

1 Informationsvisualisierung

1.1 Einführung und Definitionen

Die Informationsvisualisierung befasst sich mit der Darstellung abstrakter Daten, um dem Menschen das Verständnis und die Analyse dieser Daten zu erleichtern.

Informationsvisualisierung

Informationsvisualisierung ist die Nutzung computergestützter, interaktiver, visueller Repräsentationen von **abstrakten Daten**, um die **Kognition zu verstärken** (amplify cognition).

- **Abstrakte Daten:** Daten ohne inhärente räumliche Struktur (z. B. Finanzdaten, Textsammlungen, Softwarestrukturen). Es gibt keine direkte Abbildung auf eine Geometrie, weshalb künstliche visuelle Strukturen geschaffen werden müssen.
- **Ziel:** Daten „enthüllen“ (To reveal data). Visualisierung hilft dabei, Muster, Trends und Ausreißer zu erkennen, die in rein numerischen Tabellen verborgen bleiben.

1.2 Das Referenzmodell (Visualization Pipeline)

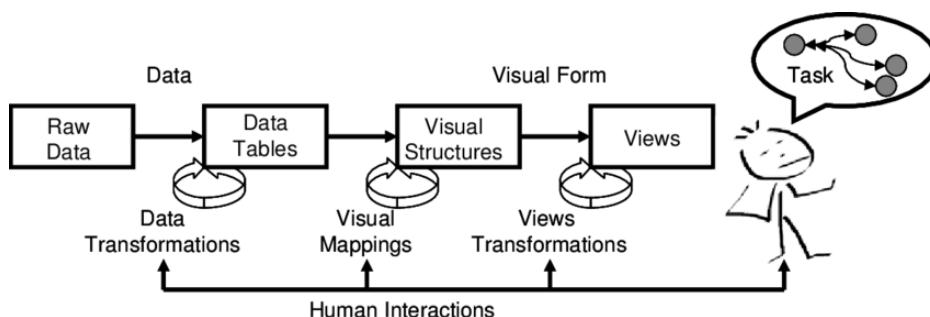
Das Referenzmodell nach Card et al. (1999) beschreibt den Prozess der Umwandlung von Rohdaten in eine visuelle Form, die vom Benutzer interpretiert werden kann.

Pipeline-Schritte

1. **Raw Data (Rohdaten):** Daten in ihrem ursprünglichen, oft proprietären Format.
2. **Data Tables (Datentabellen):** Strukturierte Relationen (Fälle × Variablen) inklusive Metadaten.
3. **Visual Structures (Visuelle Strukturen):** Räumliche Substrate, grafische Symbole (Marks) und deren Eigenschaften (Farbe, Größe).
4. **Views (Ansichten):** Spezifische grafische Darstellungen mit Parametern wie Zoom, Ausschnitt (Clipping) oder Skalierung.

1.2.1 Transformationen im Referenzmodell

- **Data Transformations:** Aufbereitung der Rohdaten (Filtern, Normalisieren, Aggregieren).
- **Visual Mappings:** Abbildung der strukturierten Daten auf visuelle Kanäle (z. B. ein Wert wird auf die x-Position abgebildet).
- **View Transformations:** Interaktive Modifikation der Ansicht (z. B. Rotation im 3D-Raum oder Zoom).



1.3 Daten- und Visualisierungstypen

Unterschiedliche Datentypen erfordern spezialisierte Visualisierungstechniken:

- **Zeitliche Daten:** Darstellung von Trends über die Zeit (z. B. Linien- oder Flächendiagramme).
- **Geographische Daten:** Daten mit Ortsbezug (z. B. Choroplethenkarten für Inzidenzwerte).
- **Multivariate Daten:** Daten mit vielen Attributen pro Datensatz (z. B. Scatterplot-Matrizen oder Parallele Koordinaten).
- **Hierarchien und Netzwerke:** Darstellung von Beziehungen (z. B. Baumdiagramme, Node-Link-Diagramme).

1.4 Interaktionstechniken

Interaktion ist essenziell, um große Datenmengen explorativ zu untersuchen. Sie greift in verschiedene Schritte der Visualisierungspipeline ein:

Pipelineschritt	Interaktionstechnik
View	Zoom, Pan (Verschieben), Highlight
Visuelle Abbildung	Farbschemaänderung, Datenauswahl für Abbildung
Datentransformation	Änderung der Normalisierung, Filtern, Editieren

1.5 Gutes Informationsdesign und Fehlerquellen

Effektive Visualisierungen vermeiden Verzerrungen und Fehlinterpretationen. Häufige Probleme in der Praxis sind:

- **Skalierungsfehler:** Unregelmäßige Zeitabstände auf Achsen suggerieren falsche Trends (z. B. bei der Darstellung von Schuldenwachstum).
- **Perspektivische Verzerrung:** 3D-Effekte erschweren den Vergleich von Größen (z. B. 3D-Tortendiagramme oder perspektivische Balken).
- **Visuelle Überladung (Clutter):** Zu viele Kanten in Netzwerkdiagrammen führen zu unleserlichen „Hairballs“.
- **Größenverhältnisse:** Flächen müssen proportional zu den Datenwerten wachsen ($A \propto r^2$). Wenn nur der Radius proportional zum Wert skaliert wird, wirkt die Fläche quadratisch vergrößert.

Anscombe's Quartet

Ein klassisches Beispiel dafür, warum Visualisierung wichtig ist: Vier Datensätze mit identischen statistischen Kennwerten (Mittelwert, Varianz, Korrelation), die aber völlig unterschiedliche Verteilungen aufweisen, wenn sie grafisch dargestellt werden.

1.6 Wichtige Formeln und Konzepte

In der Prüfungsvorbereitung ist insbesondere die korrekte Abbildung von Datenwerten auf visuelle Kanäle relevant:

- **Lie Factor (Lügenfaktor):** Das Verhältnis der im Bild dargestellten Änderung zur tatsächlichen Änderung in den Daten.

$$\text{Lie Factor} = \frac{\text{Größe des Effekts in der Grafik}}{\text{Größe des Effekts in den Daten}}$$

- Ein idealer Wert ist 1. Werte > 1.05 oder < 0.95 deuten auf eine erhebliche Verzerrung hin.