

What: Data Abstraction

Matakuliah Visualisasi Data (Data visualization)

Ahmad Luky Ramdani Ira Safitri Dimas Dwi Randa

Program Studi Data Sains
Fakultas Sains
Institut Teknologi Sumatera

Bandung, February 9, 2026

Gambaran Umum Abstraksi Data

Abstraksi Data

Tujuan utama abstraksi data adalah menerjemahkan data dari **bahasa domain** (spesifik bidang) menjadi **bahasa visualisasi** (abstrak).

Komponen Utama dalam The Big Picture:

- **What Data?:** Mengidentifikasi tipe dataset (Table, Network, Field, atau Geometry).
- **Data Types:** Membedakan elemen dasar seperti *Items*, *Attributes*, *Links*, *Positions*, dan *Grids*.
- **Attribute Types:** Menentukan apakah data bersifat *Categorical* atau *Ordered* (Ordinal/Quantitative).
- **Data Semantics:** Memahami arti dari data dalam konteks dunia nyata untuk menentukan metode pemrosesan yang tepat.

Mengapa Perlu Abstraksi Data?

- **Bahasa Universal:** Abstraksi memungkinkan kita menggunakan kosakata yang sama untuk data yang berasal dari domain berbeda (misal: "nasabah" dalam bank dan "pasien" dalam RS keduanya diabstraksikan sebagai *Items*)[cite: 5].
- **Tujuan:** Menentukan tipe data apa yang tersedia agar kita bisa memilih teknik visualisasi yang paling efektif[cite: 5].
- **Pertanyaan Kunci:** "Data apa saja yang sedang dilihat pengguna?" (*What?*) [cite: 5].

Mengapa Semantik dan Tipe Data Penting?

Why Data Semantics and Types Matter?

Memahami **Semantik** (Makna) dan **Tipe** (Struktur) adalah langkah awal untuk menghindari kesalahan interpretasi visual.

Semantik (Makna):

- Memberikan konteks pada angka mentah.
- Menentukan peran item (siapa subjeknya, apa atributnya).
- Contoh: Angka 10 bisa berarti \$10 (Uang) atau Kelas 10 (Tingkat).

Tipe (Struktur Matematika):

- **Categorical:** Hanya label (Apple, Orange).
- **Ordered:** Memiliki urutan.
 - *Ordinal:* Urutan logis (S, M, L).
 - *Quantitative:* Angka bisa dihitung (10kg, 20kg).

Data Mentah vs. Semantik (Makna)

"Raw data has no meaning without semantics."

- **Data Mentah:** 14, 2.6, 30, 30, 15, 100001.
- **Pertanyaan Abstraksi:**
 - Apakah ini *Geometry* (titik 3D)?
 - Apakah ini *Network* (node, links, dan weights)?
- Pentingnya Konteks untuk Menentukan Tipe Atribut
 - **Interpretasi 1 (Spasial):** Jika angka-angka tersebut adalah "dua titik di ruang 3D", maka kita mengabstraksikannya sebagai Geometry atau Position.
 - **Interpretasi 2 (Network):** Jika ada angka yang berarti "links" dan "weight", maka kita mengabstraksikannya sebagai Network/Trees dengan atribut pada links.
 - **Interpretasi 3 (Kategorikal):** Kata "Basil" bisa berarti nama tanaman (kategori makanan), nama lokasi (kategori spasial), atau nama tikus lab (kategori subjek).

Data Mentah vs. Semantik (Makna)

- **Pentingnya Metadata:** Kita membutuhkan informasi tambahan untuk menentukan apakah sebuah atribut bersifat **Categorical**, **Ordinal**, atau **Quantitative**.

Kesimpulan untuk Desainer

Visualisasi data bukan sekadar memetakan angka ke gambar, tetapi memetakan **makna (semantics)** ke dalam **abstraksi data** yang tepat.

Data Mentah vs. Semantik (Makna)

- Tujuan utama dari Abstraksi Data dalam desain visualisasi adalah menghilangkan ambiguitas
- Sebelum membuat grafik, desainer harus menerjemahkan data mentah tersebut menjadi:
 - Items: Apa entitas utamanya? (Misal: Tikus lab bernama Basil).
 - Attributes: Apa yang diukur? (Misal: 7 kali percobaan, S sebagai arah selatan).
- Tanpa mengetahui apakah "S" berarti "South" (Arah/Siklik) atau "Small" (Ordinal), desainer tidak bisa memilih saluran visual (visual channels) yang tepat.

Matching Principles: Ekspresifitas & Efektivitas

- **Prinsip Ekspresifitas:** Jangan gunakan tanda visual yang menyiratkan hubungan yang tidak ada pada data.
 - *Kesalahan umum:* Menggunakan garis penghubung pada data kategori (menyiratkan tren berkelanjutan).
- **Prinsip Efektivitas:** Gunakan saluran visual yang paling akurat untuk tipe data yang paling penting.
 - Data kuantitatif paling efektif menggunakan **Posisi** pada sumbu yang sama.
 - Data kategori paling efektif menggunakan **Warna (Hue)** atau **Bentuk**.

"The visual system has its own rules; the data must be mapped to respect those rules."

Dataset Types and Attributes

Tabel ini merangkum apa saja yang menjadi elemen pembentuk dalam setiap tipe dataset utama:

| Dataset Type | Data Types | Availability |
|---------------------|----------------------------------|---------------------|
| Tables | Items, Attributes | Umum digunakan |
| Networks | Items (Nodes), Links, Attributes | Relasional / Pohon |
| Fields | Grids, Attributes | Spasial / Kontinu |
| Geometry | Items, Positions | Bentuk Fisik 2D/3D |

Data Types

Munzner membagi Data Type menjadi empat tipe utama:

- **Items:** Entitas individu yang sedang diamati (Baris dalam tabel, Node dalam grafik).
- **Attributes:** Karakteristik atau fitur dari sebuah item (misal: Nama, Umur, Gaji).
- **Links:** Hubungan antar items, terutama pada data bertipe *network*.
- **Positions:** Data lokasi secara spasial (garis lintang/bujur)[cite: 6].

Tipe-Tipe Dataset (Dataset Types)

Munzner membagi dataset menjadi empat tipe utama:

- 1 **Tables:** Data dalam bentuk baris (*items*) dan kolom (*attributes*).
- 2 **Networks & Trees:** Data yang memiliki hubungan (*links*) antar titik (*nodes*).
- 3 **Fields:** Data yang kontinu di ruang fisik (sering ditemukan dalam data spasial atau medis seperti MRI).
- 4 **Geometry:** Informasi tentang bentuk fisik 2D atau 3D.

Dataset Types - Date Type

| Name | Age | Shirt Size | Favorite Fruit |
|---------|-----|------------|----------------|
| Amy | 8 | S | Apple |
| Basil | 7 | S | Pear |
| Clara | 9 | M | Durian |
| Desmond | 13 | L | Elderberry |
| Ernest | 12 | L | Peach |
| Fanny | 10 | S | Lychee |
| George | 9 | M | Orange |
| Hector | 8 | L | Loquat |
| Ida | 10 | M | Pear |
| Amy | 12 | M | Orange |

item: person

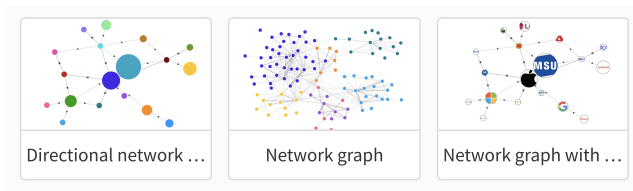
- **Flat Table:** Struktur paling umum dalam visualisasi.
- **Items:** Baris mewakili individu/entitas (Contoh: Orang).
- **Attributes:** Kolom mewakili dimensi data (Contoh: Umur, Nama).

Dataset Types - Date Type

| | | USA | | | Europe | | |
|-------|------|-----------------------|------------------------|-------|-----------------------|------------------------|-------|
| | | During last 12 months | Earlier than 12 months | Never | During last 12 months | Earlier than 12 months | Never |
| Men | 1997 | 9.5% | 8.3% | 82.2% | 9.2% | 11.3% | 79.5% |
| | 2002 | 16.6% | 7.8% | 75.6% | 14.4% | 8.2% | 77.4% |
| | 2012 | 12.2% | 22.9% | 64.9% | 13.7% | 18.5% | 67.8% |
| Women | 1997 | 12.1% | 7.0% | 80.9% | 12.6% | 8.8% | 78.6% |
| | 2002 | 15.5% | 8.4% | 76.1% | 17.6% | 8.2% | 74.2% |
| | 2012 | 13.4% | 23.6% | 63.0% | 14.9% | 16.9% | 68.2% |
| Total | 1997 | 10.8% | 7.7% | 81.5% | 11.0% | 10.0% | 79.0% |
| | 2002 | 16.0% | 8.1% | 75.9% | 16.0% | 8.2% | 75.8% |
| | 2012 | 12.8% | 23.3% | 63.9% | 14.3% | 17.7% | 68.0% |

- **Multidimensional Table:** Struktur paling umum dalam visualisasi.
- Indexing based on multiple keys seperti: jenis kelamin, tahun

Dataset Types - Date Type



- **Networks:** Fokus pada hubungan (*links*) antar entitas (*nodes*).
- **Trees:** Jaringan hierarkis tanpa siklus. often have roots and are directed

Dataset Types: Fields

Definisi

Dataset **Fields** mengandung atribut yang diukur secara kontinu di seluruh ruang fisik.

Elemen Pembentuk:

- **Grids:** Strategi pengambilan sampel untuk membagi ruang kontinu menjadi titik-titik diskrit.
- **Position:** Lokasi spasial tempat data diambil (misal: koordinat x, y, z).
- **Attributes:** Nilai kuantitatif yang diukur pada posisi tersebut (misal: suhu, tekanan).

Contoh Fields

Data Medis (MRI/CT-Scan), Data Meteorologi (Peta Tekanan Udara), dan Simulasi Aliran Fluida (Aerodinamika).

Fields vs Tables

Tables

Data bersifat diskrit/independen

Baris adalah item

Fokus pada hubungan atribut

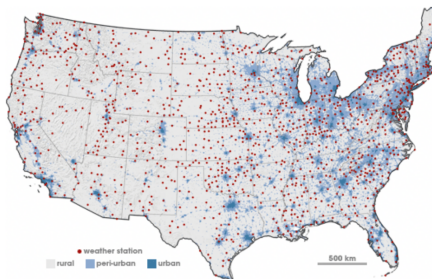
Fields

Data bersifat kontinu

Posisi ditentukan oleh grid

Fokus pada distribusi spasial

*Kunci utama Fields adalah **Interpolasi**: Menentukan nilai di antara titik sampel.*



Attribute Types

Atribut adalah variabel atau kolom data yang diukur. Munzner membaginya berdasarkan sifat matematikanya:

- ① **Categorical (Nominal):** Tidak ada urutan. Hanya untuk identitas/label.
- ② **Ordered:** Memiliki urutan yang jelas.
 - **Ordinal:** Ada urutan, tapi jarak antar nilai tidak terukur secara pasti.
 - **Quantitative:** Nilai numerik di mana operasi aritmatika bersifat valid.

Tabel: Perbandingan dan Contoh Atribut

| Tipe Atribut | Operasi Logis | Contoh Konkret |
|---------------------|--------------------|---------------------------------------|
| Categorical | = (Sama/Beda) | Nama Kota, Jenis Sensor, Warna Mata |
| Ordinal | <, > (Urutan) | Peringkat (1, 2, 3), Ukuran (S, M, L) |
| Quantitative | +, -, ×, / (Jarak) | Suhu, Pendapatan, Jumlah Populasi |

Penting bagi Desainer:

Jangan menggunakan saluran visual yang memiliki urutan (seperti gradasi gelap ke terang) untuk data **Categorical**, karena otak manusia akan secara otomatis mencari "siapa yang lebih tinggi" padahal data tersebut setara.

Latihan: Identifikasi Tipe Atribut

Tentukan tipe atribut untuk data berikut:

- **Nomor HP:** ... (Tips: Apakah nomor HP bisa dijumlahkan?)
- **Tingkat Kepedasan (Level 1-5):** ...
- **IPK Mahasiswa:** ...
- **Provinsi asal:** ...

Nomor HP adalah **Categorical (Nominal) meskipun berupa angka, karena ia hanya label unik.*

Ordering Direction

Untuk atribut yang terurut (*Ordered*), kita harus memahami ke mana arah data tersebut mengalir:

- 1 **Sequential:** Data mengalir dari satu ujung ke ujung lainnya (rendah ke tinggi).
- 2 **Diverging:** Data memiliki titik tengah yang krusial dan mengalir ke dua arah berlawanan.
- 3 **Cyclic:** Data yang nilainya berulang kembali ke titik awal.

Detail Karakteristik dan Contoh

- **Sequential (Satu Arah)**

- Contoh: Jumlah penduduk, tinggi badan, harga saham.
- *Warna*: Satu warna dengan intensitas berbeda (Light to Dark).

- **Diverging (Dua Arah)**

- Contoh: Perubahan suhu, laba vs rugi, korelasi (-1 hingga +1).
- *Warna*: Dua warna berbeda (misal: Merah - Putih - Biru).

- **Cyclic (Berulang)**

- Contoh: Jam, bulan, arah angin (derajat).
- *Warna*: Palet melingkar (*Circular color maps*).

Pesan Utama

Menggunakan palet **Sequential** untuk data **Diverging** akan menyembunyikan titik tengah yang penting (misal: kita tidak bisa melihat dengan cepat mana perusahaan yang rugi).

Mengapa Ini Penting?

Kesalahan Umum:

Menggunakan warna pelangi (*Rainbow colormaps*) untuk data sequential.

Akibatnya: Otak sulit menentukan mana yang lebih besar/kecil karena tidak ada urutan intensitas warna yang alami.

Solusi Terbaik:

Sesuaikan palet dengan struktur data.

→ Diverging = Menonjolkan ekstrem.

→ Sequential = Menonjolkan gradasi.

Ringkasan Abstraksi Data

Pesan Utama

Sebelum membuat grafik, kita harus mengidentifikasi:

- 1 Apa tipe dataset-nya? (Table, Network, dsb)
- 2 Apa saja atributnya? (Categorical vs Ordered)
- 3 Bagaimana karakteristik datanya? (Sequential vs Diverging)

Pilihan visualisasi yang salah (misal: menggunakan gradasi warna untuk data kategori) akan membingungkan audiens secara kognitif.