

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 8, 2026

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).

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

- **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.

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.