

Resum Esquemàtic de la Teoria per a l'Examen de Visualització de Dades

Aquest informe presenta un resum esquemàtic i detallat dels conceptes fonamentals en la visualització de dades, organitzats per facilitar la comprensió i la recuperació ràpida d'informació clau per a finalitats d'estudi. S'aborden les tècniques de processament de dades, la visualització multivariable, els principis del color i la precisió visual, l'honestat en la representació, l'ús de mapes, la distinció entre visualitzacions 2D i 3D, les directrius per a la visualització científica i l'art del *storytelling* amb dades, així com la interactivitat i l'experiència d'usuari.

I. Reducció de Dimensionalitat i Processament de Dades

La reducció de dimensionalitat (DR) és un procés essencial en la visualització de dades, que busca transformar conjunts de dades d'alta dimensió a un espai de menor dimensió, conservant la majoria dels coneixements significatius de les dades originals.¹ L'objectiu primordial de la DR és preservar l'estructura significativa d'un conjunt de dades, utilitzant menys atributs per representar els elements.¹ Aquesta estratègia és fonamental per gestionar la complexitat en la visualització, permetent l'exploració de dades i l'aprenentatge automàtic.¹ En reduir les dimensions a 2D o 3D, es facilita la visualització de les mostres, la qual cosa al seu torn permet detectar clústers i identificar valors atípics.¹

Les tècniques generals de DR es divideixen en diverses categories: la **Selecció de Característiques** (*Feature Selection*), que implica triar un subconjunt de característiques rellevants per representar el conjunt de dades de manera òptima; l'**Eliminació de Característiques** (*Feature Elimination*), que redueix l'espai de característiques descartant variables; i l'**Enginyeria de Característiques** (*Feature Engineering*), que transforma les dades brutes en característiques més representatives.¹ Dins la selecció de característiques, s'inclouen mètodes com el Missing Value Ratio, Low Variance Filter, High Correlation Filter, Random Forest, Backward Feature Extraction i Forward Feature Selection.¹ Altres tècniques de DR es

basen en components o factors (com Factor Analysis, Principal Component Analysis - PCA, i Independent Component Analysis) o en projecció (com ISOMAP, t-SNE i UMAP).¹

Un aspecte important a considerar és la compensació inherent en la reducció de dimensionalitat per a la visualització. Les tècniques de DR busquen preservar la informació, però la forma en què ho fan varia. Per exemple, PCA prioritza la variància global, mentre que t-SNE s'enfoca en l'estructura local. Aquesta diferència implica que optimitzar una d'aquestes propietats pot comprometre l'altra. Si l'objectiu és identificar patrons amplis i valors atípics, PCA podria ser més adequat. No obstant això, per trobar clústers densos i relacions de veïnatge, t-SNE és superior, encara que amb l'avertència que les relacions globals entre clústers en t-SNE poden manca de significat. L'elecció de la tècnica de DR, per tant, no és arbitrària i ha d'alinejar-se amb la pregunta analítica específica que es busca respondre, la qual cosa impacta directament en l'honestedat i precisió de la visualització resultant.

Anàlisi de Correlació i Correlogrames

Quan es treballa amb més de 3 o 4 variables quantitatives, la Scatterplot Matrix (SPLOM) es torna difícil de manejar.¹ En aquests escenaris, resulta més útil quantificar l'associació entre parells de variables i visualitzar aquestes quantitats en lloc de les dades brutes.¹ Per a això, s'utilitza el

Coefficient de Correlació (R), que es calcula mitjançant la fórmula $R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$.¹ El valor de R oscil·la entre -1 i 1, indicant el grau de correlació entre dues variables.¹ Un R igual a 0 significa que no hi ha associació, mentre que un R d'1 o -1 indica una associació perfecta.¹ Si R és major que 0, les variables estan correlacionades positivament (valors majors en una coincideixen amb majors en l'altra); si R és menor que 0, estan anticorrelacionades (valors majors en una coincideixen amb menors en l'altra).¹

Els **Correlogrames** són una eina visual per representar l'estructura de correlació de les variables d'entrada.¹ La intensitat del color i la forma de l'el·lipse en un correlograma estan directament vinculades al coeficient de correlació.¹ Per convenció, les correlacions positives ($R > 0$) es mostren en blau, i les anticorrelacions ($R < 0$) es representen en vermell.¹ Un color més fosc o intens i una el·lipse més estreta indiquen

una correlació més forta, ja sigui positiva o negativa.¹

Variància, Covariància i Matriu de Covariància

Abans d'aprofundir en les tècniques de reducció de dimensionalitat, és crucial comprendre els conceptes de variància i covariància. La **Variància** representa la variació dels valors en una única variable, indicant com de dispersos estan els valors entre si.¹ La seva fórmula és

$\text{var}(x) = N^{-1} \sum (x_i - \bar{x})^2$.¹ A diferència de la variància, la

Covariància es calcula entre dues variables diferents i el seu propòsit és indicar com aquestes dues variables varien juntes.¹ La fórmula de la covariància és

$\text{cov}(x,y) = N^{-1} \sum (x_i - \bar{x})(y_i - \bar{y})$.¹ La covariància pot variar entre $-\infty$ i $+\infty$ i el seu signe indica la direcció de la relació lineal entre les variables.¹ És important destacar que la covariància és zero per a variables independents.¹

La distinció entre covariància i correlació és fonamental: la covariància mesura l'extensió i direcció en què dues variables canvien en conjunt i és afectada pel canvi d'escala, mentre que la correlació mesura tant la força com la direcció de la relació lineal i no és influenciada pel canvi d'escala, amb un rang entre -1 i 1.¹

La **Matriu de Covariància** s'utilitza per representar els valors de covariància de cada parell de variables en dades multivariades.¹ En aquesta matriu simètrica, la diagonal mostra la variància de cada variable (la covariància entre la mateixa variable amb si mateixa).¹ Els valors dins la matriu de covariància proporcionen informació sobre la magnitud i direcció de la distribució de dades multivariades en un espai multidimensional.¹ Una covariància positiva indica que els valors de les variables tendeixen a augmentar o disminuir junts; una covariància negativa, que una variable tendeix a augmentar mentre l'altra disminueix; i una covariància zero, que no hi ha relació lineal.¹

La comprensió d'aquests conceptes estadístics és un pont fonamental cap a la codificació visual efectiva. La visualització de dades no és merament una qüestió estètica; està profundament arrelada en la comprensió estadística. La forma en què es codifiquen visualment les relacions, com el color i la forma en els correlogrames o els eixos en els gràfics PCA, són representacions directes d'aquestes propietats

estadístiques subjacents. Una sòlida comprensió d'aquests fonaments estadístics és, per tant, crucial tant per crear visualitzacions precises com per interpretar-les correctament, ja que una mala interpretació de la variància o la correlació pot portar a una lectura errònia dels patrons visuals, la qual cosa subratlla la importància de la precisió visual.

Anàlisi de Components Principals (PCA)

L'**Anàlisi de Components Principals (PCA)** és una tècnica àmpliament utilitzada per examinar dades d'alta dimensió, operant sota un paradigma d'aprenentatge no supervisat.¹ Es recomana usar PCA quan es necessita reduir el nombre de variables sense poder identificar quines eliminar per complet, quan es desitja assegurar que les variables siguin independents entre si, o quan s'està còmode amb el fet que la variable independent resultant sigui menys interpretable.¹

El mecanisme de PCA consisteix a introduir un nou conjunt de variables, denominades **Components Principals (PCs)**, que són combinacions lineals de les variables originals, estandarditzades a mitjana zero i variància unitària.¹ Aquests PCs s'ordenen per variància: el PC 1 representa la direcció de la major variància de les dades, mentre que el PC 2 representa la major part de la variància restant ortogonal al PC 1. Aquest procés s'estén per obtenir el nombre de components necessaris per cobrir la quantitat de variància desitjada.¹

La relació de PCA amb la matriu de covariància és directa: els autovalors (*eigenvalues*) de la matriu de covariància representen la magnitud de la dispersió en la direcció dels components principals, i els autovectors (*eigenvectors*) mostren la direcció.¹ En casos on la covariància és zero (variables independents), els autovalors són iguals als valors de variància.¹

El procés de PCA segueix una sèrie de passos definits:

1. Prendre la matriu de característiques $X \in \mathbb{R}^{N \times P}$, on $N > P$.¹
2. Calcular el vector de mitjanes per a cada dimensió.¹
3. Calcular la matriu de covariància.¹
4. Calcular els autovectors i autovalors corresponents per a cada dimensió.¹
5. Ordenar els autovectors per autovalors decreixents i triar P autovectors amb els majors autovalors per formar una nova matriu $W \in \mathbb{R}^{P \times P}$.¹
6. Usar aquesta matriu d'autovectors per transformar les mostres al nou subespai:

$$Z=XW.^1$$

Entre les fortaleeses de PCA, es destaca que és un mètode molt interpretable, on cada PC està ben definit i és ortogonal a les altres dimensions. A més, permet obtenir la variància explicada per cada PC per seleccionar un nombre apropiat de dimensions.¹ No obstant això, PCA té debilitats importants: tendeix a ser altament afectat per valors atípics (*outliers*) i funciona millor només amb dades contínues.¹

Anàlisi Discriminant Lineal (LDA)

L'**Anàlisi Discriminant Lineal (LDA)** és una tècnica d'aprenentatge supervisat el propòsit de la qual és separar o discriminar de la millor manera possible les mostres en el conjunt d'entrenament segons el seu valor de classe.¹ La idea fonamental darrere de LDA, proposada per Ronald Fisher, és trobar un nou espai de característiques per projectar les dades de manera que es maximitzi la separabilitat de les classes.¹ Això s'aconsegueix maximitzant la diferència entre les mitjanes de les classes (centroïdes de classe), normalitzada per una mesura de la variabilitat dins de la classe, i minimitzant la separació de les mostres dins de cada classe.¹ LDA assumeix una distribució Gaussiana de les dades.¹

Els passos per implementar LDA són els següents:

1. Calcular el vector de mitjanes d -dimensional per a les diferents classes del conjunt de dades.¹
2. Calcular la matriu de dispersió (*Scatter matrix*), tant entre classes com dins de la classe.¹
3. Ordenar els autovectors per autovalors decreixents i triar k autovectors amb els majors autovalors per formar una matriu W de dimensió dxk .¹
4. Usar aquesta matriu d'autovectors per transformar les mostres al nou subespai mitjançant la multiplicació de matrius: $Y=XW$.¹

En comparar PCA i LDA, s'observa que ambdós són tècniques de transformació lineal comunament utilitzades per a la reducció de dimensionalitat.¹ No obstant això, PCA és un algorisme no supervisat que busca els eixos de components ortogonals de màxima variància en un conjunt de dades, mentre que LDA és un algorisme supervisat l'objectiu del qual és trobar el subespai de característiques que optimitza la

separabilitat de classes.¹

Malgrat els seus avantatges, LDA presenta diverses debilitats: no funciona bé si el disseny no està balancejat (és a dir, si el nombre d'objectes en les diferents classes és molt diferent).¹ També pot tenir un rendiment deficient si la distribució de les dades és significativament no Gaussiana.¹ A més, és sensible al sobreajust (

overfit) i no és aplicable (o és inferior) per a problemes no lineals.¹

T-Distributed Stochastic Neighbor Embedding (t-SNE)

T-Distributed Stochastic Neighbor Embedding (t-SNE) és una tècnica no lineal de reducció de dimensionalitat, utilitzada principalment per visualitzar conjunts de dades d'alta dimensió.¹ El seu mecanisme clau consisteix a visualitzar dades d'alta dimensió assignant a cada punt de dades una ubicació en un mapa 2D o 3D, amb el benefici fonamental d'intentar preservar l'estructura local de les dades.¹ Això significa que les distàncies i agrupacions dels punts observades en dimensions baixes (com un gràfic de dispersió 2D) són el més properes possible a les distàncies en l'espai d'alta dimensió, ja que t-SNE optimitza les distàncies tenint en compte l'estructura local.¹

La principal diferència entre t-SNE i PCA rau en el fet que t-SNE intenta desconvolucionar les relacions entre veïns en dades d'alta dimensió, sent particularment efectiu per a conjunts de dades altament no lineals, on PCA podria distorsionar les distàncies (exemple del "swiss roll").¹ Mentre que PCA és una tècnica matemàtica, t-SNE és probabilística.¹ Els algoritmes lineals de reducció de dimensionalitat com PCA es concentren a col·locar punts de dades dissimilars lluny els uns dels altres en una representació de menor dimensió. En contrast, per representar dades d'alta dimensió en un manifold no lineal de baixa dimensió, és essencial que els punts de dades similars es representin a prop, la qual cosa t-SNE aconsegueix i PCA no.¹

No obstant això, t-SNE presenta diverses debilitats: és computacionalment costós i pot trigar diverses hores en conjunts de dades de milions de mostres, a diferència de PCA que es completa en segons o minuts.¹ A més, a causa de la seva naturalesa probabilística, diferents execucions amb els mateixos hiperparàmetres poden produir resultats diferents, la qual cosa requereix observar múltiples gràfics abans de realitzar qualsevol avaluació.¹ Una altra limitació és que t-SNE no aconsegueix preservar

l'estructura global entre clústers; les distàncies d'incrustació entre clústers manquen de significat i la distribució global dels clústers és aleatòria.¹ Malgrat això, t-SNE està dissenyat per capturar relacions polinòmiques complexes entre característiques, cosa que PCA no pot fer a causa de la seva linealitat.¹

Visualització Tomogràfica (Isocontorns, Superfícies, Renderitzat de Volum)

La **visualització tomogràfica** es refereix a les tècniques utilitzades per mesurar i segmentar estructures internes d'objectes 3D, com es veu en l'anàlisi d'estructures craniofacials en embrions de ratolí.¹ Dins d'aquest camp, els

isocontorns o contorns de mapes tomogràfics són línies que connecten punts d'igual valor. Quan les línies de contorn estan molt separades, indiquen un pendent més suau, mentre que si estan molt juntes, mostren un pendent més pronunciat.¹ Les elevacions de punts, per la seva part, representen les altures entre aquestes línies de contorn.¹

Existeixen mètodes que redueixen la dimensionalitat de les dades de 3D a 2D per a la seva visualització. Un **gràfic de superfície** (*surface plot*) mostra el límit del conjunt de dades, però no revela la seva part interior.¹ Un

pla de tall (*slice plane*) exhibeix una secció 2D en una coordenada específica, com una mateixa coordenada Y.¹ D'altra banda, una

isosuperfície (*isosurface*) representa un valor escalar específic, ignorant tots els punts de volum que no compleixen amb aquest valor.¹

El **renderitzat de volum** (*volume rendering*) és una tècnica l'objectiu principal de la qual és crear una imatge 2D que reflecteixi, en cada píxel, les dades escalars dins d'un conjunt de dades 3D donat.¹ El desafiament principal en aquest procés rau en l'elecció de la funció de mapeig que transforma un conjunt complet de valors escalars (per als vòxels al llarg del raig de visió) a un sol píxel en la imatge 2D resultant.¹

Finalment, la **Tomografia Computada (CT)** utilitza raigs X i un model matemàtic, la Transformada de Radon, per reconstruir una imatge 3D d'un objecte a partir de les seves projeccions.¹ L'objectiu de la CT és recuperar una imatge detallada del cos (valors de densitat) a partir de les dades de raigs X sobre un nombre finit de línies.¹

La representació d'informació 3D en visualitzacions 2D presenta un desafiament inherent. Atès que moltes d'aquestes tècniques impliquen una reducció de la dimensionalitat de 3D a 2D, es produeix una pèrdua o transformació d'informació. Aquesta limitació subratlla la dificultat de transmetre eficaçment estructures 3D complexes en un medi 2D, com una pantalla o paper. L'elecció de la tècnica de visualització en aquest context es converteix en una decisió sobre quin aspecte de les dades 3D és més crític comunicar en 2D, establint les bases per a una discussió més profunda sobre els avantatges i desavantatges de les visualitzacions 2D enfront de les 3D.

Taula 1: Comparativa de Tècniques de Reducció de Dimensionalitat

Característica	PCA (Anàlisi de Components Principals)	LDA (Anàlisi Discriminant Lineal)	t-SNE (t-Distributed Stochastic Neighbor Embedding)
Propòsit	Reduir variables, trobar eixos de màxima variància.	Separar (discriminar) classes, maximitzar separabilitat entre grups.	Visualitzar conjunts de dades d'alta dimensió, preservar estructura local.
Linealitat	Lineal	Lineal	No Lineal
Supervisió	No Supervisat	Supervisat (requereix etiquetes de classe)	No Supervisat
Objectiu Principal	Maximitzar la variància global de les dades.	Optimitzar la separabilitat de classes.	Preservar les relacions de veïnatge (estructura local).
Fortaleses	Molt interpretable, PCs ortogonals, variància explicada.	Maximitza la diferència entre mitjanes de classes, minimitza variabilitat intra-classe.	Efectiu per identificar clústers locals, captura relacions complexes.
Debilitats	Sensible a <i>outliers</i> , funciona millor amb dades contínues.	Sensible a dades no balancejades, distribucions no Gaussianes, <i>overfit</i> , no apte per a problemes no lineals.	Computacionalment costós, resultats variables entre execucions, no preserva estructura global entre clústers.
Cost	Baix	Moderat	Alt

Computacional			
Naturalesa	Matemàtica	Matemàtica	Probabilística
Consistència	Consistent (mateixos resultats amb mateixes dades/paràmetres)	Consistent (mateixos resultats amb mateixes dades/paràmetres)	Pot variar entre execucions (requereix múltiples observacions)

II. Visualització Multivariable i Avançada

Els conjunts de dades grans presenten un desafiament fonamental en la visualització: contenen molta més informació de la que pot mostrar-se eficaçment en un sol gràfic.¹ Per abordar aquesta complexitat, es recorre a solucions com l'agrupació de variables en un únic panell o, per a conjunts de dades més complexos, la creació de figures amb múltiples panells, on cadascun mostra un subconjunt específic de les dades.¹ Aquesta necessitat de manejar la "maledicció de la dimensionalitat" s'estén més enllà del processament de dades i afecta directament el disseny de la visualització. Fins i tot després de la reducció de dimensionalitat, el gran nombre de variables restants o la complexitat de les seves relacions segueix sent un desafiament significatiu per a la visualització directa, imposant una càrrega cognitiva considerable a l'observador. Això impulsa el desenvolupament de tècniques de visualització multivariable especialitzades que abstracten o organitzen les dades de manera més digerible, fins i tot si es perd algun detall. La visualització, en aquest sentit, es converteix en un acte de curiosa selecció i presentació, no només de descàrrega de dades brutes.

Visualització de Distribucions Múltiples

En escenaris on es desitja visualitzar múltiples distribucions simultàniament, és útil conceptualitzar les dades en termes d'una **variable de resposta** (la variable les distribucions de la qual es volen mostrar, com lifeExp o el value de mètriques d'Iris) i **variables d'agrupació** (que defineixen subconjunts de dades amb distribucions distintes de la variable de resposta, com continent/year o species/petal/sepal).¹

Per a aquest propòsit, s'empren diverses tècniques de visualització ¹:

- **Boxplots:** Mostren la distribució de la variable de resposta per a cada grup.¹
- **Violin Plots:** Similars als boxplots, però exhibeixen la densitat de la distribució.
- **Strip Charts:** Mostren cada punt de dades individualment.
- **Stacked Histograms:** Histogrames apilats per comparar distribucions.
- **Overlapping Densities:** Presenten densitats superposades per comparar distribucions.¹
- **Ridgeline Plot:** Mostren múltiples distribucions de densitat en un format apilat.
- **Sina Plots:** Una variació dels strip charts que ajusta els punts per visualitzar la densitat.
- Les distribucions també poden organitzar-se al llarg dels eixos vertical o horitzontal per facilitar la comparació.¹

Visualització de Proporcions Múltiples

El desafiament de visualitzar proporcions s'accentua quan aquestes s'especifiquen segons múltiples variables d'agrupació.¹ Per a això, existeixen diverses tècniques de visualització ¹:

- **Multiple Pie Charts:** Tendeixen a ser ineficients en espai i sovint enfosqueixen les relacions entre les proporcions.¹
- **Grouped Bars:** Funcionen bé sempre que el nombre de condicions comparades sigui moderat.¹
- **Stacked Bars:** Poden ser eficaços per a un gran nombre de condicions.¹ Permeten ajustos de posició com dodge (costat a costat), fill (apilar i normalitzar altura), jitter (afegir soroll per evitar superposició), nudge (moure etiquetes) i stack (apilar).¹
- **Stacked Densities:** Són apropiades quan les proporcions canvien al llarg d'una variable contínua.¹ L'apilament desglossa les dades per més d'una variable categòrica que componen el tot ¹, com s'il·lustra amb el conjunt de dades diamonds usant `geom_density(position="fill")`.¹ El paràmetre `adjust` en `geom_density` permet ajustar l'ample de banda.¹
- **Mosaic Plots:** Assumeixen que cada nivell d'una variable d'agrupació pot combinar-se amb cada nivell d'una altra.¹ Són similars als gràfics de barres apilades, però tant les altures com els amplex de les àrees ombrejades varien, sent proporcionals al nombre de casos de cada combinació possible de variables categòriques.¹ Es construeixen subdividint successivament els eixos X i Y.¹
- **Treemaps:** Funcionen bé fins i tot si les subdivisions d'un grup són completament

distintes de les subdivisions d'un altre, a diferència dels mosaic plots.¹ Aniden recursivament rectangles, on l'àrea de cadascun és proporcional a la quantitat de casos.¹ Requereixen una variable numèrica contínua per a l'àrea (value), una per al color de farciment (parent), una per a l'etiqueta (id) i el grup pare (parent).¹ La seva implementació en R es realitza amb el paquet treemapify, utilitzant funcions com geom_treemap(), geom_treemap_subgroup_border(), geom_treemap_subgroup_text() i geom_treemap_text().¹

- **Parallel Sets (PSP):** Són més efectius que els mosaic plots o treemaps quan hi ha més de dues variables d'agrupació.¹ Mostren com el conjunt de dades total es desglossa per cada variable categòrica individual, utilitzant bandes ombrejades per il·lustrar les relacions entre els subgrups.¹ No són adequats per a dades numèriques (per als quals s'utilitzen els PCP) i les barres horitzontals en els PSP mostren la freqüència absoluta de cada categoria.¹

Visualització de Relacions i Correlacions

Per entendre com es relacionen dues o més variables quantitatives¹, s'empren tècniques com la

Scatterplot Matrix (SPLOM) i el **Bubble Plot**. La SPLOM utilitza múltiples diagrames de dispersió organitzats en una matriu per determinar i visualitzar la correlació entre una sèrie de variables.¹ Permet identificar relacions per parells, la seva naturalesa, la presència de valors atípics i l'agrupament per grups.¹ No obstant això, es torna difícil de manejar amb més de 3 o 4 variables quantitatives.¹ El Bubble Plot, per la seva part, utilitza cercles o altres formes per representar dades, on la mida i el color del punt poden codificar variables addicionals simultàniament.¹

Figures Multi-Panell (Small Multiples, Figures Compostes)

Les figures multi-panell es classifiquen en dues categories principals¹:

- **Small Multiples** (també coneguts com "trellis plot" o "faceting"): Són gràfics que consisteixen en múltiples panells disposats en una quadrícula regular. Cada panell mostra un subconjunt diferent de les dades, però tots els panells utilitzen el

mateix tipus de visualització.¹ La idea clau és segmentar les dades segons una o més dimensions, visualitzar cada segment per separat i organitzar les visualitzacions en una quadricula, etiquetant columnes, files o panells individuals amb els valors de les dimensions de les dades.¹ És fonamental evitar l'ús de diferents escales en els eixos en panells separats, o almenys cridar l'atenció del lector sobre això si és necessari, i organitzar els panells en un ordre lògic i significatiu.¹

- **Compound Figures:** Combinen diversos panells independents ensamblats en una disposició arbitrària. Aquests panells poden mostrar visualitzacions completament diferents o fins i tot conjunts de dades distints, amb l'objectiu de transmetre un punt general unificador.¹ Per assegurar la claredat, és crucial mantenir la consistència entre colors, símbols i fonts, i els eixos han d'estar alineats per evitar males interpretacions.¹

Gràfics de Coordenades Paral·leles (PCP)

Els **Gràfics de Coordenades Paral·leles (PCP)** són una tècnica de visualització utilitzada per analitzar dades numèriques multivariades, permetent traçar elements de dades individuals a través de moltes dimensions.¹ A diferència del pla cartesià clàssic, cada variable numèrica té el seu propi eix, disposat en paral·lel, verticalment i equidistant.¹ Cada element de dades es representa mitjançant segments de línia connectats, derivats d'un conjunt connectat de punts, un en cada eix.¹

La fortalesa dels PCP resideix en la seva capacitat per comparar variables que poden ser completament diferents, amb rangs i fins i tot unitats distintes.¹ En general, moltes línies paral·leles indiquen una relació positiva, mentre que moltes línies que es creuen (formes de X) suggereixen una associació negativa.¹ No obstant això, els PCP solen ser difícils d'entendre per a audiències no tècniques a causa del típic "embolic de línies".¹ No es recomana mostrar massa variables numèriques alhora, i els PCP no són adequats per a dades categòriques; en el seu lloc, s'han d'usar els Parallel Sets Plots (PSPs), on les barres horitzontals mostren la freqüència absoluta de cada categoria.¹

Per millorar la llegibilitat dels PCP, s'apliquen transformacions típiques¹:

- **Escalat (Scaling):** Transforma les dades brutes a una nova escala comuna, crucial per comparar variables amb diferents unitats.¹
- **Ordre dels Eixos (Axis Order):** Optimitzar l'ordre dels eixos verticals pot disminuir

el desordre (*clutter*) en minimitzar el nombre de creuaments entre sèries.¹

- **Ressaltat** (*Highlighting*): Ressaltar una mostra o grup d'interès específic pot evitar la superposició excessiva de línies.¹

La importància de seleccionar la tècnica de visualització adequada segons el tipus de variable i la pregunta analítica és un principi rector. L'efectivitat d'una visualització depèn en gran mesura de la seva alineació amb el tipus de dades (numèriques, categòriques, proporcionals) i la tasca analítica específica (comparar distribucions, mostrar parts d'un tot, identificar relacions). Utilitzar un tipus de gràfic incorrecte, com un PCP per a dades categòriques, pot portar a una interpretació errònia o a una manca d'eficiència. Això reforça el concepte de "precisió visual", suggerint que la visualització de dades és una ciència de decisions deliberades per comunicar el missatge correcte.

Taula 2: Tipus de Paletes de Color i les seves Aplicacions

Tipus de Paleta de Color	Tipus de Dades Ideal	Propòsit/Aplicació	Característiques Clau
Categòrica/Qualitativa	Dades nominals (categories sense ordre)	Diferenciar categories discretes.	Grans diferències en to (hue), saturació i/o lluminositat per facilitar la distinció. Evitar gradients continus que suggereixin ordre. Màxim 12 colors.
Quantitativa Seqüencial	Dades ordinals o quantitatives que progressen de baix a alt.	Mostrar una progressió ordenada de valors.	Variació consistent en lluminositat i/o saturació d'un mateix to o tons adjacents. Pot ser contínua o discretitzada.
Quantitativa Divergent	Dades quantitatives amb un punt central significatiu (ex. zero, mitjana).	Ressaltar desviacions positives i negatives des d'un valor central.	El color central (normalment blanc o neutre) ha de representar el valor central. Els colors divergeixen cap a tons distints en cada extrem.

Agrupada	Múltiples categories principals amb subcategories.	Organitzar visualment dades jeràrquiques o agrupades.	Utilitza diferents tons per a les categories principals i variacions en saturació i lluminositat dins de cada to per a les subcategories.
Mapes Coroplètics Bivariats	Dues variables quantitatives superposades espacialment.	Mostrar la interacció o relació entre dues distribucions espacials.	Esquema de color amb dues escales superposades en dues direccions, sovint formant una matriu (ex. 3x3) de combinacions de color.

III. Principis de Color i Precisió Visual

L'elecció del color en la visualització de dades va més enllà de l'estètica; és un component crític per a la integritat i accessibilitat de la informació. La forma en què els humans percebem el color difereix fonamentalment de com les computadores el processen.¹ Els humans tenim tres tipus de cons que reaccionen a diferents espectres de llum, i la nostra percepció es basa en la lluminositat, el to i la saturació.¹ En contrast, les computadores processen bandes de freqüència molt estretes de vermell, verd i blau, i calculen la llum de manera lineal.¹ La Teoria del Procés Oponent d'Ewald Hering explica com el cervell combina els senyals per subtracció per crear canals de diferència de color (vermell-verd, groc-blau) i un canal de luminància (blanc i negre).¹ Aquesta teoria també explica per què certs colors, com el groc, poden percebre's amb major intensitat a causa de la sobreexcitació dels receptors.¹ A més, la nostra percepció és inherentment relativa; els judicis de color no són absoluts i el context que envolta els elements visuals pot modular la discriminabilitat i la precisió.¹

Per garantir la precisió i la coherència, es recomana utilitzar **models de color perceptivament correctes** com HSV (Hue, Saturation, Value), HCL (Hue, Chroma, Luminance) i CIELab (L^* , a^* , b^*).¹ Aquests models són preferibles perquè s'acosten més a la forma en què l'ull humà percep el color, i s'han d'usar escales de color amb interpolació no lineal per evitar distorsions.¹

Canals de Color i la seva Efectivitat

El color és un dels canals visuals més potents per codificar dades.¹ Es classifica en:

- **Canals de Magnitud** (per a atributs ordenats o dades quantitatives): Inclouen la posició en una escala comuna (la més efectiva), la posició en una escala no alineada, la longitud, la inclinació/angle, l'àrea, la profunditat, la luminància de color, la saturació de color, la curvatura i el volum (la menys efectiva).¹
- **Canals d'Identitat** (per a atributs categòrics): Inclouen la regió espacial, el to de color (hue), el moviment i la forma.¹

La luminància i la saturació són canals de magnitud, mentre que el to és un canal d'identitat.¹ Aquests tenen poca o cap interferència amb els canals de posició.¹ La jerarquia dels canals visuals implica que el color ha de usar-se judiciosament, complementant en lloc de reemplaçar codificacions més efectives. Encara que el color és un canal poderós, no és el més efectiu per a comparacions quantitatives precises. Una excessiva dependència del color per a judicis quantitius exactes, quan la posició o la longitud serien més clares, constitueix una deficiència en el disseny. Per tant, el color ha d'emprar-se per donar suport a la codificació principal, per exemple, per ressaltar categories o afegir una capa addicional d'informació, en lloc de ser l'únic mitjà per transmetre diferències quantitatives precises, especialment en dades ordenades.

Problemes d'Escales Arc de Sant Martí

L'ús d'una paleta de color incorrecta, com les escales arc de sant martí, pot mostrar efectes en la visualització que no estan presents en les dades, la qual cosa representa un risc significatiu de desinformació.¹ Les escales arc de sant martí no tenen en compte com percebem el color i manquen d'un ordre inherent, la qual cosa les fa inadequades per representar dades ordenades.¹ La seva variació no lineal trenca la linealitat de l'escala a causa de la sobreexcitació de receptors en certs punts, la qual cosa pot generar l'aparició de límits o "bandes" perceptuals que no existeixen en la realitat.¹ Això dificulta la distinció de regions i porta a una pèrdua de detall.¹ Per totes aquestes raons, es recomana encareidament evitar l'ús de l'escala arc de sant martí.¹

Propietats del Canal de Color (Contrast, Lluminositat, Saturació, To)

Les propietats del canal de color són crucials per a una visualització efectiva. El **to** (*hue*), la **saturació** i la **lluminància/claredat/brillo** són components fonamentals.¹ El

contrast facilita la detecció: un major contrast fa que els elements siguin més fàcils de percebre, mentre que un menor contrast els dificulta.¹ Un contrast homogeni en el to, però amb variació en el to i la lluminositat, facilita la identificació.¹ És important tenir en compte el

contrast simultani, on el context que envolta els elements modula la percepció del color.¹

Paletes i Esquemes de Color (Categòrics, Seqüencials, Divergents, Agrupats)

Una directriu general per a la selecció de colors és evitar intentar seleccionar colors individualment; en el seu lloc, es recomana utilitzar paquets de R com grDevices, colorRamps, RcolorBrewer i colorspace.¹ El paquet

colorspace, basat en els models HCL i HSV, és particularment recomanat pel seu atractiu estètic i les seves opcions de personalització.¹

Els tipus de paletes i esquemes de color es trien en funció del tipus de dades ¹:

- **Categòrica/Qualitativa:** S'utilitza per diferenciar categories discretes.¹ És important evitar colors que formin un gradient continu, ja que podrien confondre's amb valors ordinals.¹ Es prefereixen grans diferències en to, saturació i/o lluminositat, i es recomana no usar més de 12 colors.¹
- **Quantitatiu Seqüencial:** Ideal per a dades que progressen de baix a alt, podent ser continus o discretitzats (ordinals).¹
- **Quantitatiu Divergent:** Dissenyat per a dades amb un punt central significatiu (per exemple, el zero).¹ El color central (normalment blanc) ha de representar aquest valor central, i els colors divergeixen cap a tons distints en cada extrem. Pot ser continu o discretitzat.¹
- **Agrupada:** Un esquema de color per a múltiples categories principals (diferents

tons) amb passos de saturació i lluminositat dins de cada to.¹

- **Mapes Coroplètics Bivariats:** Utilitzen un esquema de color amb dues escales superposades en dues direccions, sovint resultant en una matriu de 3x3.¹

Les escales perceptivament lineals són aquelles que varien de forma consistent a través del rang de valors.¹ Es recomana usar escales perceptivament correctes, com les de la família Viridis (Viridis, Magma, Plasma, Inferno, Cividis, Mako, Rocket, Turbo).¹

L'elecció del color no és merament estètica; és un component crític per a la integritat i l'accessibilitat de les dades. Si el color es tria incorrectament, pot distorsionar activament les dades, creant patrons falsos o ocultant els reals, i a més pot excloure segments de l'audiència. Això eleva la selecció del color d'una decisió de disseny a una qüestió d'integritat de les dades, la qual cosa implica una responsabilitat ètica per garantir l'accessibilitat.

Disseny per a Deficiències de Color

És fonamental dissenyar visualitzacions que siguin accessibles per a persones amb daltonisme.¹ Per a això, es poden utilitzar eines i simuladors com

Colorblindness per a R, Coblis (www.color-blindness.com) i Color Brewer 2.0 (colorbrewer2.org).¹

Semàntica del Color

L'assignació de color ha de basar-se en el seu significat inherent per evitar confusions i millorar la comprensió. Per exemple, es poden usar paletes càlides per a altes temperatures, una escala blau-vermella per a temperatures negatives a positives, o el verd per a dades de vegetació.¹ És crucial evitar usos enganyosos, com emprar una escala contínua com si fos categòrica, usar més de 12 categories en una mateixa paleta, o presentar rànquings sense un ordre clar.¹ Una bona pràctica és seguir el principi de "No data, no color", és a dir, no assignar color a regions que manquen de dades.¹

IV. Honestitat i Precisió en la Visualització

La precisió visual és un pilar fonamental en la comunicació de dades. Les representacions visuals, fins i tot si es basen en dades precises, poden inherentment distorsionar la veritat si s'ignoren els principis de disseny. Les decisions de disseny en si mateixes comporten missatges implícits i poden alterar activament la realitat. Fins i tot amb dades correctes, una visualització mal dissenyada pot "mentir". Això estableix un vincle causal directe entre el disseny i la interpretació, la qual cosa subratlla la dimensió ètica de la visualització de dades.

Manipulació d'Eixos (Escala, Línia Base Zero, Etiquetes, Eixos Duals)

La manipulació dels eixos és una font comuna de deshonestitat visual. L'elecció de l'escala pot emfatitzar la variació o la tendència de manera enganyosa.¹ Un

eix Y truncat, que no comença en zero o per sota del valor mínim significatiu, pot magnificar petits canvis, fent-los semblar més dramàtics del que realment són.¹ La directriu és mostrar l'eix complet per no magnificar descensos o ascensos.¹

Per a les **gràfiques de barres**, que codifiquen la **mida**, és **imprescindible** que tinguin una **línia base zero** per representar amb precisió les quantitats.¹ No fer-ho és incorrecte i pot generar una falsa impressió de les magnituds relatives.¹ En contrast, els gràfics de línies, que codifiquen la

posició dels punts, no necessàriament necessiten una línia base zero.¹ Les

etiquetes en els eixos són **imprescindibles** per a la claredat i comprensió de les dades.¹

Es recomana **evitar els eixos duals** ja que són problemàtics i poden portar fàcilment a conclusions equivocades.¹ Això es deu al fet que suggereixen una correlació implícita entre les línies que pot no existir, i només són acceptables en casos molt específics.¹

Agregacions Equivalents i Dades Faltants

Per assegurar la precisió, les **agregacions** han de ser **equivalents** en comparar dades, evitant missatges enganyosos.¹ Per exemple, en comparar inversions de la indústria abans i després de regulacions, els períodes de temps han de ser comparables.¹

Quant a les **dades faltants**, és crucial **marcar-les explícitament** en la visualització.¹ Això implica mantenir l'espai per a les marques absents o usar marques distintes per indicar la seva absència, en lloc de simplement ometre-les, la qual cosa podria generar una falsa sensació de continuïtat o completitud.¹

Precisió en Gràfiques d'Àrea i Barres Apilades

Les **barres apilades no alineades** són problemàtiques perquè dificulten la comparació de percentatges i la identificació de punts de referència com el 50%.¹ Es recomana alinear-les a una base comuna o usar altres representacions.

Les **gràfiques d'àrea** (*area charts*) també poden ser enganyoses.¹ És molt difícil identificar quina categoria és responsable d'un augment, i els pics en les capes superiors tendeixen a magnificar-se, distorsionant la percepció del canvi.¹ En molts casos, es recomana no usar gràfics d'àrea i, en el seu lloc, calcular índexs o utilitzar gràfics de línies separats per a cada categoria.¹

Un aspecte important a considerar és que les configuracions predeterminades del programari de visualització poden ser una font de deshonestitat. Molts programes poden establir eixos automàticament o usar gràfics d'àrea sense el context adequat, la qual cosa pot portar a representacions enganyoses. Confiar únicament en aquestes configuracions sense un pensament crític sobre el disseny pot resultar en visualitzacions que, sense intenció maliciosa, distorsionen la veritat. Per tant, és fonamental que els usuaris avaluin críticament la sortida visual de les seves eines i anul·lin les configuracions predeterminades quan sigui necessari per garantir la precisió i l'honestedat.

Selecció de Mètriques Rellevants (Taxes vs. Totals)

En visualitzar dades, és essencial triar **mètriques que siguin veritablement rellevants** per al missatge.¹ Un error comú és mostrar totals (per exemple, el nombre total d'assassinats en ciutats) en un mapa geogràfic, la qual cosa sovint es correlaciona amb la mida de la població i pot portar a la conclusió errònia que les ciutats més grans són intrínsecament més perilloses.¹ La directriu és usar

taxes (per exemple, la taxa d'assassinats per cada 100.000 persones) en lloc de totals per a comparacions significatives entre entitats de diferents mides.¹ Els mapes de perfil geogràfic que són essencialment només mapes de població es consideren una "molèstia" comuna (

pet peeve).¹

Representació d'Incertesa i Error (Alternatives a Boxplots)

La representació de la incertesa i l'error és crucial per a una visualització honesta. Els **Box and Whisker Plots (Boxplots)**, encara que útils per mostrar la distribució, tenen limitacions¹: la caixa es percep com una unitat, no mostren tota la distribució, i l'associació de mida a quantitat pot ser enganyosa (una caixa petita pot representar una quantitat gran).¹ A més, poden fer que grups distints semblin gairebé idèntics.¹

Com a alternatives als boxplots, se suggereixen¹:

- **Violin Plots:** Mostren la densitat de la distribució.
- **Gradient Plots.**
- **Bean Plots:** Similars als violin plots però mostren els punts individuals.
- **Beeswarm Plots:** Mostren cada punt de dades individualment, evitant la superposició.

La inclusió d'**estimacions d'incertesa o barres d'error** és una pràctica fonamental que millora la presa de decisions i proporciona un millor reflex de la comprensió científica.¹ Per als gràfics de línies, es recomanen les

bandes d'error (*Error Bands*) i els **fancharts** com a alternatives per representar la incertesa.¹

Taula 3: Errors Comuns en Precisió Visual i Com Evitar-los

Error Comú / Pràctica Enganyosa	Per què és Enganyós	Pràctica Correcta / Alternativa
Eix Y Truncat (no comença en zero o valor mínim)	Magnifica petits canvis, distorsiona la magnitud de les variacions.	Mostrar l'eix complet , incloent el zero o el rang complet de dades.
Línia Base No Zero per a Barres	Les barres codifiquen mida; sense línia base zero, la magnitud relativa és incorrecta.	Les barres han de tenir una línia base zero . (Els gràfics de línies, que codifiquen posició, no la requereixen necessàriament).
Eixos Duals	Suggereixen una correlació implícita que pot no existir; són difícils d'interpretar.	Evitar eixos duals en la majoria dels casos. Si és indispensable, usar amb extrema precaució i explicacions clares.
Barres Apilades No Alineades	Dificulta la comparació de percentatges i la identificació de proporcions (ex. 50%).	Alinear les barres apilades a una base comuna o usar alternatives que facilitin la comparació de proporcions.
Magnificació en Gràfics d'Àrea	Els pics en capes superiors es magnifiquen, fent difícil identificar la contribució real de cada categoria.	No usar gràfics d'àrea en molts casos. Considerar gràfics de línies separats o índexs per a cada categoria.
Totals en Mapes Geogràfics (en lloc de taxes)	Els mapes de totals sovint reflecteixen la densitat de població, no la taxa real del fenomen, portant a conclusions errònies.	Usar taxes (ex. per 100.000 habitants) per a comparacions significatives entre entitats de diferent mida.
Boxplots sense context complet	No mostren tota la distribució, poden fer que grups semblin idèntics o que la mida de la caixa s'associï erròniament a la quantitat.	Usar Violin Plots, Bean Plots, Beeswarm Plots o incloure la distribució completa per a una representació més precisa.

Absència d'Etiquetes en Eixos	Impedeix la comprensió clara de les dades i les seves unitats.	Les etiquetes en els eixos són imprescindibles .
Dades Faltants no indicades	Crea una falsa sensació de continuïtat o completitud en les dades.	Marcar explícitament on falten dades, mantenint l'espai o usant marques distintes.

V. Visualització amb Mapes

Els mapes són eines poderoses per a la visualització de dades geoespacials.¹ Permeten que audiències no tècniques compreguin i analitzin dades de manera més senzilla.¹ No obstant això, crear un bon mapa no és una tasca senzilla, ja que les persones sovint recorden l'aparença del mapa més que les dades que representa.¹ Abans de dissenyar un mapa, és crucial respondre a dues preguntes fonamentals: Quin és el propòsit del mapa? i Quin és el cost/impacte esperat del resultat?.¹

La saliência visual de l'àrea geogràfica en un mapa pot induir a error en la interpretació si no es normalitza mitjançant mètriques rellevants. Per exemple, els mapes coroplètics, encara que útils per mostrar patrons, poden ser enganyosos si la mida d'una regió no es correlaciona amb les dades que se li atribueixen. Això es deu a una tendència cognitiva comuna: les àrees més grans atrauen naturalment més atenció i es perceben implícitament com a "més" del que s'està mapejant, fins i tot si les dades subjacents (com la densitat o la taxa) són menors. Això estableix un vincle causal directe entre la codificació visual (mida de l'àrea) i la possible mala interpretació. Per evitar això, és crucial distingir quan un mapa coroplètic és apropiat (per exemple, per a patrons de votació política on l'àrea és rellevant) enfront de quan es necessita un mapa de símbols proporcionals (per a recomptes bruts) o un mapa basat en taxes (per a comparacions normalitzades).

Fonaments i Tipus de Mapes

Existeixen diversos tipus de mapes per a diferents propòsits ¹:

- **Mapa de Punts (Point Map):** És la representació més simple, col·locant un punt

en cada ubicació corresponent a la variable.¹ Són útils per mostrar patrons de distribució i densitat, però requereixen dades d'ubicació precises (geocodificació) i els punts poden superposar-se segons el nivell de zoom.¹

- **Mapes de Símbols Proporcional** (*Proportional Symbol Maps / Bubble Maps*): Utilitzen cercles o altres formes la mida i el color dels quals representen variables de dades addicionals.¹ També poden patir de superposició i requereixen geocodificació precisa.¹
- **Mapes de Clúster** (*Cluster Maps*): Similars als mapes de símbols proporcional, però en fer zoom, poden revelar punts més petits representats pels punts més grans.¹
- **Diagrames de Voronoi** (*Voronoi Diagrams*): S'empren per trobar la distància mínima necessària per assolir un punt o fita, ajudant a comprendre la proximitat i les característiques de distància.¹ Són útils, per exemple, per triar una nova ubicació per a una escola lluny de les existents.¹
- **Mapes Coroplètics** (*Choropleth Maps*): Divideixen una àrea en seccions (per exemple, límits polítics o geogràfics) i omplen cada secció amb un to o ombra diferent per representar una variable o un rang de valors.¹ Són excel·lents per a patrons i tendències generals, però s'ha de tenir precaució en usar-los amb àrees que difereixen molt en mida, ja que la mida d'una regió pot no tenir relació amb les dades.¹
- **Mapa de Punts** (*Dot Map*): Cada punt representa una instància o quantitat específica.¹
- **Mapa d'Isopletes** (*Isopleth Map*): Mostra dades contínues utilitzant línies que connecten punts d'igual valor.¹
- **Mapes Esquemàtics** (*Schematic Maps*): S'utilitzen per simplificar navegacions i sistemes complexos, com metros o xarxes elèctriques.¹
- Existeixen molts altres tipus de mapes, com els mapes de fons amb R i ggplot2, mapes de bombolles interactius, cartogrames (on la geometria de les regions es distorsiona per representar una variable), mapes hexbin (regions representades com hexàgons o densitat 2D) i mapes de connexió (que mostren rutes entre ubicacions).¹

Consideracions per a Mapes Interactius

Les APIs de Google Maps (Google Earth, Google Maps Images, Google Places) ofereixen eines per construir programes de mapeig visual interactius.¹ Els mapes són

eines poderoses per al

storytelling, com s'observa en exemples com "The desirability of Boston streets" o "Pinellas County is the worst place in Florida to be black and go to public school".¹ No obstant això, el seu impacte pot veure's esbiaixat pel fet que l'usuari sovint s'enfoca en l'aparença del mapa en si, més que en les dades que representa.¹ Aquesta dualitat suggereix que, si bé els mapes són atractius i poden transmetre narratives potents, el seu fort atractiu visual pot, en ocasions, eclipsar la informació quantitativa que es pretén comunicar. El medi pot convertir-se en el missatge, desviant potencialment l'atenció dels coneixements numèrics subjacents. Per això, les visualitzacions basades en mapes requereixen un disseny acurat per garantir que les dades segueixin sent el més important. Els dissenyadors han d'equilibrar l'atractiu estètic amb la claredat de les dades, possiblement utilitzant elements interactius o anotacions per guiar l'atenció de l'usuari cap a la informació quantitativa.

VI. Visualització 2D vs. 3D

L'elecció entre visualitzacions 2D i 3D és una decisió crítica que ha de basar-se en la tasca de l'espectador i les capacitats perceptives humanes, no en la mera disponibilitat d'eines 3D.

Desafiaments Perceptuals de 3D (Profunditat, Oclusió, Distorsió)

La intuïció comuna que experimentem el món en 3D és enganyosa; en realitat, els humans veiem en **2.05D**, i la inexactitud dels nostres judicis de profunditat no és sorprenent.¹ Només podem percebre la capa exterior d'un món 3D, veient un únic punt al llarg de cada raig.¹

L'**occlusió** és un desafiament inherent a les visualitzacions 3D, on parts de l'escena poden quedar ocultes per altres.¹ Encara que la navegació interactiva pot ajudar a resoldre l'estructura 3D de les parts ocloses, això requereix temps i imposa una càrrega cognitiva considerable a l'usuari.¹

La **distorsió de perspectiva** és un altre fenomen problemàtic: els objectes distants

semblen més petits i canvien la seva posició planar en el pla de la imatge.¹ Aquesta distorsió, juntament amb l'oclusió, fa que els gràfics de barres 3D siguin intrínsecament més difícils d'interpretar que els seus equivalents 2D.¹

El sistema perceptual humà està fonamentalment mal equipat per realitzar judicis quantitatius precisos en 3D sobre una pantalla 2D. La projecció de dades 3D en una pantalla 2D per a anàlisis quantitatives introdueix biaixos perceptuals i precisions inherents que són difícils de superar. El cervell lluita per comparar amb precisió longituds, àrees o posicions en un espai 3D distorsionat. Això constitueix un fort argument contra l'ús gratuït de 3D per a dades que són inherentment 2D o que podrien representar-se eficaçment en 2D. Subratlla que el 3D ha de reservar-se únicament quan l'estructura geomètrica és la principal informació que es necessita.

Criteris per Triar entre 2D i 3D

La decisió d'usar 3D ha de ser deliberada:

- **Quan usar 3D:** Només quan la tasca de l'espectador requereix fonamentalment comprendre l'**estructura geomètrica tridimensional** d'objectes o escenes.¹ Exemples inclouen animacions facials en temps real, on la combinació de dades 2D i 3D pot reduir el cost computacional.¹ En aquests casos, una vista 3D amb **controls de navegació interactius** (per ajustar el punt de vista) permet als usuaris construir un model mental útil de l'estructura del conjunt de dades més ràpidament.¹
- **Quan NO usar 3D** (per a dades 2D): Un gràfic 3D que mostra dades 2D no afegeix valor i, de fet, afegeix soroll (*clutter*), fent el gràfic més difícil de llegir.¹ Si no és necessari per a la comprensió de l'estructura 3D, no s'ha d'usar 3D.¹

La navegació interactiva és una solució necessària, però insuficient, per als problemes inherents de la visualització 3D. Encara que la interactivitat pot mitigar alguns dels desafiaments de la visualització 3D, com l'oclusió, no els elimina per complet. Al contrari, desplaça la càrrega de la interpretació a l'usuari, qui ha de participar activament per resoldre les ambigüitats visuals. Això suggereix que, si bé la interactivitat permet l'exploració, comporta un cost en termes de càrrega cognitiva per a l'usuari. Una visualització 3D veritablement efectiva minimitza la necessitat d'una interacció extensa per resoldre ambigüitats fonamentals, la qual cosa implica que la interactivitat en 3D és una eina per a l'exploració, no una solució per a les deficiències

perceptuals inherents.

S'han d'aplicar **mètriques de qualitat** generals, aplicables tant a 2D com a 3D ¹:

- Usar **tons de color suaus**, excepte on es vulgui cridar l'atenció, ja que els colors brillants poden ser esgotadors.
- Assegurar l'**accessibilitat al color** utilitzant eines com Coblis per verificar la compatibilitat amb el daltonisme.
- **Pivotar el gràfic** (ex., usar gràfics de barres horitzontals) quan les etiquetes de categoria són llargues per facilitar la lectura.
- Triar un **ample de barra apropiat**: barres més primes amb suficient espai en blanc perquè siguin lleugerament més gruixudes que l'espai entre elles, evitant que estiguin massa juntes (el cervell podria intentar avaluar àrea vs. longitud).
- **Eliminar l'encapçalament de l'eix X i afegir etiquetes de dades** si aquestes últimes ja proporcionen la informació necessària, fent l'encapçalament redundant.

VII. Visualització Científica i Mètriques de Qualitat

La visualització científica és una forma especialitzada de comunicació que prioritza la transferència d'informació sobre l'embelliment estètic. Les decisions de disseny s'inclinen fortament cap a la claredat, la precisió i l'eficiència. Qualsevol element de disseny que no contribueixi directament a la comprensió es considera "soroll" o "clutter". Això significa que el context de la visualització és primordial; una visualització per a una publicació científica tindrà prioritats de disseny diferents a les d'una revista d'estil de vida, la qual cosa subratlla la importància de l'audiència i el propòsit en les decisions de disseny.

Directrius Clau per a Publicacions Científiques

L'objectiu principal de qualsevol gràfic en el context de publicacions científiques i presentacions és **transmetre informació de manera efectiva**.¹ Per a això, es proposen deu directrius clau ¹:

1. **Simplicitat**: Crear el gràfic més simple que transmeti la informació desitjada.¹

2. **Codificació:** Considerar el tipus d'objecte i atribut de codificació utilitzats.¹
3. **Focus:** Centrar-se en visualitzar patrons o detalls segons el propòsit del gràfic.¹
4. **Rangs d'Eixos:** Seleccionar rangs d'eixos significatius.¹
5. **Taxes de Canvi:** Utilitzar transformacions de dades i relacions d'aspecte del gràfic per emfatitzar les taxes de canvi en sèries de temps.¹
6. **Punts Superposats:** Traçar punts superposats de manera que les diferències de densitat siguin aparents en gràfics de dispersió (aquesta directriu també aplica a mapes).¹
7. **Sèries de Temps:** Usar línies en connectar dades seqüencials en gràfics de sèries de temps.¹
8. **Agregació:** Agregar conjunts de dades grans de manera significativa.¹
9. **Comparació de Variables:** Mantenir els rangs dels eixos el més similars possible en comparar variables.¹
10. **Esquema de Color:** Seleccionar un esquema de color apropiat basat en el tipus de dades (en referència a la classe de color).¹

A més, la inclusió d'**estimacions d'incertesa o barres d'error** és una pràctica fonamental que millora la presa de decisions i proporciona un millor reflex de la comprensió científica.¹

Reducció de Soroll Visual (Clutter)

El *clutter* es defineix com una col·lecció desordenada d'entitats gràfiques que no són necessàries per comprendre la informació.¹ El seu impacte és significatiu, ja que el

clutter enfosqueix l'estructura de les dades i dificulta la identificació de patrons, relacions i estructures.¹

Per reduir el *clutter*, es poden aplicar diversos mètodes:

- **Ordre de Dimensions:** Variar l'ordre de les dimensions pot reduir el *clutter* sense disminuir el contingut d'informació.¹
- **Eliminació d'Ítems Comuns de Clutter:** Això inclou efectes 3D en dades 2D, línies de quadrícula fosques (s'han d'usar línies grises suaus o eliminar-les), l'ús excessiu de colors brillants o negres, i l'ús innecessari de text en majúscules.¹

Procés de Conceptualitzar, Crear, Criticar i Tallar (4 Cs)

El procés de les "4 C" (Conceptualitzar, Crear, Criticar i Tallar) representa un enfocament cíclic i iteratiu per al disseny de visualitzacions, emfatitzant el refinament i la centralitat en l'usuari.

- **Conceptualitzar:** Implica crear un esborrany (*rough plot*) de la infografia, definir un títol enganxós i descriptiu, establir un punt de partida i un punt final (conclusió o crida a l'acció), i definir la ruta visual que es desitja que el lector segueixi (hipòtesi).¹
- **Crear:** Consisteix a elaborar la història de les dades, incloent informació interessant i sorprenent, i personalitzant la presentació.¹
- **Criticar:** Se centra en la consistència de la història de les dades (temps verbal, terminologia, disseny) i en demanar a una altra persona que revisi la visualització abans de publicar-la.¹
- **Tallar (Cut):** Implica eliminar informació supèrflua o innecessària, mantenint només l'essencial i interessant. L'objectiu és no aclaparar el lector amb massa text o imatges, i considerar si la informació extra podria anar en una altra infografia o si una part es comunicaria millor com a imatge, reorganitzant per millorar el flux.¹ Es proporcionen exemples que comparen infografies de cervesa, mostrant-ne una amb molt *clutter* enfront d'una de clara i concisa.¹

Aquest procés és un flux de treball pràctic que ensenya que el disseny de visualitzacions efectives és un procés de millora contínua i refinament, on la retroalimentació i la simplificació són clau. Això també es connecta amb el tema de la usabilitat en emfatitzar les proves i la iteració.

VIII. Storytelling en Visualització de Dades

El *storytelling* en la visualització de dades és una elecció estratègica per superar la càrrega cognitiva i l'apatia, transformant el consum passiu de dades en un compromís actiu. La narració estructurada d'una sèrie de fets i accions, dutes a terme per un o més personatges, és una tècnica de comunicació que crea un vincle emocional amb l'audiència i ajuda a que el missatge destaquï.¹ Aplicat a la visualització de dades, el

storytelling maximitza l'impacte en la comunicació de dades.¹ El seu propòsit principal és atraure i mantenir l'atenció de l'audiència, transmetre un missatge de forma memorable a través de l'emoció, dirigir l'interès i connectar amb l'audiència, i comunicar una idea embolicada en emoció, com ho descriu Robert McKee.¹

La informació en brut, fins i tot si està ben visualitzada, pot ser aclaparadora o poc atractiva per a moltes audiències. El *storytelling* actua com un pont cognitiu, proporcionant estructura i ganxos emocionals per fer que la informació complexa sigui digerible i rellevant. Aborda la tendència humana inherent a buscar narratives i significat, la qual cosa significa que la capacitat d'elaborar una narrativa convincent al voltant de les dades és una habilitat d'alt nivell que millora l'impacte pràctic de les visualitzacions, especialment en contextos professionals on les parts interessades necessiten informació clara i processable.

Estructures Narratives (Viatge de l'Heroi, Tres Actes)

Les històries posseeixen característiques universals.¹ El

Viatge de l'Heroi, popularitzat per Joseph Campbell com el "monomite", descriu una estructura fonamental que comparteixen els mites: l'heroi inicia la seva aventura des del món ordinari cap a una regió de prodigis, s'enfronta a forces fabuloses, obté una victòria decisiva i torna amb dons per a la seva comunitat.¹ Aquesta estructura ha tingut una gran influència i ha donat lloc a variacions com el "Story Circle" de Dan Harmon.¹

Una altra estructura narrativa fonamental és l'**Estructura Clàssica de Tres Actes** d'Aristòtil ¹:

1. **Plantejament (Introducció):** Es presenta el personatge i el seu món, i un incident incitador condueix al conflicte.
2. **Nus (Conflicte):** Es desenvolupen els esforços per resoldre el conflicte, culminant en un clímax.
3. **Desenllaç (Resolució):** Es presenta el resultat del conflicte.

Les transicions entre aquests actes estan marcades per situacions de conflicte que generen tensió, un element clau per mantenir l'atenció.¹ Aquesta estructura serveix com un model general per a la comunicació.¹

Tipus de Narrativa (Anotació, Narració, Història)

No totes les visualitzacions requereixen ser una història completa; existeixen diverses estratègies per comunicar eficaçment amb dades.¹ Per guiar l'usuari, es poden emprar les següents estratègies:

1. **Anotació:** Consisteix en text i marques que ajuden a interpretar la visualització.¹
2. **Narració:** Implica anotacions connectades en una seqüència lògica.¹ Pot ser **tancada** (una estructura seqüencial amb un missatge definit) o **oberta** (una visualització exploratòria que serveix com a eina per buscar narratives).¹
3. **Història:** Una narració que incorpora drama o conflicte per provocar una resposta emocional.¹

La necessitat de contar una història s'accentua quan l'audiència no té un context previ sobre les dades o la visualització, quan es requereix guiar-la per evitar conclusions errònies, o quan no existeix un interès previ en el tema, el qual pot generar-se a través del *storytelling*.¹

L'element de "conflicte", tradicionalment central en el *storytelling*, es tradueix en la visualització de dades com la posada de relleu de discrepàncies o problemes, la qual cosa impulsa l'interès de l'usuari. En la narració de dades, el "conflicte" no és necessàriament un drama humà, sinó una discrepància, un problema o un patró inesperat dins de les dades. Ressaltar aquests "conflictes", com la inexactitud d'un model, un valor atípic o una tendència divergent, crea una tensió narrativa que impulsa l'audiència a buscar una resolució o comprensió. Això proporciona una aplicació concreta dels principis abstractes del *storytelling* a les dades. Els dissenyadors poden identificar "conflictes" en els seus conjunts de dades i utilitzar-los com a motors narratius, fent que les seves visualitzacions siguin més atractives i impactants en emmarcar-les com a problemes que han d'entendre's o resoldre's.

Planificació de la Història (Storyboard)

El **storyboard** és el primer pas crucial en la planificació d'una història de dades.¹ Permet organitzar les idees en seccions i subseccions, reduint cada part a la seva

mínima expressió (una acció, una frase).¹ Això facilita visualitzar el flux de la història i realitzar canvis. Pot ser multinivell, permetent desenvolupar l'esquema de la història en subseccions.¹ Les eines utilitzades són similars a les de planificació de qualsevol narrativa, com guions o taulers.¹ Un exemple pràctic d'aplicació es troba en la visualització de dades d'una granja eòlica, on s'il·lustra el procés des dels prototips inicials fins a la imatge final, destacant els desafiaments i les solucions per comunicar la precisió del model.¹

IX. Interactivitat i Animació

La interactivitat transforma l'observador en un usuari, atorgant-li un paper més prominent i proactiu en la percepció de la informació, amb menys dependència de la guia del visualitzador.¹ Aquesta capacitat expandeix els límits físics del que es pot mostrar en un espai determinat, abastant una major varietat d'anàlisis per satisfer diverses necessitats d'informació.¹ La interactivitat facilita la manipulació de dades per respondre a diferents preguntes, amplifica el control i la personalització de l'observador, i amplia l'interès a diversos grups d'usuaris, especialment a les noves generacions, acostumades a un consum d'informació més actiu.¹

Abans d'incorporar tècniques interactives, és fonamental estudiar les necessitats, els mitjans disponibles i les opcions utilitzables.¹ Els conceptes clau en interactivitat inclouen l'

Esdeveniment (una acció de l'usuari, com un clic o una pulsació de tecla), el **Control** (la reacció a un esdeveniment, com mostrar un pop-up o ressaltar una opció) i l'**Acció** (l'operació a realitzar en l'aplicació, com filtrar o ressaltar informació).¹ Els tipus de controls abasten des de contenidors (Frames, Panels) fins a components específics com botons, caselles de verificació, etiquetes, camps de text, lliscadors, menús desplegable, barres de cerca i selectors de data.¹

Factors que Influeixen en la Incorporació d'Interactivitat

Diversos factors determinen la incorporació de tècniques interactives ¹:

- **Restriccions:** La tecnologia disponible i les competències de l'observador o l'audiència.¹
- **Objectius:** L'experiència d'usuari desitjada i les opcions interactives disponibles.¹
- **Representació de Dades:** Quins gràfics podrien beneficiar-se de la interactivitat per facilitar la interpretació.¹
- **Disseny Fiable:** Assegurar que el sistema funcioni com s'espera i que la integritat de les dades es mantingui.¹
- **Disseny Accessible:** Minimitzar les pulsacions de tecles o clics del ratolí per a una interactivitat discreta i útil.¹
- **Disseny Elegant:** Garantir una interacció fluida i millorar la jugabilitat.¹

Categories d'Interacció en Visualització de Dades

Les interaccions es classifiquen en diverses categories ¹:

- **Filtratge:** Permet als usuaris especificar quines dades desitgen veure o ocultar.¹ Exemples inclouen la selecció d'esports en resultats olímpics, diferències de gènere en doctorats, espècies d'arbres a Nova York i anàlisi de mercat amb lliscadors.¹
- **Selecció:** Ressalta valors d'interès sense eliminar-los, sovint modificant atributs de color o reordenant dades.¹ Exemples són el ressaltat de premis Nobel per país, noms de bebès populars al llarg dels anys, reformes de compensació laboral per estat, histogrames del cens per nació (brushing), morts per armes als EUA i resultats bancaris (ordenació per columna).¹
- **Participació/Col·laboració:** Permet als usuaris contribuir amb les seves pròpies dades per personalitzar la representació visual.¹ Exemples inclouen "Quants anys tens?" (comparant l'edat de l'usuari amb la de celebritats), "Com de bé coneixes la teva àrea?" (un qüestionari sobre el veïnat), "Quanta sucre en el menjar?" (un qüestionari sobre el contingut de sucre) i "Recordes on estava dividida Alemanya?" (traçant la frontera).¹
- **Resum:** Proporciona diferents nivells d'anotacions de dades sota demanda de l'usuari, permetent a l'usuari controlar la percepció de les dades.¹ Exemples inclouen un mapa de calor de creuaments de personatges dels Venjadors, la retòrica dels presidents dels EUA i visualitzacions de Bloomberg amb pop-ups de "com llegir aquest gràfic".¹
- **Observació i Exploració:** Mostra múltiples vistes o diferents nivells de detall de forma dinàmica, especialment quan no tota la informació pot mostrar-se

simultàniament.¹ Exemples inclouen "Earth" (patrons de vent, clima, oceà amb zoom/pan), "Racial Dot Map" (densitat de població amb zoom geomètric), "2020 Census Demographic data Map Viewer" (dades demogràfiques amb zoom/pan) i "OECD Better Life Index" (drill-down en països).¹ Aquesta categoria també inclou l'exploració lineal a través de seqüències narratives, com "Killing the Colorado" (un viatge pas a pas de la degradació del riu) i "What's Really Warming the World?" (una seqüència de causes del canvi climàtic).¹

Animació

L'**animació** mostra dades amb una dimensió temporal de forma dinàmica, permetent apreciar patrons que no es poden veure en imatges estàtiques.¹ És una seqüència d'imatges amb canvis entre imatges consecutives al llarg de l'espai i el temps, creant la il·lusió de moviment en el cervell.¹ Encara que genera un fort impacte visual en la visualització de dades, no tots els usuaris la aprecien.¹

El propòsit de l'animació en la visualització de dades és múltiple ¹:

- **Incrementar la Dimensionalitat:** Afegint el temps com una dimensió a les dades.
- **Proporcionar Informació Diferent:** Oferint coneixements no disponibles en gràfics estàtics.
- **Qualitat Estètica:** Millorant l'atractiu visual.
- **Coneixement:** Proporcionant informació contextual fora de l'àrea d'atenció específica.
- **Transició:** Guiant l'usuari entre diferents gràfics.
- **Descripció Funcional:** Relacionant el comportament que representa un objecte animat (per exemple, el desplaçament).
- **Èmfasi:** Cridant l'atenció sobre un element visible o un procés específic.
- **Expressió:** Enriquant l'experiència de l'usuari amb l'aplicació o tasca.
- **Canvi:** Mostrant com un objecte o procés evoluciona amb el temps.
- **Visualització Directa:** Establir correspondències entre atributs de moviment (fase, freqüència) i variables del conjunt de dades.
- **Associació:** Utilitzant seqüències de moviment per establir relacions entre grups d'informació.

En considerar l'animació, és important tenir en compte la durada entre els fotogrames: una animació massa lenta pot resultar avorrida, mentre que una massa

ràpida pot provocar la pèrdua de detalls.¹ L'animació també pot distreure del missatge.¹ És crucial que l'usuari tingui control per iniciar i aturar l'animació, a través d'esdeveniments i controls com l'inici automàtic en carregar la pàgina, botons de reproducció/pausa/aturada/repetició i lliscadors per controlar els fotogrames.¹ Exemples d'animació inclouen "Breathing Earth" (simulant zones de vegetació global), jugadors de la NFL (altura i pes al llarg del temps) i la taquilla global per gènere (carrera de gràfics de barres animats).¹

Consells Generals per a Interactivitat i Animació

La gestió del temps és crucial en implementar solucions interactives.¹ No s'han d'invertir temps i recursos en tècniques interactives o innovadores si no aporten valor a la comprensió.¹ És fonamental centrar-se en el que és important i rellevant, qüestionant sempre si la interactivitat és necessària i aporta valor al projecte.¹

X. Usabilitat i Experiència d'Usuari (UX)

L'avaluació dels sistemes de visualització de dades és fonamental, especialment quan són interactius.¹ Els termes

Usabilitat i Experiència d'Usuari (UX) s'empren per avaluar aquests sistemes, i encara que sovint es confonen, UX abasta la usabilitat com un dels seus components.¹

Usabilitat

La **Usabilitat** es defineix com un atribut de qualitat que avalua la facilitat d'ús d'aplicacions, llocs web, interfícies i sistemes per part de l'usuari.¹ Inclou cinc components clau de qualitat ¹:

- **Aprenentatge** (*Learnability*): Facilitat amb la qual els usuaris aprenen les tasques.
- **Eficiència**: Rapidesa amb la qual els usuaris realitzen les tasques.

- **Memorabilitat:** Facilitat per recordar i reproduir tasques després d'un temps sense usar el sistema.
- **Errors:** Nombre i gravetat dels errors, i capacitat de recuperació del sistema.
- **Satisfacció:** Grau de satisfacció general amb l'ús del sistema.

És important diferenciar entre **Utilitat**, **Usabilitat** i **Útil** ¹:

- **Utilitat:** Es refereix a la funcionalitat, és a dir, si el sistema fa el que els usuaris esperen.
- **Usabilitat:** Se centra en com de fàcil i satisfactori és l'ús del sistema.
- **Útil:** És una combinació d'Utilitat i Usabilitat.

La usabilitat és crucial perquè és una condició necessària perquè els usuaris adoptin i continuïn utilitzant un sistema.¹ Els usuaris tendeixen a abandonar sistemes complicats, que fallen amb freqüència, no permeten realitzar accions desitjades, causen desorientació o tenen informació d'ajuda poc clara.¹ La usabilitat està estretament lligada a la productivitat.¹

En el context de la visualització de dades, la "**Usabilitat de Dades**" es refereix a la qualitat de les dades i la informació, associada a tres principis ¹:

- **Fiabilitat de Dades:** Qualitat del processament de dades, assegurant confiança en les dades amb els seus intervals d'error.
- **Estabilitat de Dades:** Minimitzar l'impacte negatiu de transformacions o processaments, com la reducció de dimensionalitat.
- **Suport a la Presa de Decisions:** La representació ha de ser prou clara perquè els usuaris prenguin decisions informades.

Per avaluar la usabilitat, l'enfocament principal és l'usuari, mitjançant observació i qüestionaris.¹ Dos qüestionaris estàndard són:

- **SUS (System Usability Scale):** Una eina ràpida per mesurar la usabilitat de programari, llocs web i aplicacions. Consta de 10 preguntes amb una escala Likert de 5 punts. És simple, fàcil d'entendre i proporciona resultats fiables, distingint entre sistemes útils i no útils. Avalua l'efectivitat (si l'usuari pot realitzar accions) i l'eficiència (l'esforç de l'usuari).¹ La puntuació es calcula com $SUS = (X + Y) * 2.5$, on X és la suma de les puntuacions de preguntes imparells - 5, i Y és 25 - la suma de les puntuacions de preguntes parells. Una puntuació inferior a 51 indica possibles problemes.¹
- **PSSUQ (Post-Study System Usability Questionnaire):** Àmpliament utilitzat per mesurar la satisfacció de l'usuari. La versió 3.0 té 16 preguntes amb una escala Likert de 7 punts, proporcionant indicadors de satisfacció general, utilitat del

sistema (SYSUSE), qualitat de la informació (INFOQUAL) i qualitat de la interfície (INTERQUAL).¹ L'elecció entre SUS i PSSUQ depèn dels aspectes a avaluar i la fatiga de l'avaluador.¹

Experiència d'Usuari (UX)

L'**Experiència d'Usuari (UX)** estudia l'experiència global d'un usuari en interactuar amb un sistema.¹ Abasta indicadors objectius, estètics i emocionals, incloent usabilitat, interfície d'usuari, experiència d'interacció, disseny interactiu i experiència del client.¹ Les definicions d'UX emfatitzen que és una conseqüència de l'estat intern de l'usuari (predisposicions, expectatives, necessitats, motivació, estat d'ànim), les característiques del sistema (complexitat, propòsit, usabilitat, funcionalitat) i el context d'interacció (entorn organitzacional/social, significat de l'activitat, voluntarietat d'ús).¹

Els mètodes d'avaluació d'UX inclouen ¹:

- **Avaluacions d'Experts:** Realitzades per especialistes.
- **Estudis de Laboratori:** Similars a les proves d'usabilitat, útils en la fase de prototipat.
- **Estudis de Camp:** Els participants utilitzen el sistema en el seu entorn natural.
- **Qüestionaris:** Els usuaris responen preguntes per formar una opinió.
- **Mètodes Mixtos:** Combinació de diferents mètodes per obtenir indicadors més complets.
- **Mesures Psicofisiològiques:** Mesures objectives com la freqüència cardíaca, la transpiració i els moviments dels músculs facials per avaluar emocions positives o negatives, idealment amb dispositius no invasius.¹

Models d'UX

Els models d'avaluació d'UX es basen en dos aspectes principals: la usabilitat i l'experiència/emotivitat.¹ Elements externs són crucials per entendre l'UX com una experiència subjectiva i dinàmica ¹:

- **Usuari:** Les seves competències (novell, experimentat, expert) i característiques

emocionals (estatus social, objectius personals, necessitats afectives) poden influir en l'UX.¹

- **Context:** L'entorn físic (il·luminació, connexió a internet) i el context sociocultural (valors, actituds, experiència prèvia) tenen un paper.¹
- **Aspectes Temporals:** La qualitat pragmàtica (usabilitat) tendeix a augmentar amb el temps, mentre que la qualitat hedònica (emotivitat) pot disminuir a mesura que la novetat desapareix.¹ L'avaluació de l'UX es realitza típicament quan un sistema està acabat, amb menys estudis a llarg termini.¹

Els models d'UX més utilitzats inclouen ¹:

- **Hassenzahl UX:** Basat en tres factors (Usuari, Sistema, Context), distingeix entre atributs pragmàtics (utilitat i funcionalitat) i hedònics (estètica, comportament, funcionalitat intuïtiva, estimulació, identificació, evocació).¹ El qüestionari Attrakdiff, amb 28 criteris, s'utilitza per a aquest model.¹
- **User Experience Questionnaire (UEQ):** Avalua productes interactius amb 26 indicadors bipolars en 7 escales, agrupats en 6 categories: atractiu, perspiciuïtat, eficiència, fiabilitat, estimulació i novetat.¹
- **meCUE:** Una avaluació modular dels components clau de l'UX, que busca crear una mesura estàndard. Es basa en el model CUE, que defineix característiques instrumentals i no instrumentals, similars als atributs pragmàtics i hedònics, i que influeixen en les respostes emocionals de l'usuari.¹ El qüestionari meCUE té tres mòduls validats (percepcions del producte, emocions i conseqüències de l'UX) i és flexible, permetent seleccionar mòduls segons el que es necessiti mesurar.¹

Models d'UX per a Visualització de Dades

Per als sistemes de visualització de dades, és crucial distingir entre característiques instrumentals/pragmàtiques i no instrumentals/hedòniques ¹:

- **Instrumentals:** Inclouen utilitat (satisfacer necessitats, dades útils), completitud (informació correcta i suficient), percepció (informació comprensible amb mínim esforç), confiança (informació vàlida i precisa) i intuïció (informació intuïtiva i fàcil d'entendre).¹
- **No Instrumentals:** Abasten l'estètica (visualment agradable) i l'atractiu (que atregui l'usuari).¹

El pes de l'avaluació entre característiques instrumentals i no instrumentals depèn de

l'ús del sistema.¹ En sistemes basats en accions, com els de visualització de dades, les qualitats instrumentals són més importants, ja que l'objectiu és optimitzar el rendiment de la tasca.¹

El model i qüestionari **meCUE** es considera el més adequat per avaluar sistemes de visualització de dades ¹ perquè:

- Atorga un pes suficient a les característiques no instrumentals, alhora que emfatitza la usabilitat.
- El seu qüestionari és flexible.
- Valora els aspectes visuals i estètics.
- El model i el qüestionari funcionen per a diferents tipus d'aplicacions.

Conclusions

La visualització de dades és un camp multidisciplinari que exigeix una comprensió profunda tant dels principis estadístics i computacionals com de la percepció humana i el disseny. La reducció de dimensionalitat és un pas fonamental per manejar la complexitat de les dades, però la seva elecció ha de ser estratègica, considerant si es prioritza l'estructura global o local de les dades. La precisió visual és un pilar inquebrantable, ja que fins i tot les visualitzacions basades en dades correctes poden distorsionar la veritat si els principis de disseny, com l'ús adequat dels eixos, les línies base i les paletes de color, no es respecten. La manipulació subtil dels elements visuals pot portar a interpretacions errònies, la qual cosa subratlla la responsabilitat ètica del dissenyador.

El color, encara que un canal visual potent, ha d'usar-se amb judici, complementant codificacions més efectives i considerant sempre l'accessibilitat per a persones amb deficiències de color. Els mapes, eines poderoses per al *storytelling* geoespacial, requereixen una acurada normalització de les dades per evitar que la saliència visual de l'àrea geogràfica indueixi a error. La visualització 3D, malgrat el seu atractiu, presenta desafiaments perceptuals inherents que limiten la seva precisió per a judicis quantitatius en pantalles 2D, per la qual cosa el seu ús ha de reservar-se per a quan la comprensió de l'estructura geomètrica és essencial.

El *storytelling* emergeix com una estratègia vital per transformar el consum passiu de dades en un compromís actiu, superant la càrrega cognitiva i l'apatia de l'audiència.

En emmarcar les dades en narratives estructurades que incloguin elements de "conflicte" (discrepàncies o problemes en les dades), es genera tensió i s'impulsa l'interès de l'usuari. La interactivitat i l'animació, si bé amplien les possibilitats d'exploració i mostren el dinamisme temporal, han d'incorporar-se amb un propòsit clar, evitant la distracció i prioritzant la comprensió de l'usuari. Finalment, la usabilitat i l'experiència d'usuari (UX) són mètriques crítiques per avaluar l'eficàcia d'una visualització. Models com meCUE permeten una avaluació holística, equilibrant la funcionalitat instrumental amb l'estètica i l'emotivitat, assegurant que la visualització no només sigui precisa, sinó també atractiva i fàcil d'usar.

Per a la preparació de l'examen, es recomana un enfocament esquemàtic i basat en paraules clau. El domini de les definicions, fórmules i comparatives entre tècniques (ex. PCA vs. LDA vs. t-SNE) és crucial. La capacitat d'identificar i corregir errors comuns de precisió visual (eixos truncats, línies base, ús de totals vs. taxes) serà fonamental. Comprendre el propòsit i les limitacions de cada tipus de visualització (inclosos els mapes i les representacions 2D/3D) i l'aplicació adequada dels principis del color garantirà una resposta precisa. Finalment, la integració del *storytelling* i la interactivitat com a eines per millorar la comunicació de dades, en lloc de mers adorns, completarà la comprensió experta de la matèria.

Obras citadas

1. fecha de acceso: enero 1, 1970,