

Data visualisatie

- Introductie data visualisatie
 - Uitleg
 - Opdracht
- Data verwerving en voorbereiding
 - Uitleg
 - Opdracht

Introductie data visualisatie

In dit hoofdstuk ontdek je waarom datavisualisatie onmisbaar is om complexe data begrijpelijk te maken. Je leert welke doelen visualisaties dienen—verkennen, verklaren of overtuigen—en welke grafische ontwerpprincipes zorgen dat jouw visualisaties helder en visueel aantrekkelijk zijn.

Uitleg

Wat is data visualisatie?

Data visualisatie is het omzetten van bijvoorbeeld getallen en tekstgegevens naar beeldvormen zoals grafieken, diagrammen of interactieve dashboards of een grafische poster. Dit maakt trends, patronen en afwijkingen direct inzichtelijk voor een breed publiek.

- **Verkennen (Exploration):** Ontdek verborgen patronen en relaties in ongestructureerde data. Bijvoorbeeld een scatterplot waarin outliers zichtbaar worden.
- **Verklaren (Explanation):** Leg een dataset uit aan je publiek, met bijvoorbeeld een staafdiagram dat groei per jaar toont.
- **Overtuigen (Persuasion):** Gebruik data om een standpunt te onderbouwen, bijvoorbeeld een infographic die aantoont dat energiebesparing loont.

Doel en doelgroep

Elke visualisatie begint met een vraag: wat wil je laten zien en aan wie? Bedenk:

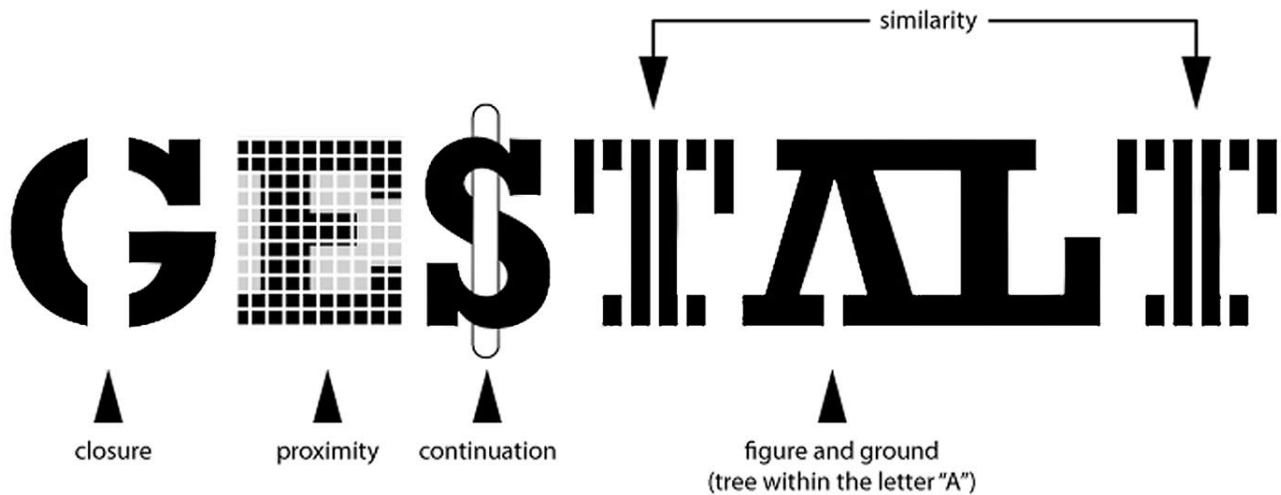
- **Doelstelling:** Wil je informeren, overtuigen of inzicht bieden?
- **Doelgroep:** Zijn het vakgenoten, managers of het algemene publiek? Pas hier je taal en complexiteit op aan.

Voorbeeld: Een grafiek voor management toont alleen hoogover groeicijfers, terwijl een wetenschappelijke publicatie diepgaand statistische details en betrouwbaarheidsintervallen bevat.

Basisprincipes van grafisch ontwerp

Goede visualisaties volgen ontwerpprincipes om informatie efficiënt over te brengen:

- **Gestalt-wetten:** Zorg dat gerelateerde elementen visueel gegroepeerd zijn.
Bijvoorbeeld: balken met dezelfde kleur in een staafdiagram vormen één groep.
- **Hiërarchie:** Gebruik grootte en contrast om belangrijke informatie te benadrukken.
Belangrijke datapunten kunnen groter worden weergegeven.
- **Kleurgebruik:** Kies een beperkt kleurenpalet en let op toegankelijkheid (kleurcontrast).
Gebruik kleuren consequent: één kleur per categorie.
- **Typografie:** Heldere lettertypes en voldoende lettergrootte. Gebruik vet of cursief spaarzaam om nadruk te leggen.
- **Witruimte:** Houd ruimte vrij tussen grafiekelementen om rommel te voorkomen en de leesbaarheid te vergroten.



Veelvoorkomende valkuilen

- **3D-effecten:** Laten data vervormen door verkeerde perspectieven.
- **Te veel kleuren:** Leiden af en maken interpretatie lastiger.
- **Onjuiste assen:** Niet vanaf nul starten kan trends vervormen.
- **Geen labeling of legende:** Gebruikers moeten zelf raden wat de lijnen of kleuren betekenen.

Praktijkvoorbeelden

- **Slechte visualisatie:** Een taartdiagram met meer dan zes segmenten. Het herkennen van kleine verschillen in hoeken en kleuren is bijna onmogelijk.
- **Goede visualisatie:** Een gestapelde staafdiagram met duidelijke kleuren en legende. Elk segment is verklaard, en de totale hoogte toont het totaal per categorie.

Opdracht

Schrijf een analyse (minimaal 600 woorden) van drie visualisaties uit verschillende contexten (bijv. nieuws, marketing, wetenschap). Voor elke visualisatie behandel je:

1. **Context:** Bron en publiek.
2. **Doelstelling:** Verkennen, verklaren of overtuigen.
3. **Ontwerpprincipes:** Beschrijf ten minste drie toegepaste principes.
4. **Feedback:** Geef twee verbeterpunten per visualisatie, inclusief een korte schets of omschrijving van je alternatieve ontwerp.

Lever je verslag in als PDF met afbeeldingen van de visualisaties en eigen annotaties.

Data verwerving en voorbereiding

In dit hoofdstuk leer je waar je data kunt vinden, hoe je deze ophaalt en voorbereidt voor visualisatie. Je maakt kennis met verschillende datatypen, bronnen en technieken voor het opschonen van data.

Uitleg

Datatypen en -structuren

- **Kwantitatieve data:** Numerieke waarden (bijv. aantallen, meetwaarden). Te visualiseren met histogrammen, boxplots.
- **Kwalitatieve data:** Categorieën of tekst (bijv. landen, productnamen). Te visualiseren met staafdiagrammen, woordwolken.
- **Gestructureerde data:** Tabellen in CSV, databases. Helpt bij kolom- en rijstructuur.
- **Ongestructureerde data:** Tekstbestanden, logs, afbeeldingen. Vereist extractie of transformatie.

Data bronnen

- **Open data portals:** CBS, Eurostat, World Bank.
- **API's:** RESTful endpoints, JSON- of XML-response.
- **Web scraping:** Automatisch gegevens ophalen van webpagina's met tools als BeautifulSoup.
- **Eigen data:** Sensoren, enquêtes of experimenten.

Data importeren

- In Python: gebruik `pandas.read_csv()`, `pandas.read_json()`, `sqlalchemy` voor databases.
- In Excel: data import wizard, Power Query voor complexere bewerkingen.

Data cleaning technieken

- **Ontbrekende waarden:** Identificeren (`isnull()`), verwijderen (`dropna()`) of imputeren (gemiddelde, mediaan).
- **Outliers:** Detectie via statistische methoden (IQR, z-score), visuele inspectie (boxplot).
- **Duplicaten:** Identificatie (`duplicated()`), verwijderen (`drop_duplicates()`).
- **Normalisatie/standaardisatie:** Schalen van kolommen (min-max, z-score) voor vergelijkbare eenheden.

Data profiling

Maak een eerste verkennende analyse:

- **Statistische samenvatting:** `describe()` toont mean, median, min-max.
- **Distributies bekijken:** Histogrammen, density plots.
- **Relaties onderzoeken:** Scatterplot matrix, correlatiematrix.

Opdracht

1. **Dataset kiezen:** Selecteer een dataset van een open data portal of via een API.
2. **Importeren:** Laad de dataset in Python (pandas) of Excel.
3. **Data cleaning:** Voer de volgende stappen uit en lever op:
 - Identificeer en behandel ontbrekende waarden.
 - Detecteer en behandel outliers.
 - Verwijder duplicaten.
 - Normaliseer één of twee kolommen.
4. **Data profiling:** Maak een overzicht met:
 - Statistische samenvatting (`describe()`).
 - Twee histogrammen voor numerieke kolommen.
 - Een correlatiematrix als heatmap.
5. **Documentatie:** Schrijf kort (300–500 woorden) over je reinigingsproces en uitdagingen.

Bundel alles in een verslag met de belangrijkste grafieken en een toelichting bij elke stap.