

# 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 18, 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).

# 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 - Tabel

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

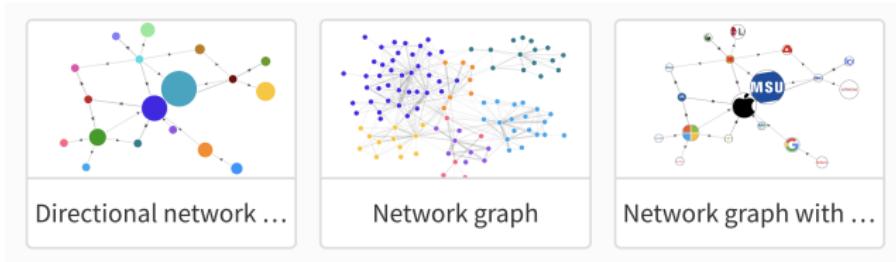
- **Flat Table:** Struktur paling umum dalam visualisasi.
- **Items:** Baris mewakili individu/entitas (Contoh: Orang).
- **Attributes:** Kolom mewakili dimensi data (Contoh: Umur, Nama).

# Dataset Types - Tabel

		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 - Network



- **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. Dataset tipe Fields digunakan ketika kita ingin merepresentasikan nilai yang ada di setiap titik dalam suatu ruang fisik. Nilai ini tidak "terputus", melainkan mengalir.

Contoh Dunia Nyata: Tekanan udara di atmosfer, suhu di dalam mesin mobil, atau aliran darah dalam pembuluh darah.

# Dataset Types - Fields

## Elemen Pembentuk:

- **Grids:** Strategi pengambilan sampel untuk membagi ruang kontinu menjadi titik-titik diskrit. Hal ini, karena kita tidak mungkin mengukur data di setiap titik atom di ruang nyata, kita menggunakan Grid untuk membagi ruang tersebut menjadi sel-sel kecil.
  - Fungsi: Menentukan di mana data diukur.
  - Jenis: Bisa berupa uniform grid (kotak-kotak rapi seperti piksel) atau irregular grid (bentuk yang menyesuaikan lekukan objek).
- **Position:** Lokasi spasial tempat data diambil (misal: koordinat x, y, z). Posisi adalah lokasi spasial yang menjadi tempat pengambilan sampel data.
  - Dalam Fields, posisi sering kali bersifat implisit berdasarkan struktur grid.
  - Misal: Pada peta cuaca, posisi didefinisikan oleh koordinat Lintang dan Bujur.

# Dataset Types - Fields

- **Attributes:** Nilai kuantitatif yang diukur pada posisi tersebut (misal: suhu, tekanan).
  - Atribut ini biasanya bersifat Quantitative (angka).
  - Contoh: Pada grid cuaca, atributnya bisa berupa Suhu ( $25^{\circ}\text{C}$ ) atau Kecepatan Angin (10 km/jam).

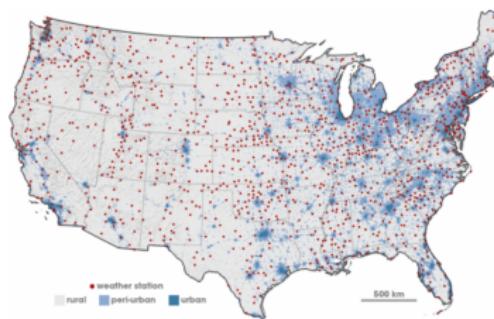
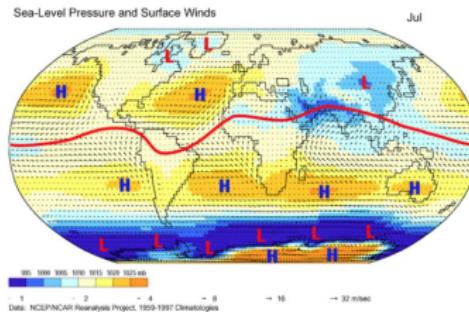
## Visualisasi Fields (Interpolasi)

Kunci utama dari dataset Fields adalah Interpolasi. Karena data hanya diambil di titik-titik grid, komputer harus "menebak" nilai di antara titik tersebut agar terlihat mulus (kontinu).

## Contoh Fields

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

# Dataset Types - Fields



# Fields vs Tables

Tables	Fields
Data bersifat diskrit/independen	Data bersifat kontinu
Baris adalah item	Posisi ditentukan oleh grid
Fokus pada hubungan atribut	Fokus pada distribusi spasial

# Dataset Types - Geometry

## Definisi

Geometry adalah tipe yang paling spesifik karena berkaitan langsung dengan bentuk fisik atau representasi spasial benda di dunia nyata.

- Elemen Utama Geometry. Dataset Geometry terdiri dari dua elemen kunci:
  - Items: Merupakan entitas objek fisik (misalnya: sebuah gedung, organ jantung, atau komponen mesin).
  - Positions: Informasi koordinat spesifik dalam ruang 2D atau 3D yang membentuk objek tersebut.

# Dataset Types - Geometry

- Karakteristik Penting
  - Spasial secara Intrinsik: Tidak seperti tabel (di mana kita bisa menukar posisi baris tanpa mengubah makna), pada Geometry, posisi adalah data utama. Jika posisi diubah, bentuk objek akan hancur atau maknanya hilang.
  - Bentuk (Shape): Dataset ini menyimpan informasi tentang titik (points), garis (lines), kurva, hingga poligon yang membentuk permukaan atau volume objek.
  - Visualisasi Langsung: Karena datanya sudah berupa bentuk fisik, tugas utama visualisasi sering kali adalah memproyeksikan bentuk 3D tersebut ke layar 2D dengan pencahayaan dan sudut pandang yang tepat agar mudah dipahami manusia.
- Contoh Penggunaan.
  - Arsitektur dan Teknik: Model CAD (Computer-Aided Design) untuk rancangan bangunan atau jembatan.
  - Medis: Model 3D hasil rekonstruksi organ tubuh dari data CT-Scan (tulang, pembuluh darah).

# Perbedaan Utama: Geometry vs. Fields

Fields	Geometry
Nilai yang mengalir di ruang	bentuk fisik objek itu sendiri
Contoh: Suhu udara di dalam ruangan	Contoh: Bentuk furnitur di ruangan
Data diambil pada titik-titik <i>grid</i>	Data berupa titik, garis, atau poligon

*"Fields memberitahu kita apa yang terjadi DI DALAM ruang, sedangkan Geometry memberitahu kita APA isi ruang tersebut."*

# Attribute Types

## Defnisi

Tipe Atribut adalah fondasi paling krusial bagi seorang desainer visualisasi. Atribut adalah variabel atau kolom data yang diukur, dan tipenya menentukan saluran visual (channels) mana yang efektif digunakan (seperti warna, posisi, atau ukuran).

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.

# Attribute Types

## ① Categorical (Kategorikal / Nominal)

Atribut kategorikal digunakan untuk membedakan identitas atau keanggotaan dalam suatu grup.

- **Sifat:** Hanya bisa dibandingkan apakah suatu item "sama" atau "berbeda" (= atau  $\neq$ ). Tidak ada urutan yang melekat secara alami.
- **Contoh:** Jenis kelamin, nama negara, jenis sensor, merek mobil.
- **Strategi Visual:** Gunakan Spatial Grouping (pengelompokan posisi), Color Hue (warna yang berbeda), atau Shape (bentuk)

## ② Ordered (Terurut)

Atribut ini memiliki urutan atau peringkat yang jelas. Munzner membaginya lagi menjadi dua:

- **Ordinal**  
Memiliki urutan logis, tetapi jarak atau interval antar nilai tidak dapat dihitung secara pasti atau tidak seragam.
- **Quantitative (Kuantitatif)**  
Atribut berupa angka murni di mana jarak antar nilai bersifat konsisten dan bermakna secara matematis.

# Attribute Types

- **Ordinal**

- Sifat: Kita bisa menentukan mana yang "lebih besar" atau "lebih kecil" (< atau >), tetapi tidak bisa melakukan operasi aritmatika.
- Contoh: Ukuran baju (S, M, L, XL), tingkat kepuasan (Puas, Cukup, Tidak Puas), tingkat kepedasan.
- Strategi Visual: Gunakan Color Saturation (kepekatan warna) atau Size (ukuran).

- **Quantitative (Kuantitatif)**

- Sifat: Kita bisa melakukan operasi aritmatika seperti penjumlahan atau perkalian (+, -, ×, /).
- Contoh: Berat badan, harga saham, suhu, jumlah populasi.
- Strategi Visual: Gunakan saluran yang paling akurat bagi mata manusia: Position on common scale (Posisi pada sumbu X atau Y)

# 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

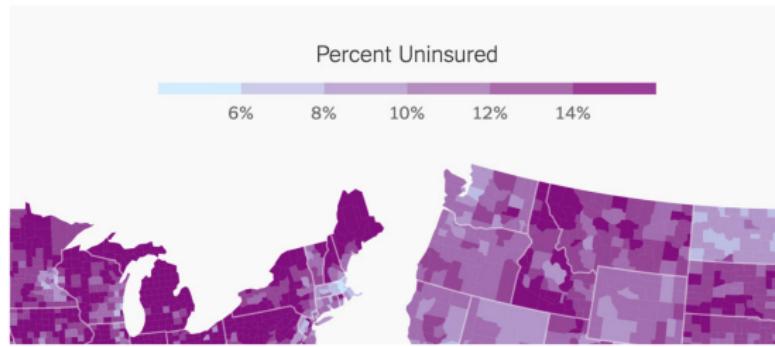
Ordering Direction membahas tentang struktur atau pola aliran dari data yang terurut (*Ordered Data*). Memahami arah ini sangat penting karena menentukan jenis palet warna atau skala visual yang harus digunakan agar tidak menyesatkan pembaca.

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.

# Ordering Direction - Sequential

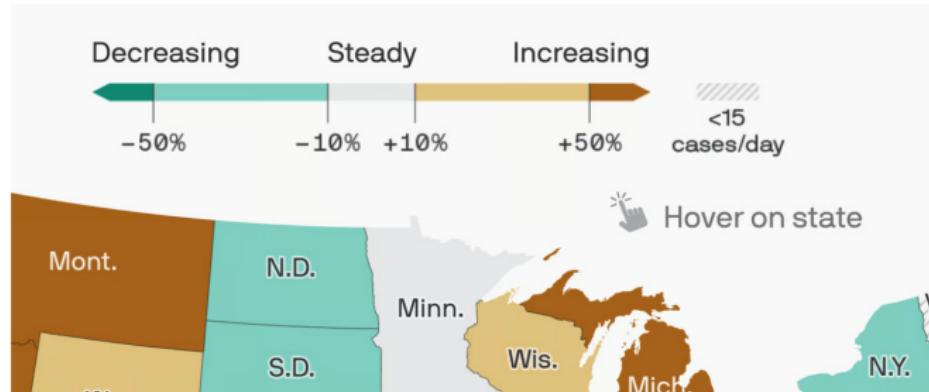
- Sequential (Sekuensial). Data yang mengalir dalam satu arah, biasanya dari nilai minimum ke nilai maksimum.
  - Karakteristik: Tidak memiliki titik tengah yang dianggap sebagai netral. Fokusnya adalah intensitas atau besaran dari rendah ke tinggi.
  - Contoh: Jumlah penduduk (0 hingga jutaan), usia (0 hingga 100 tahun), atau harga barang.
  - Penerapan Visual: Menggunakan gradasi satu warna yang berubah intensitasnya (misal: biru muda ke biru tua).



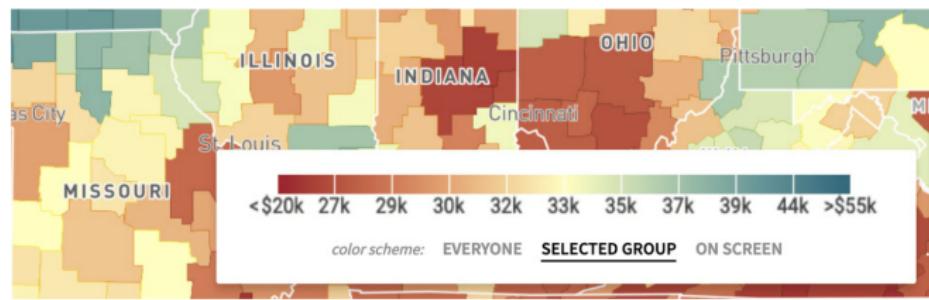
# Ordering Direction - Divergen

- Diverging (Divergen). Data yang memiliki dua arah berlawanan yang bertemu atau berpangkal di satu titik tengah yang krusial (titik nol atau nilai referensi).
  - Karakteristik: Fokusnya adalah seberapa jauh nilai menyimpang dari titik tengah ke arah positif atau negatif.
  - Contoh: Laba dan rugi perusahaan (titik nol adalah impas), suhu udara (titik tengah  $0^{\circ}\text{C}$ ), atau hasil survei (titik tengah "Netral").
  - Penerapan Visual: Menggunakan dua warna yang kontras untuk masing-masing arah (misal: merah untuk negatif, biru untuk positif) yang keduanya memudar ke arah warna netral (seperti putih atau abu-abu) di tengah.

# Ordering Direction - Divergen



Diverging, classed color scale, Axios



Diverging, unclassed color scale, Opportunity Atlas

# Ordering Direction - Cyclic

- Cyclic (Siklik)

Data yang urutannya berulang kembali ke titik awal setelah mencapai akhir periode.

- Karakteristik: Tidak memiliki nilai "paling tinggi" atau "paling rendah" secara absolut karena sifatnya yang melingkar.
- Contoh: Waktu (jam dalam sehari), bulan dalam setahun, atau arah mata angin (derajat kompas).
- Penerapan Visual: Menggunakan palet warna melingkar (Circular color maps) di mana warna pada titik  $360^\circ$  sama dengan warna pada titik  $0^\circ$ .

# Ordering Directio - Karakteristik dan Contoh

## ① Sequential (Satu Arah)

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

## ② Diverging (Dua Arah)

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

## ③ 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).

# When to use sequential?

## Use a sequential color scale for a more intuitive reading

- Kemudahannya untuk langsung dipahami tanpa harus berpikir keras atau melihat legenda secara detail.
- Intuisi Alami (Tanpa Legenda)  
Skala sekuensial mengikuti pola "terang ke gelap" (atau sebaliknya). Otak manusia secara alami mengasosiasikan warna yang lebih gelap atau lebih pekat dengan nilai yang lebih tinggi/besar, dan warna yang lebih terang dengan nilai yang lebih rendah
- Bebas Ambiguitas  
Skala sekuensial menghilangkan kebingungan ini karena hanya ada satu alur intensitas warna. tidak terdapat masalah "Manakah yang lebih tinggi? Biru atau Merah?", "Apakah Merah berarti buruk/rendah atau justru sangat tinggi?"
- Fokus pada Satu Cerita Utama  
Sekuensial: Cocok jika Anda ingin menekankan nilai tertinggi (puncak data). Fokusnya adalah "di mana konsentrasi terbesarnya?"

# when to use diverging color scales?

# 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.