casos - Anscombe's Quartet (1973)

9			3					
	1		2		3		4	
	Χ	Υ	Χ	Υ	Χ	Υ	Χ	Υ
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

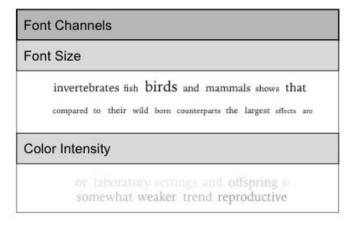
Ejemplo de casos - Anscombe's Quartet (1973)

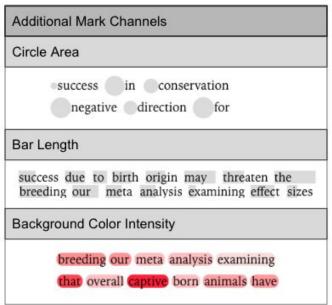


Ejemplo de casos - Anscombe's Quartet (2007)

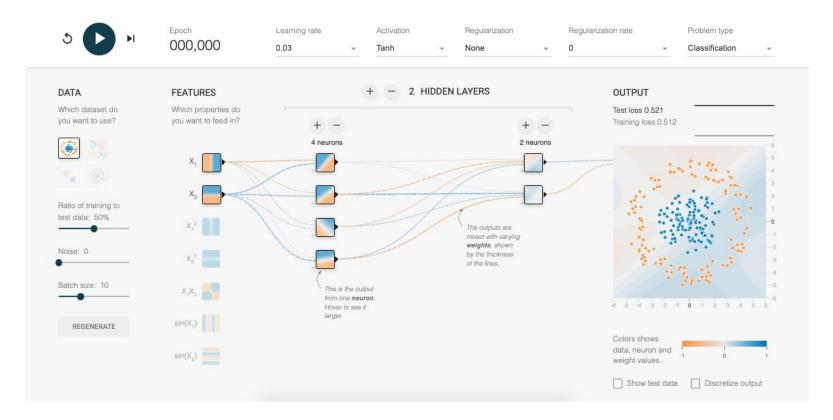


Ejemplo de casos - Palabras más relevantes para un modelo de *Deep Learning*





Ejemplo de casos - Jugar con una red neuronal



37

Casos de usos

En este curso, nos enfocaremos más en...

- Explorar o entender un dataset
- Explicar resultados de un modelo de ML.
- Calibrar modelos de ML.
- Comparar y seleccionar modelos de ML.
- Enseñar conceptos de ML.

Percepción, memoria y visión

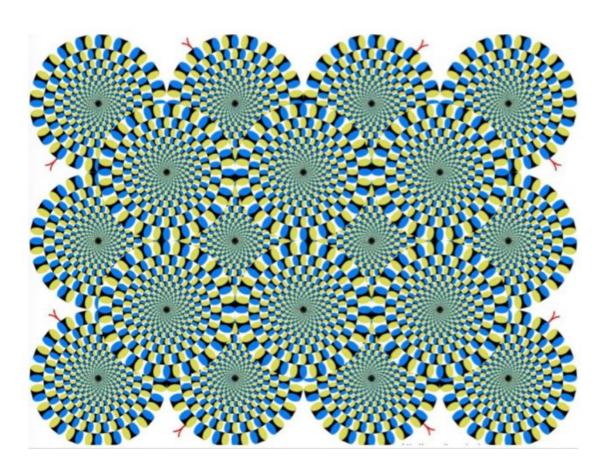
Veremos temas como:

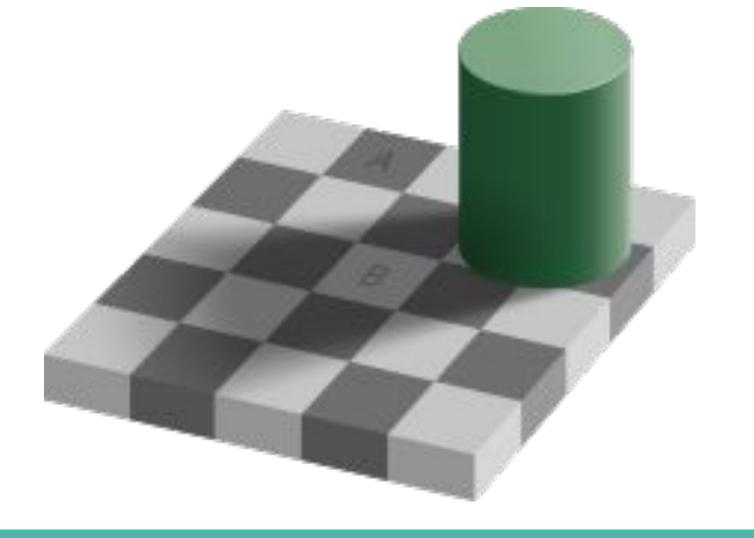
- Leyes de Gestalt
- Efecto contraste
- Estimación de magnitud
- Visión

La percepción es la manera en la que el cerebro de un organismo **interpreta los estímulos** sensoriales que recibe a través de los sentidos para formar una impresión consciente de la realidad física de su entorno.

• La percepción es un proceso fuertemente **afectado** por el **aprendizaje**, la **memoria** y la **atención**.

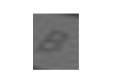
• El estudio de ilusiones e imágenes ambiguas ha demostrado que nuestros cerebros intentan (de forma subconsciente, incluso) darle **sentido** al *input* que **recibimos**.









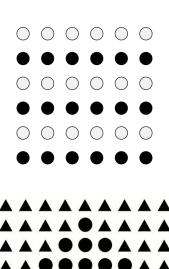


Percepción - *Grouping (Gestalt Laws)*

Esta teoría, que fue desarrollada por psicólogos alemanes de los años 1920, intenta describir cómo percibimos a través de grupos. Para lograr esto, nuestro cerebro aplica varios principios

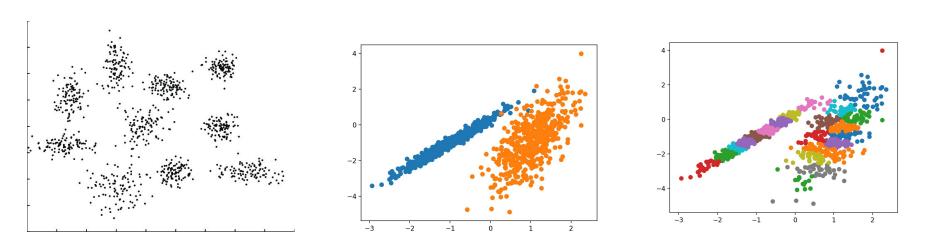




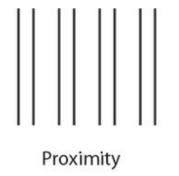


Proximidad

Similaridad



Y hay más..



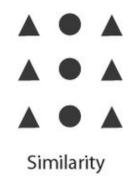
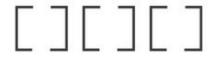
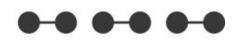




Figure-ground





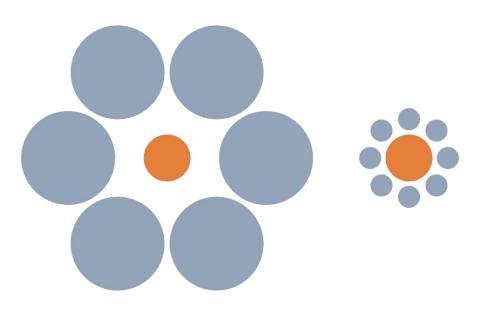


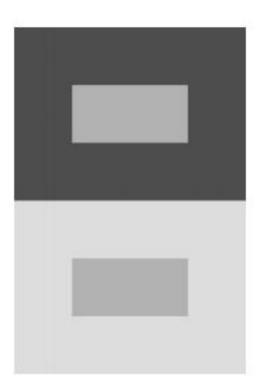
Closure

Connectedness

Link de interes: https://www.usertesting.com/blog/gestalt-principles

Contrast effects

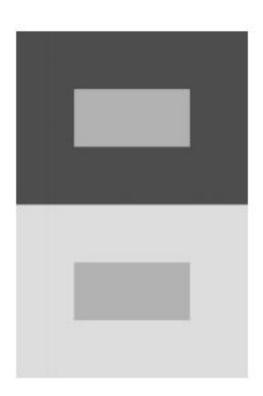




Contrast effects

La forma en que percibimos un objeto depende de los **objetos que lo rodean o lo visto en un instante anterior.**

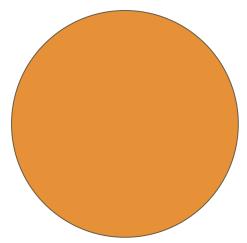
En visualización, hay canales (como el color y tamaño) que se ven afectados fuertemente por este efecto y hace que su comparación no sea tan efectiva.



Estimación de magnitud

¿Cuántas veces es más grande el círculo de la derecha? (En términos del área)















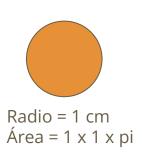






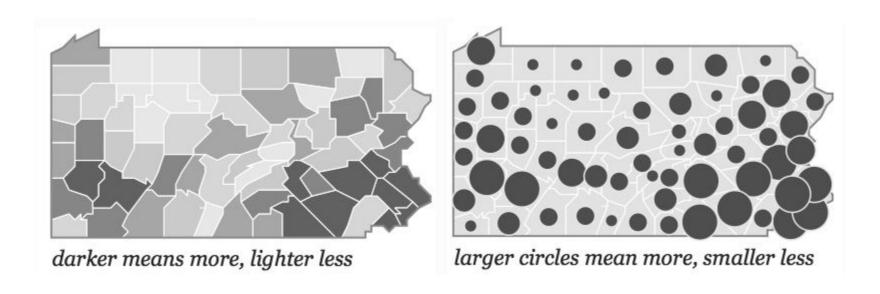
Estimación de magnitud

¿Cuántas veces es más grande el círculo de la derecha? (En términos del área)

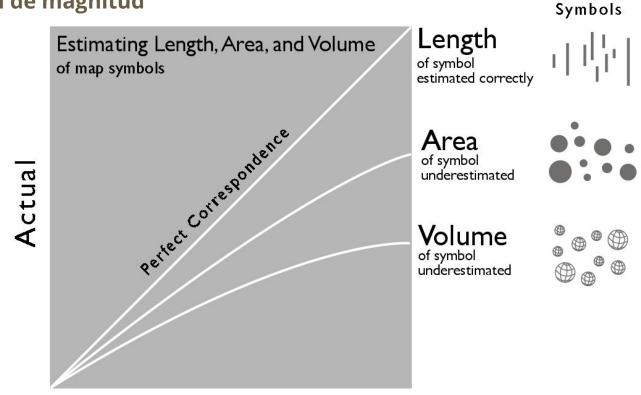




Estimación de magnitud

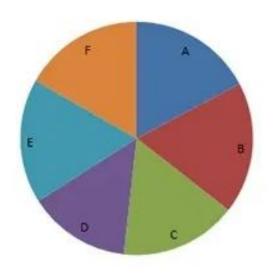


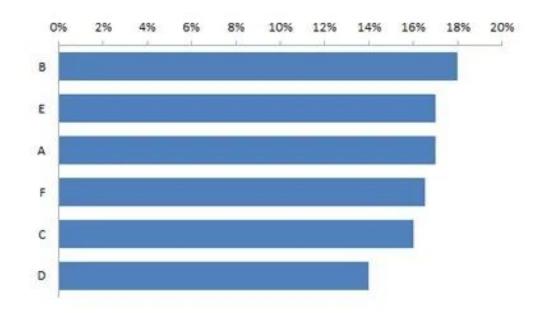
Estimación de magnitud



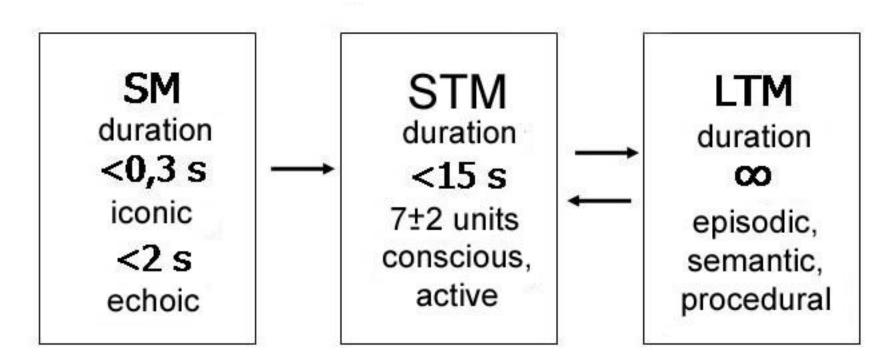
Perceived

Pie Chart VS Bar Chart



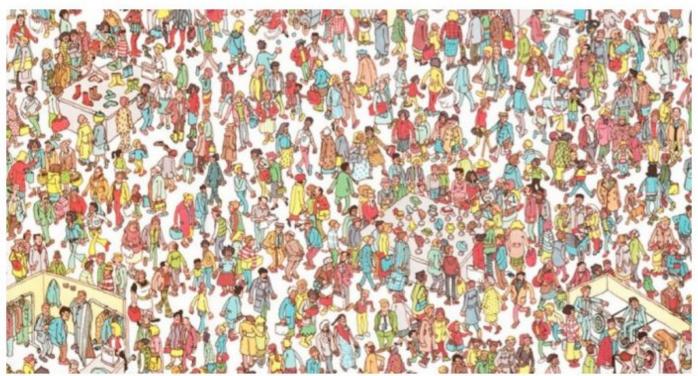


Estructura de la memoria y sus procesos

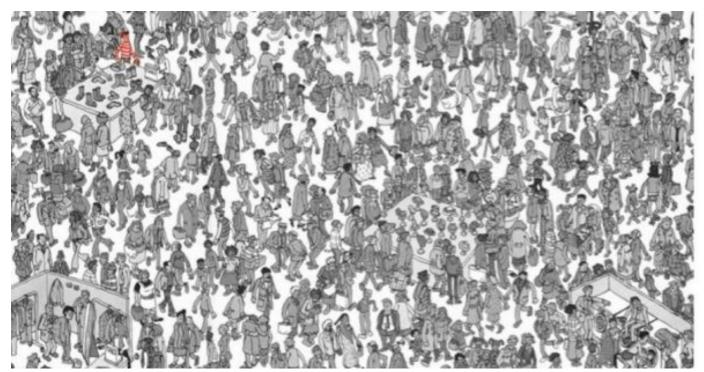


¿Dónde está Wally?

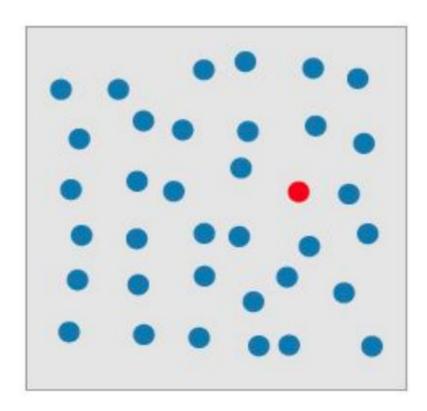




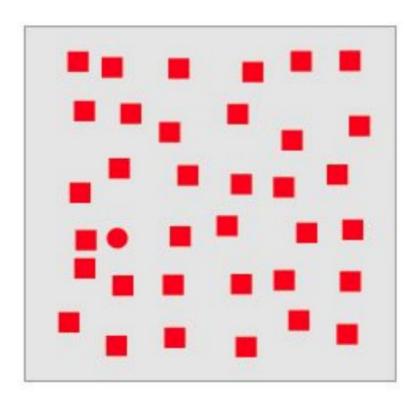
¿Dónde está Wally?



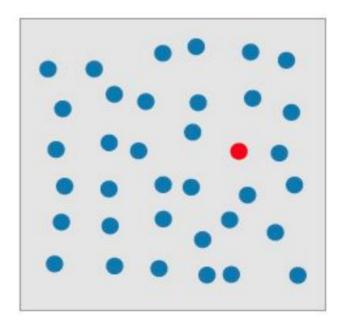
Busca el círculo rojo

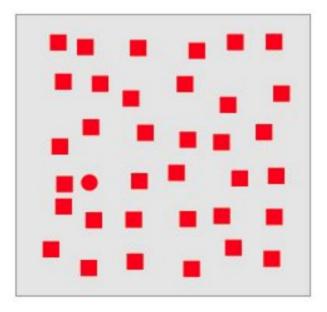


Busca el círculo rojo



Color V/S Forma.





Memoria sensorial

- Se procesan rápidamente los estímulos en forma paralela.
- Existen propiedades visuales que pueden ser procesadas en forma pre-atentiva.
- En el diseño de visualizaciones e interfaces estas propiedades pueden aprovecharse para reducir los tiempos de procesamiento del usuario.

AEPPUOTKXTQIEFB

A E P P U O T K X T Q I E F B

S O L L U N A T I E R R A V E R D E

AEPPUOTKXTQIEFB

A E P P U O T K X T Q I E F B

S O L L U N A T I E R R A V E R D E

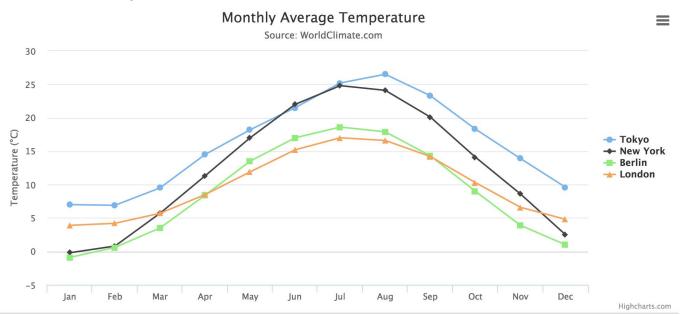
SOL LUNA TIERRA VERDE

Memoria de corto plazo

- En este sistema la información se procesa de forma activa y consciente.
- Las personas solo podemos <u>recordar 7 ± 2 pedazos de información</u>. (Ley de George Miller, 1956)
- Nos permite estimar y comparar.
- En visualización, es necesario tener en cuenta esta capacidad de recordar información para no saturar al usuario con tanta información a presentar en una visualización.

Memoria de corto plazo

• No esperemos que el usuario memorice todo de la visualización, sino los elementos más importantes.



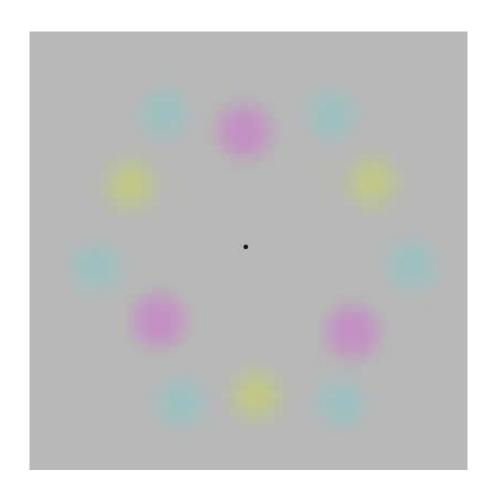
Memoria de largo plazo

- Consolidación por práctica y asociación semántica
- Por ejemplo: Leer, escribir, operaciones básicas

Visión

Si fijamos los ojos en el punto central de la figura. Lo esperado es que en un momento dejemos de ver los colores alrededor.

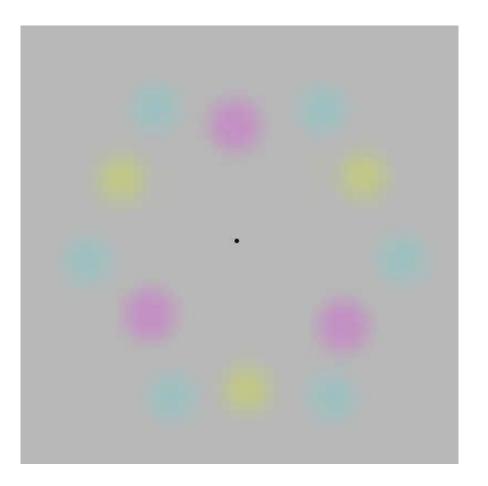
Esto ocurre por la **fatiga ocular.**



Visión

Algunas causas de la **fatiga ocular** son:

- Disminución de la frecuencia del parpadeo.
- Lugar de trabajo mal iluminado, resolución de la pantalla deficiente, reflejos y brillo excesivos.
- Estrés.

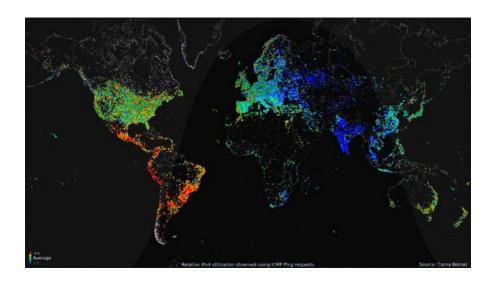


Visión - Fatiga ocular

En visualización, hay que tener cuidado con no provocar la fatiga retinal al usuario.

Para enfrentar este problema, se aconseja:

 Controlar el uso de animaciones y la cantidad de información que se transmite simultáneamente.



Visión - Fatiga ocular

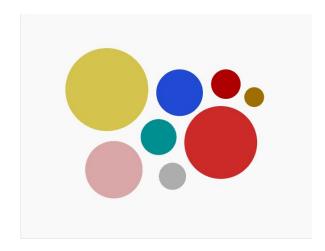
En visualización, hay que tener cuidado con no provocar la fatiga retinal al usuario.

Para enfrentar este problema, se aconseja:

 Se aconseja utilizar colores de tonos pastel en lugar de colores vivos como el rojo, fucsia, amarillo y naranja en una misma gráfica.



NOT IDEAL



BETTER

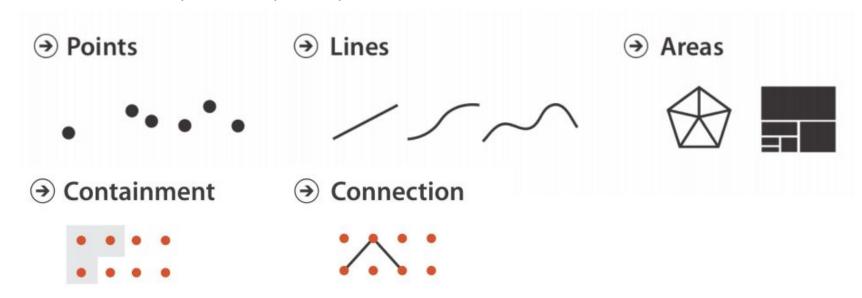
Percepción, memoria y visión

Resumen

- Definición
- Principios de Gestalt
- Efecto Contraste
- Estimación de Magnitud
- Memoria
- Visión

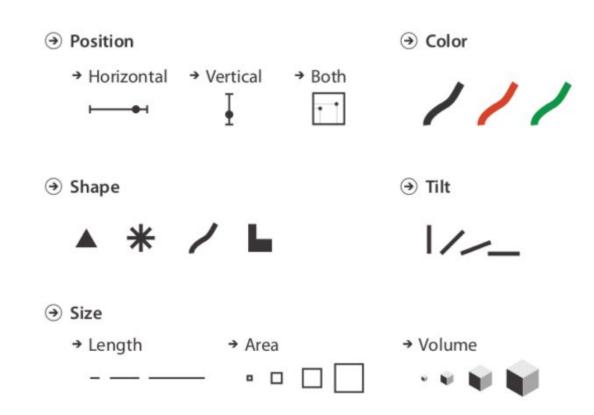
Marcas

• Elemento **geométrico básico**, que puede ser clasificado según el número de dimensiones espaciales que requiera.



Canal

 Permite controlar la apariencia de las marcas, independientemente de la dimensionalidad de este elemento primitivo



Marcas y canales - Ejemplo

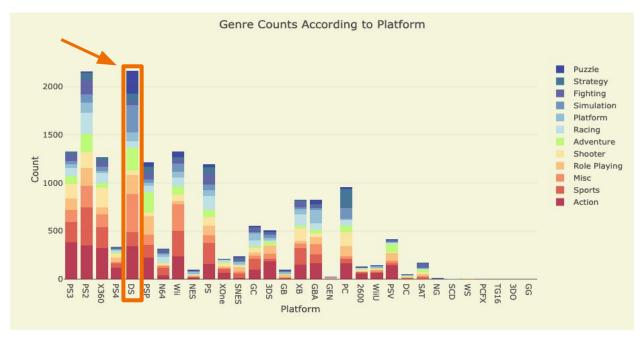
Nuestra información la vamos a representar por un área circular (marca).



Podemos personalizar (canales) estos puntos con:

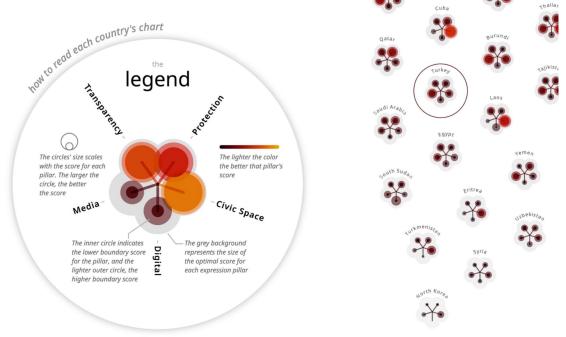
		Tamaño Borde	
•			

Glifos → Objeto gráfico **complejo** que representa algún dato, y con una estructura interna **formada por múltiples marcas** primitivas que también representan otra información.



Fuente: Video Game Sales Exploratory Data Analysis | Kaggle

Glifos → Objeto gráfico **complejo** que representa algún dato, y con una estructura interna **formada por múltiples marcas** primitivas que también representan otra información.



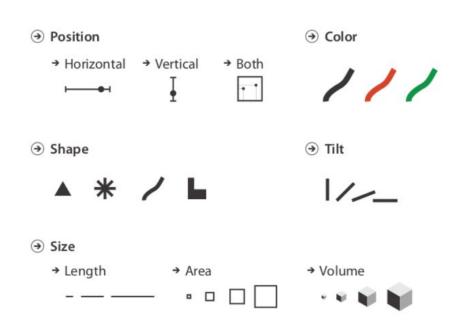
Fuente: The Freedom of Expression

Tipos de canales

- El sistema de percepción humano tiene dos tipos de modalidades:
 - El identity channel permite discernir información sobre qué es algo o dónde se encuentra.
 - El magnitude channel, por otra parte, nos permite saber cuánto de ese algo existe.
- Con estas dos modalidades, podemos saber, por ejemplo:
 - o ¿Qué figura es? ¿Un círculo, un triángulo, una cruz o un heptágono? [identity]
 - ¿De qué hue/matiz es? ¿Rojo, verde, caqui o gris? [identity]
 - ¿Cuánta saturación tiene ese azul? ¿Celeste, azul marino o turquesa? [magnitude]
 - ¿Dónde está? ¿En qué región se encuentra la marca? [identity]
 - ¿Qué tan larga es aquella línea con respecto a esta? [magnitude]
 - ¿Cuánto espacio hay entre ambos rectángulos? [magnitude]

¿Cómo usarlos?

- No todos los canales son iguales: los mismos datos codificados con dos canales visuales distintos resultará en información diferente.
- Dos principios guían el uso de canales visuales: expresividad y efectividad



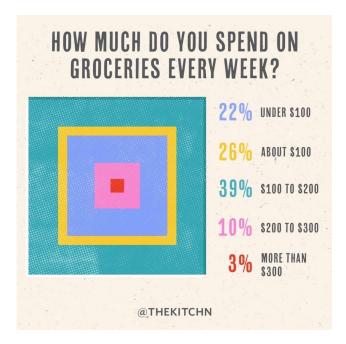
Principio de expresividad

- Debe haber **coherencia** entre el tipo de canal (magnitud, identidad) con la semántica del atributo (cuantitativo, ordinal, categórico).
 - Los datos ordenados deben ser mostrados de tal forma que nuestro sistema perceptual los perciba como ordenados; inversamente, debe ocurrir lo mismo con los datos no ordenados.

Principio de efectividad

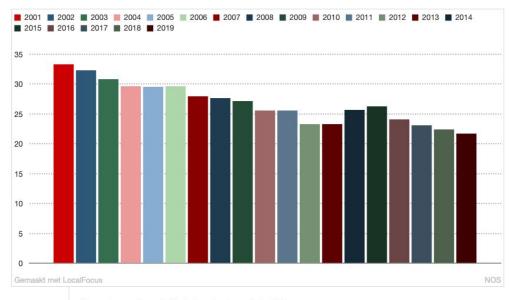
 Dicta que los atributos más importantes deben ser codificados con los canales más efectivos, para que sean más perceptibles

Marcas y canales - Ejemplos



El uso de área no el más efectivo en esta situación

No cumple el **principio de efectividad**



Percentage rokers in Nederland volgens het CBS

Los años presentan un orden, pero la matiz de color no. ¿Por qué el verde va después que el azul y el rojo?

85

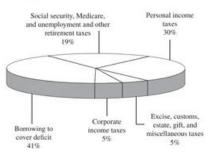
No cumple el **principio de expresividad**

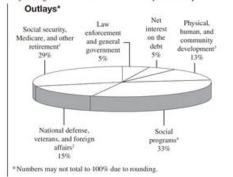
Fuente: https://viz.wtf/

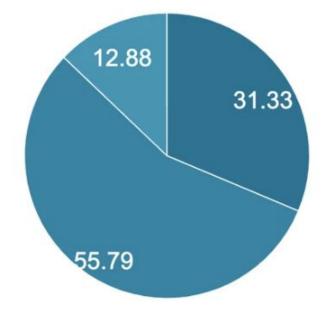
Marcas y canales - Ejemplos

Major Categories of Federal Income and Outlays for Fiscal Year 2021

Income and Outlays. These pie charts show the relative sizes of the major categories of federal income and outlays for fiscal year 2021. Income







Public

Private nonprofit

Private for-profit

El uso de volumen + perceptiva no el más efectivo para comparar los valores

No cumple el **principio de efectividad**

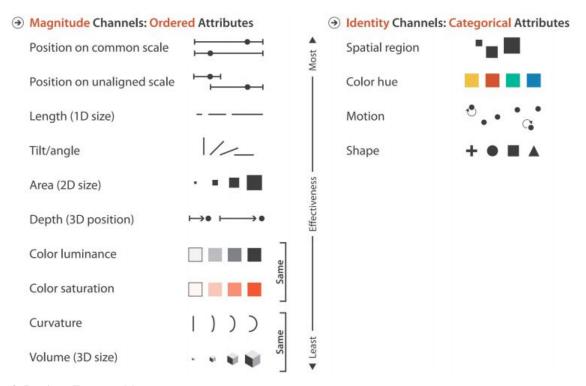
Se ocupa una gama de azules para 3 datos categóricos que no tienen una relación de orden.

No cumple el **principio de expresividad**

Fuente: r/dataisugly (izquierda) y r/dataisugly (derecha)

Eficiencia de canales

Según el tipo de canal, se construyó un ranking de estos

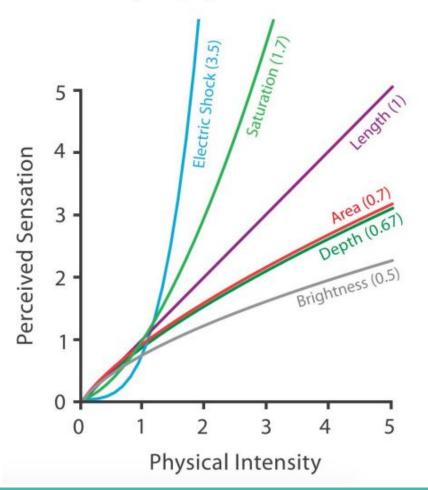


- Para analizar el espacio de encodings posibles, hay que entender ciertas características de estos canales visuales.
 - ¿Cómo se justifica este ranking?
 - ¿Por qué hay canales mejores que otros?
 - ¿Cuánta información puede codificar un canal?
 - ¿Pueden ser usados de forma independiente o podría haber interferencia entre ellos?
- Responderemos a estas preguntas, estudiando ciertos criterios:
 - El criterio de *accuracy*.
 - El criterio de *discriminability*.
 - El criterio de separability.
 - La habilidad de ofrecer visual popout

Accuracy (Stevens's power law 1975)

Modificación de un canal VS cuánto se **percibe** el cambio en dicho canal

Steven's Psychophysical Power Law: S= I^N

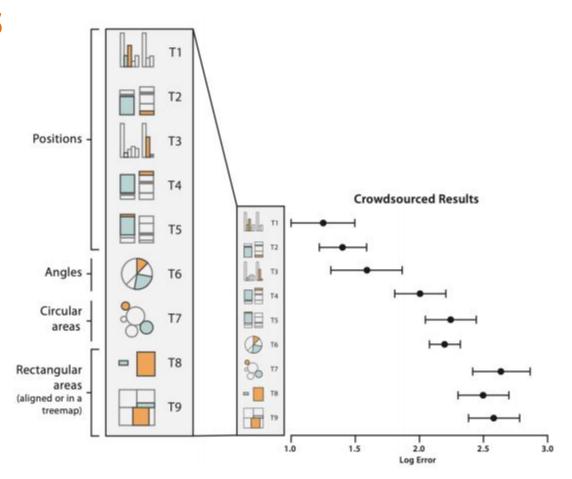


Accuracy

Cleveland & McGill (1984) Heer & Bostock (2010)

Estudio para clasificar canales visuales empíricamente por su efectividad para transmitir valores cuantitativos.

2010 es el mismo estudio de 1984 pero realizado en <u>Amazon</u> <u>Mechanical Turk</u>

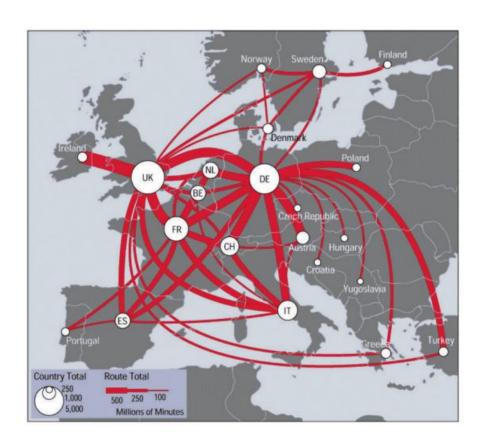


Discriminability

Es importante considerar cuánta información puede codificar un canal.

Es importante definir cuántos *bins* están disponibles para ser usados en un canal visual, en donde cada bin es un paso (o nivel) distinguible del anterior o siguiente.

Ejemplo: Ancho de línea

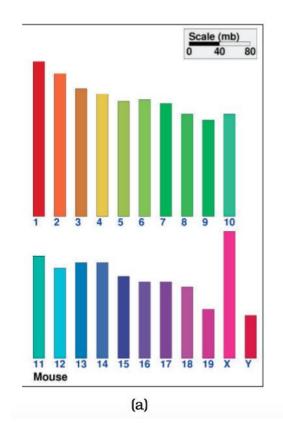


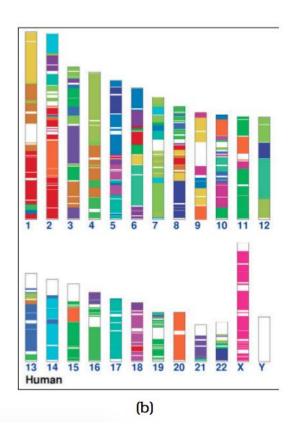
Discriminability II

Ejemplo: Colores diferentes

Página recomendada

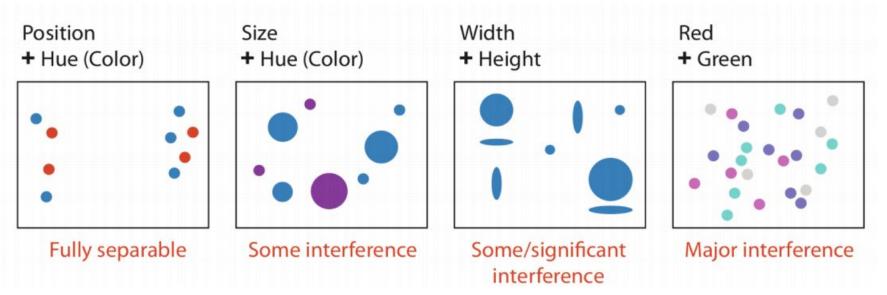
ColorBrewer





Separability

No es posible tratar a los canales de forma independiente, puesto que generalmente tendremos dependencias e interacciones entre ellos.

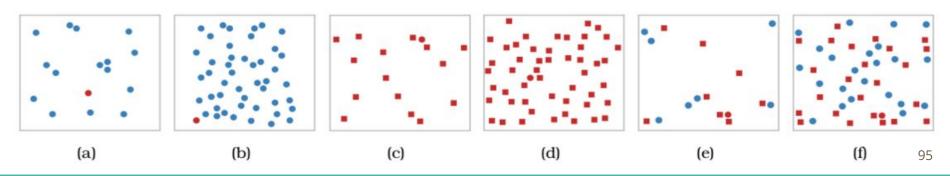


Visual popout

Muchos canales ofrecen un efecto de *popout*, donde un elemento distinto se diferencia de forma inmediata (recordemos la sección preatentiva).

Ser un canal con buen "visual popout" es que el tiempo que nos toma encontrar el objeto diferente (casi) no depende de la cantidad de los distractores.

¿Dónde está el punto **rojo**?



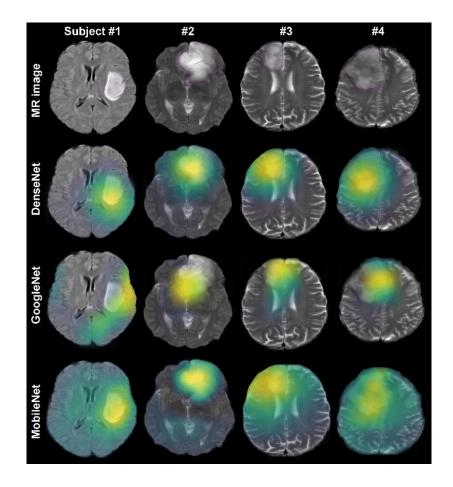
Visual popout

Indicar zonas donde las personas más hacen clicks o pasan el *mouse*.



Visual popout

Entregar el resultado de donde un modelo de *Deep Learning* identificó un tumor.

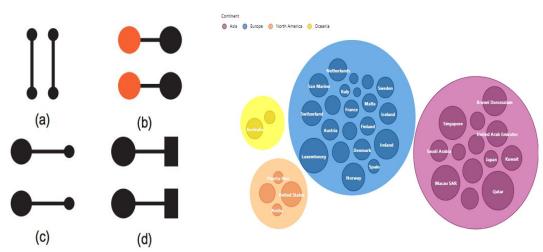


Grouping

Patrones visuales que permitan agrupar ítems (Categorías).

Aquí nos apoyamos en los **Principios de Gestalt** donde los siguientes canales son los más efectivos para agrupar ítems.

- 1. Conectividad y contención
- 2. Proximidad
- 3. Similaridad

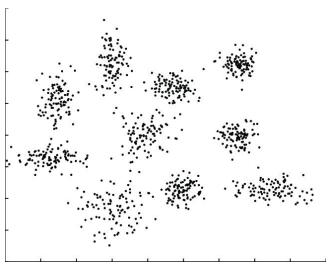


Grouping

Patrones visuales que permitan agrupar ítems (Categorías).

Aquí nos apoyamos en los **Principios de Gestalt** donde los siguientes canales son los más efectivos para agrupar ítems.

- 1. Conectividad y contención
- 2. Proximidad
- 3. Similaridad

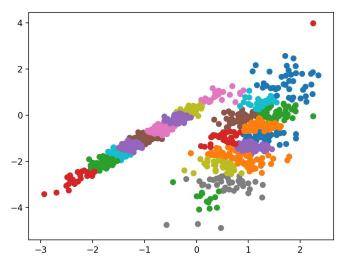


Grouping

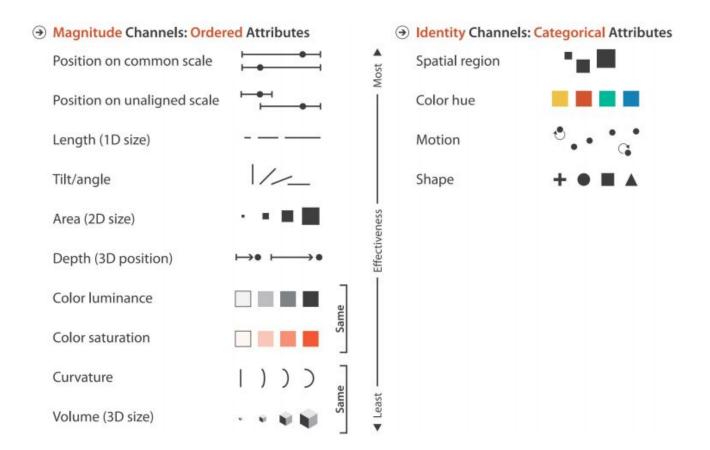
Patrones visuales que permitan agrupar ítems (Categorías).

Aquí nos apoyamos en los **Principios de Gestalt** donde los siguientes canales son los más efectivos para agrupar ítems.

- 1. Conectividad y contención
- 2. Proximidad
- 3. Similaridad



Tipos de expresividad y ranking de efectividad



Taller 1

Presentación de enunciado



Visualización de Información y Analítica Visual

Daniela Flores (<u>diflores@uc.cl</u>)
Hernán Valdivieso (<u>hvaldivieso@ing.puc.cl</u>)