

STA581 Sains Data: Data Visualization

Dr. Agus M Soleh



[https://www.stat.ipb.ac.id/agusms/
agusms@apps.ipb.ac.id](https://www.stat.ipb.ac.id/agusms/agusms@apps.ipb.ac.id)

Outline

- 1 Pendahuluan
- 2 Menyusun Grafik Data
 - Mengetahui Target Grafik Dibuat
 - Taxonomy untuk Grafik Data
 - Pewarnaan
- 3 Prinsip Keefektifan Grafik Data
 - Layout dari penyajian grafik
 - Desain dari penyajian grafik
- 4 Perkembangan Grafik Data
 - Grafik sebaran peubah diskret
 - Grafik sebaran peubah kontinu
 - Visualisasi data interaktif

Pendahuluan

- Visualisasi data dapat diidentikan dengan grafik atau diagram
- Grafik dan diagram seringkali banyak muncul pada publikasi
- Bahkan, banyak grafik dan diagram yang muncul secara periodik pada koran harian, majalah, jurnal teknik, dan laporan dari suatu perusahaan pada sektor swasta dan pemerintah.
- Fokus pada mata kuliah ini, yaitu untuk memandu prinsip-prinsip dasar dari grafik statistik, serta aturan untuk penyajian grafiknya.

Pendahuluan

- Ketika data diperoleh, maka terdapat 3 kemampuan untuk menyajikannya untuk khalayak ramai, yaitu:
 - Literacy: kemampuan untuk menyajikan informasi pada data dengan tulisan
 - Numeracy: kemampuan untuk menyajikan informasi pada data dengan angka
 - Graphicacy: kemampuan untuk menyajikan informasi pada data dengan grafik
- Fokus: Graphicacy → dapat menyajikan data dalam bentuk gambar, sehingga mudah dipahami.

Menyusun Grafik Data

- Saat ini semua orang mudah membuat grafik, tetapi kualitas dari grafik yang dihasilkan berbeda
- Perlu framework untuk memahami grafik data yang akan memperbarui palet grafik data:
 - Mengetahui target grafik dibuat
 - Taxonomy untuk grafik data
 - Pewarnaan

Mengetahui Target Grafik Dibuat

- Sangatlah penting sebelum membuat grafik, kita mengetahui target orang-orang yang akan diharapkan mengetahui makna grafik tersebut.
- Sehingga apabila kita tahu target utamanya, maka bisa diharapkan banyak informasi yang dapat diambil oleh mereka.
- Terdapat 3 kelompok besar target yang dapat membedakan bentuk grafik yang dibuat:
 - Publik/Masyarakat luas → grafik lebih sederhana dan menarik, sehingga mudah untuk dipahami
 - Pengguna data kebijakan atau administratif → grafik dapat dipahami dengan jelas dengan aspek menarik tidak terlalu ditonjolkan
 - Kolega profesional → grafik dibuat kompleks dengan penyajian yang sesederhana mungkin, sehingga dalam satu grafik dapat diambil banyak sekali informasi

Mengetahui Target Grafik Dibuat

For public

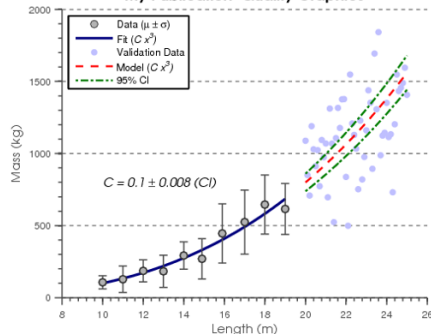


Zero Data Team | 0857 3108 4752 | zerodata-consultant.blogspot.com | @zerodataconsult

Sumber: unicef

For professional colleagues

My Publication-Quality Graphics



Taxonomy untuk Grafik Data

- Grafik data dapat dipahami dalam empat elemen dasar:
 - Isyarat Visual (*Visual Cues*)
 - Sistem Koordinat
 - Skala
 - Konteks

Isyarat Visual

- Isyarat visual adalah elemen grafis yang menarik perhatian sebagai fokus audiens.
- Terdapat 9 isyarat visual:
 - Posisi (numerik) dimana dalam hubungannya dengan hal lain?
 - Panjang (numerik) seberapa besar (dalam satu dimensi)?
 - Sudut (numerik) seberapa lebar? sejajar dengan sesuatu yang lain?
 - Arah (numerik) pada kemiringan berapa? Dalam deret waktu, naik atau turun?
 - Bentuk (kategorik) yang termasuk dalam kelompok mana?
 - Luas (numerik) seberapa besar (dalam dua dimensi)?
 - Volume (numerik) seberapa besar (dalam tiga dimensi)?
 - Bayangan (*Shade*) (lainnya) sampai sejauh mana? seberapa parah?
 - Warna (lainnya) sampai sejauh mana? seberapa parah? Waspadaai kebutaan warna merah/hijau

Sistem Koordinat

- Bagaimana data titik diatur?
- Terdapat tiga sistem koordinat yang paling umum:
 - Kartesius: Ini adalah sistem koordinat rectangular (x,y) yang sudah dikenal dengan dua sumbu tegak lurus.
 - Polar: Analog radial dari sistem Cartesian dengan titik-titik yang diidentifikasi oleh jari-jari dan sudutnya.
 - Geografis: Ini adalah sistem yang semakin penting di mana lokasi di permukaan bumi adalah lengkung, tetapi direpresentasikan dalam bidang dua dimensi.
- Pilihan yang tepat untuk sistem koordinat sangat penting dalam merepresentasikan data secara akurat

Skala

- Skala menerjemahkan nilai menjadi isyarat visual.
- Pilihan skala seringkali sangat penting.
- Pertanyaan utamanya adalah bagaimana jarak dalam grafik data diterjemahkan menjadi perbedaan kuantitas yang berarti?
- Setiap sumbu koordinat dapat memiliki skalanya sendiri dengan tiga pilihan berbeda:
 - Numerik: Kuantitas numerik paling sering ditetapkan pada skala linier, logaritmik, atau persentase.
 - Kategorikal: Peubah kategorik mungkin tidak memiliki urutan (contoh: Demokrat, Republik, atau Independen), atau mungkin ordinal (contoh: tidak pernah, mantan, atau perokok aktif).
 - Waktu: Waktu adalah besaran numerik yang memiliki beberapa sifat khusus. Pertama, karena sifat kalender → dapat dibatasi oleh serangkaian unit yang berbeda (misalnya: tahun, bulan, hari, dll.). Kedua, dapat dianggap sebagai periodik (atau siklus).
- Tersesat dengan skala itu mudah, karena berpotensi mendistorsi sepenuhnya posisi relatif titik data dalam grafik apa pun.

Konteks

- Tujuan grafik data adalah untuk membantu audiens membuat perbandingan yang bermakna
- Grafik data yang buruk dapat melakukan hal yang sebaliknya → memfokuskan perhatian pemirsa pada artefak yang tidak berarti, atau mengabaikan bagian penting dari pengetahuan yang relevan tetapi eksternal.
- Konteks dapat ditambahkan ke grafik data dalam bentuk:
 - Judul atau subtitle yang menjelaskan apa yang sedang ditampilkan
 - label sumbu yang memperjelas bagaimana unit dan skala digambarkan
 - titik atau garis referensi yang menyumbangkan informasi eksternal yang relevan.
- Walaupun demikian hindari grafik data yang berantakan dengan anotasi yang berlebihan

Pewarnaan

- Warna adalah salah satu isyarat visual yang paling tajam, tetapi paling sering disalahpahami dan disalahgunakan.
- Dalam membuat pilihan warna, ada beberapa ide kunci yang penting:
 - 1 Warna dan bayangan monokromatiknya adalah dua isyarat visual yang paling kurang dipahami. Jadi, meskipun berpotensi berguna untuk sejumlah kecil level variabel kategori, warna dan bayangan bukanlah cara yang tepat untuk merepresentasikan variabel numerik, terutama jika perbedaan kecil dalam kuantitas tersebut penting untuk dibedakan.
 - 2 Kira-kira 8 persen dari populasi memiliki beberapa bentuk buta warna. Paling umum, ini membuat mereka tidak mampu melihat warna secara akurat, terutama membedakan antara merah dan hijau juga kuning dan biru.

Pewarnaan

- Cynthia Brewer telah mengembangkan palet anti buta warna dalam berbagai corak untuk tiga jenis data numerik dalam satu variabel:
 - Sequential: Pengurutan data hanya memiliki satu arah.
 - Diverging: Pengurutan data memiliki dua arah.
 - Kualitatif: Tidak ada urutan data, dan hanya perlu warna untuk membedakan kategori yang berbeda.

Prinsip Keefektifan Grafik Data

- Grafik data adalah representasi visual dari data. Ini bisa berupa data mentah asli atau ringkasan data yang dihasilkan oleh analisis statistik, atau kombinasi keduanya.
- Salah satu prinsip yang mendasari grafik statistik yang efektif adalah bahwa harus ada rumus atau resep untuk tata letak gambar yang menyajikan informasi statistik
- Prinsip keefektifan grafik data, dapat dibagi 2 ide utama:
 - Layout dari penyajian grafik
 - Desain dari penyajian grafik

Layout dari penyajian grafik

- Dibutuhkan skala dan frame grafik
- Skala:
 - memilih skala untuk mencakup data yang ekstrem
 - memilih skala transformasi yang sesuai untuk variabel kuantitatif, seperti logaritmik, eksponensial, atau tidak sama sekali
 - memilih skala ordinal untuk variabel kategori
- Frame Grafik: layout dari gambar
 - Pilihan font dan ukuran font untuk karakter yang dicetak, termasuk karakter plot, dan ketebalan garis

Desain dari penyajian grafik

Terdapat 6 prinsip dari desain (disingkat ACCENT):

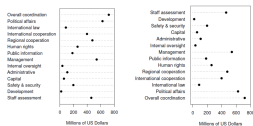
- Apprehension (Pemahaman) menyangkut kemampuan untuk melihat hubungan di antara data dengan benar → Subtansi data bukan oleh desain artistic atau lainnya
- Clarity (Kejelasan) menyangkut kemampuan untuk membedakan elemen-elemen grafik secara visual → setiap level jelas
- Consistency (Konsistensi) menyangkut kemampuan menafsirkan grafik berdasarkan keterpaparan pada grafik serupa di masa lalu atau standar universal yang wajar.
- Efficiency (Efisiensi) menyangkut kemampuan grafik untuk menggambarkan data dengan cara sesederhana mungkin.
- Necessity (Kebutuhan) berkaitan dengan kebutuhan grafik atau elemen tertentu.
- Truthfulness (Keadaan sebenarnya) menyangkut apakah data diplot dengan benar dalam sistem koordinat yang ditentukan dengan baik.

Perkembangan Grafik Data

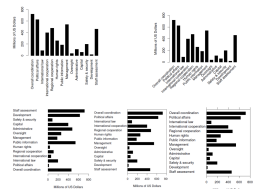
- Grafik data berkembang dari grafik yang statik ke visualisasi data interaktif
- Pada awalnya dimulai dengan melihat pola sebaran data:
 - Sebaran peubah diskret \rightarrow umumnya disebut chart
 - Sebaran peubah kontinu \rightarrow umumnya disebut plot

Grafik sebaran peubah diskret

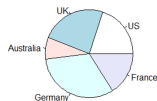
• Dotchart



• Barchart

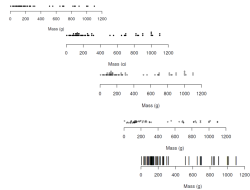


• Piechart

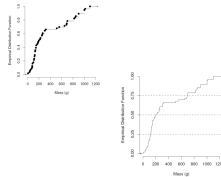


Grafik Sebaran peubah kontinu

- Dotplot



- Stemplot
- Boxplox
- EDF Plot



Visualisasi data interaktif

Umumnya menggunakan web browser, seperti:

- Rich Web content menggunakan D3.js and htmlwidgets
- Visualisasi dinamik menggunakan ggvis
- Web interaktif apps dengan Shiny

Akhir materi 4...