

Visualisasi Distribusi Data

Sumbu Vertikal dan Horizontal

Tujuan Pembelajaran

- Memahami konsep distribusi data dan pentingnya visualisasi
- Mengenal berbagai teknik visualisasi distribusi (boxplot, violin plot, histogram)
- Membedakan visualisasi vertikal dan horizontal
- Menerapkan teknik yang tepat sesuai konteks data
- Menginterpretasi hasil visualisasi distribusi data

Apa itu Distribusi Data?

Distribusi data menggambarkan bagaimana nilai-nilai dalam dataset tersebar. Ini menunjukkan:

- Pusat data (mean, median, modus)
- Penyebaran data (range, varians, standar deviasi)
- Bentuk distribusi (simetris, skewed, normal)
- Outlier dan nilai ekstrem

Mengapa Visualisasi Penting?

Pemahaman Cepat

Visualisasi mempermudah pemahaman pola dan tren dalam data

Deteksi Anomali

Mudah menemukan outlier dan nilai-nilai tidak wajar

Komunikasi Efektif

Membantu menyampaikan informasi kompleks dengan visual yang mudah dipahami

Visualisasi Sumbu Vertikal

Definisi: Data numerik diletakkan pada sumbu Y, kategori/kelompok pada sumbu X

Kegunaan: Membandingkan distribusi antar kelompok

- **Boxplot vertikal** - menunjukkan median, IQR, outlier
- **Violin plot vertikal** - menampilkan bentuk distribusi lengkap
- **Strip plot/Scatter vertikal** - setiap titik mewakili satu observasi

Boxplot Vertikal - Komponen

Bagian-bagian Boxplot:

- **Garis atas** = Max/whisker atas
- **Kotak atas** = Q3 (kuartil 3)
- **Garis tengah** = Median (Q2)
- **Kotak bawah** = Q1 (kuartil 1)
- **Garis bawah** = Min/whisker bawah
- **Titik** = Outlier



Violin Plot Vertikal

Definisi: Kombinasi boxplot dan density plot untuk menampilkan distribusi data secara detail

Keuntungan:

- Menampilkan bentuk distribusi (unimodal, bimodal, dll)
- Menunjukkan kepadatan data di setiap nilai
- Lebih informatif daripada boxplot saja

Visualisasi Sumbu Horizontal

Definisi: Data numerik diletakkan pada sumbu X, kategori/kelompok pada sumbu Y

Kegunaan: Menunjukkan tren distribusi across categories atau time series

- **Histogram horizontal** - frekuensi data dalam bins
- **Density plot horizontal** - kurva distribusi smooth
- **Ridgeline plot** - beberapa density plot terstrata

Histogram & Density Plot Horizontal

Histogram: Menampilkan frekuensi data dalam interval kelas

- ✓ Mudah dipahami | ✓ Menunjukkan skewness | ✗ Bergantung jumlah bins

Density Plot: Estimasi smooth dari distribusi probabilitas

- ✓ Lebih smooth | ✓ Mudah membaca area | ✗ Tidak menunjukkan data points

Ridgeline Plot

Definisi: Beberapa density plot yang saling tumpang tindih untuk membandingkan distribusi antar kategori

Kegunaan: Menampilkan tren perubahan distribusi (misal: data per bulan, per tahun)

Kelebihan:

- Efisien untuk membandingkan banyak kelompok
- Mencegah overlapping yang terlalu banyak
- Estetika visual yang menarik

Perbandingan Teknik Visualisasi

Panduan Memilih Teknik

- ☞ Gunakan Boxplot jika: Ingin membandingkan median & outliers cepat
- ☞ Gunakan Violin jika: Perlu melihat bentuk distribusi detail
- ☞ Gunakan Histogram jika: Data numerik tunggal, ingin lihat frekuensi
- ☞ Gunakan Ridgeline jika: Banyak kelompok, ingin bandingkan tren

Contoh Kasus: Nilai Ujian Mahasiswa

Data: Nilai ujian dari 3 kelas: A, B, C (masing-masing 40 siswa)

Pertanyaan: Kelas mana yang memiliki performa terbaik & teruniform?

Teknik yang tepat:

- Boxplot Vertikal (bandingkan median 3 kelas)
- Violin Plot (lihat distribusi detail setiap kelas)

Cara Interpretasi Visualisasi

🔍 Perhatikan:

- **Lokasi median** - nilai tengah distribusi
- **Lebar kotak/violin** - spread/varians data
- **Asimetri (skewness)** - distribusi miring ke kiri/kanan
- **Outlier/titik ekstrem** - nilai tidak wajar
- **Jumlah mode** - berapa puncak? (unimodal, bimodal, dll)

LATIHAN - SOAL 1

Soal: Anda memiliki data tinggi badan dari 200 mahasiswa di 4 angkatan (2021, 2022, 2023, 2024).

Anda ingin membandingkan distribusi tinggi badan antar angkatan.

Pertanyaan:

- a) Visualisasi apa yang paling tepat? Jelaskan alasannya!
- b) Apakah menggunakan sumbu vertikal atau horizontal? Mengapa?
- c) Apa informasi yang bisa kita dapatkan dari visualisasi tersebut?

LATIHAN - SOAL 2

Soal: Dalam sebuah boxplot vertikal, Anda melihat:

- Median berada di bagian atas kotak (lebih dekat Q3)
- Whisker bawah sangat panjang
- Terdapat beberapa outlier di bawah whisker

Pertanyaan:

- a) Apa tipe skewness dari data ini?
- b) Apa interpretasi dari pola ini terhadap data?

LATIHAN - SOAL 3

Soal: Anda punya data penjualan harian toko selama 12 bulan (500+ observasi). Ingin membandingkan pola penjualan antar bulan.

Dataset memiliki 2 bulan dengan penjualan bimodal (2 puncak), sementara bulan lain unimodal (1 puncak).

Pertanyaan:

- a) Teknik visualisasi mana yang bisa menangkap fenomena bimodal ini?
- b) Mengapa histogram atau boxplot kurang tepat?
- c) Apakah ridgeline plot cocok? Jelaskan!

LATIHAN - SOAL 4

Soal: Anda membandingkan 2 teknik visualisasi untuk data nilai ujian 3 kelas:

- **Opsi A:** Boxplot vertikal (3 box berdampingan)
- **Opsi B:** Histogram horizontal (3 histogram terpisah)

Pertanyaan:

- a) Mana yang lebih baik untuk membandingkan median 3 kelas? Mengapa?
- b) Dalam situasi apa histogram lebih berguna?
- c) Bagaimana jika kita ingin lihat KEDUANYA?

LATIHAN - SOAL 5

Soal Analisis Kritis:

Data AB-testing e-commerce menunjukkan conversion rate dari 2 versi website (A vs B) selama 4 minggu. Setiap minggu Anda punya 1000+ data poin.

Pertanyaan:



- a) Gambarlah rencana visualisasi Anda (sumbu mana untuk apa)
- b) Teknik apa yang akan digunakan? Berapa visualisasi sekaligus?
- c) Apa metrik penting yang perlu ditampilkan?

Kesimpulan

Pilih teknik visualisasi yang tepat berdasarkan:

✓ Jumlah grup yang dibandingkan

✓ Informasi apa yang ingin ditampilkan

Selamat Belajar!  

✓ Target audiens dan konteks analisis