

MODUL 8

VISUALISASI DATA DENGAN R PACKAGE : GGLOT2

A. Tujuan Praktikum

Menggunakan konsep *Grammar of Graphics* untuk membuat visualisasi data.

B. Alokasi Waktu

1 x pertemuan = 120 menit

C. Dasar Teori

R merupakan bahasa pemrograman yang terkenal akan kemampuannya dalam menghasilkan grafik atau visualisasi data dengan baik. Penting diketahui bahwa R memiliki berbagai sistem dan paket untuk pembuatan grafik, contohnya *base*, *lattice*, *ggplot*, dan lain-lain. Namun dalam modul ini kita akan fokus menggunakan sistem *ggplot* untuk membuat visualisasi data.

Sistem pembuatan grafik dengan *ggplot* dapat dilakukan dengan menggunakan paket *ggplot2* yang merupakan implementasi dari konsep *Grammar of graphic* untuk bahasa pemrograman R. Dengan memahami konsep dari *grammar of graphic*, kita dapat membuat berbagai jenis plot dengan ringkas dan mudah. Sekarang coba aktifkanlah paket *ggplot2* tersebut terlebih dahulu

```
library(_____)
```

Dalam contoh ini kita akan membuat grafik dari *dataset diamonds* yang tersedia dalam paket *ggplot2*. Kita dapat melihat isi serta dokumentasi dari *dataset* tersebut dengan menjalankan perintah *diamonds* dan *?diamonds*. Berisi informasi apakah data *diamonds*? Silakan lakukan inspeksi pada struktur data tersebut

Kita dapat menggunakan fungsi *qplot* untuk membuat grafik menggunakan *ggplot2*. Bacalah terlebih dahulu dokumentasi fungsi *qplot* dan kemudian lengkapi *script* berikut untuk membuat grafik hubungan antara berat (sumbu x), harga (sumbu y) dan kejernihan intan (variasi warna) dari dataset *diamonds*!

```
qplot(x = ____, y = price, colour = ____, data = ____)
```

Pembuatan grafik dengan menggunakan fungsi *qplot* memang relatif mudah, namun fiturnya terbatas dan kurang fleksibel. Oleh karena itu, kita akan mempelajari dan menggunakan fungsi *ggplot* untuk membuat visualisasi data dengan lebih leluasa. Grafik diatas dapat diolah kembali dengan menggunakan penulisan kode sebagai berikut:

```
ggplot(data = diamonds, mapping = aes(x = carat, y = price, colour = clarity)) +  
  geom_point()
```

Hal yang menarik dan membedakan antara pembuatan grafik menggunakan *ggplot2* dan *base* adalah bahwa keluaran fungsi *ggplot* dapat disimpan sebagai obyek R. Apa manfaatnya? Pertama adalah kita dapat dengan mudah menyimpan grafik seperti halnya menjalankan *write.csv()* pada dataset. Kedua, kita dapat dengan leluasa melakukan modifikasi pada grafik yang telah dibuat. Hal ini akan dibahas dalam subbagian kedepan.

Baris kode dalam *script* berikut menunjukkan cara untuk menyimpan grafik ke dalam obyek R bernama `plot_diamonds` dan kemudian menyimpannya dalam komputer dengan nama berkas "*diamonds.png*". Kita akan menggunakan fungsi `ggsave` yang juga berasal dari paket `ggplot2`.

```
plot_diamonds <- ggplot(data = diamonds) +  
  geom_point(mapping = aes(x = carat, y = price, colour = clarity)) # Saat  
  output disimpan ke dalam obyek R, grafik tidak otomatis dicetak pada layar  
  
plot_diamonds # Untuk mencetak grafik, Anda harus menjalankan nama obyek R  
yang sebelumnya dibuat  
  
ggsave(filename = "diamonds.png", plot = plot_diamonds)  
#> Saving 7 x 5 in image
```

Meskipun penulisan kode R untuk membuat grafik menggunakan fungsi `ggplot` relatif lebih panjang, lebih banyak kostumisasi dan pengaturan grafik yang dapat dilakukan dengan menggunakan fungsi `ggplot` dibandingkan fungsi `qplot`. Dalam modul ini kita akan membuat banyak grafik dengan menggunakan struktur penulisan kode sebagai berikut:

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Dapat ditarik kesimpulan berdasarkan struktur penulisan kode R di atas bahwa setidaknya terdapat tiga komponen utama untuk membuat grafik, yaitu:

1. *Data*
2. *Aesthetic mapping*
3. *Geometric object*

Pelajarilah dokumentasi fungsi `aes()` dan `geom_point()` (sebagai salah satu contoh *geometric object*) melalui kode berikut

```
?aes  
?geom_point
```

Selanjutnya kita akan bereksperimen membuat visualisasi data Ujian Nasional tingkat SMP di Kota Bandung yang disediakan oleh Open Data Kota Bandung. Data tersebut tersedia sebagai berkas "*un_smp.csv*" dalam subdirektori `data-raw`. Silakan impor data tersebut sebagai obyek R bernama `un_smp` menggunakan fungsi `vroom` dan `here`, kemudian tampilkan hasil impor pada layar.

```
library(_____)  
library(_____)  
un_smp <- vroom(here("_____", "_____" ))  
un_smp
```

Pelajarilah struktur data '*un_smp*' tersebut. Ada berapa observasi dan variabel yang tersedia? Apa saja nama dari setiap kolom? Data tahun berapa sajakah yang tersedia pada data tersebut?

```
_____(un_smp)  
_____(un_smp)
```

Dalam modul ini kita akan membuat visualisasi hubungan antara nilai UN mata pelajaran matematika dan bahasa Indonesia. Namun sebelum itu, penting untuk diingat bahwa dalam sistem `ggplot2` suatu grafik dibangun atas tiga komponen utama yaitu *data*, *aesthetic mapping*, dan *geometric objects*. Komponen pertama (*data*) dapat diatur dengan menggunakan baris kode berikut:

```
ggplot(un_smp)
```

Selanjutnya kita perlu mendefinisikan dimensi mana dari data yang ingin digambarkan dalam grafik. Pendefinisian ini dilakukan dalam komponen *aesthetic mapping* (`aes`). Kita diminta untuk mempelajari hubungan antara nilai UN matematika versus bahasa Indonesia. Untuk itu, kita dapat mendefinisikan variabel `bahasa_indonesia` pada sumbu x dan `matematika` pada sumbu y.

```
ggplot(un_smp, aes(x = _____, y = _____))
```

Pendefinisian sumbu x dan y telah menghasilkan *layer* baru dalam grafik. Namun, kita masih perlu mendefinisikan bentuk dari grafik tersebut melalui komponen *geometric objects* (`geom_*`) sebelum grafik tersebut dapat dibaca. Tambahkan obyek geometri berupa titik di atas *layers* yang telah dibuat sebelumnya.

```
ggplot(un_smp, aes(x = _____, y = _____)) +  
  _____()
```

Sekarang kita ingin mengetahui bagaimanaka representasi dari status sekolah (Negeri vs Swasta) pada grafik. Kita dapat menambahkan fungsi `aes` pada obyek geometri untuk melakukan hal tersebut. Dalam contoh ini kita akan menggunakan *aesthetic* berupa warna titik untuk membedakan antar status sekolah.

```
_____ (un_smp, aes(x = _____, y = _____)) +  
  _____ (aes(colour = _____))
```

Bagaimana jika Anda ingin menambahkan informasi jumlah peserta ujian yang direpresentasikan oleh ukuran titik dalam grafik tersebut?

```
_____ (un_smp, aes(x = _____, y = matematika)) +  
  _____ (aes(colour = _____, size = _____))
```

Nampaknya jumlah peserta ujian juga memiliki hubungan dengan nilai ujian, benarkah? Namun sayang sekali grafik tersebut sekarang terlihat sangat “penuh” sehingga sulit membedakan antar titik. Apakah transparansi titik-titik tersebut dapat dimodifikasi? Ya! Kita dapat menambahkan argumen “*alpha*” (nilai 0 hingga 1) pada obyek geometri yang diinginkan. (Pertanyaan: Apa yang terjadi jika Anda menambahkan argumen “*alpha*” dalam fungsi `aes` pada obyek geometri?)

```
_____ (un_smp, aes(x = _____, y = _____)) +  
  _____ (aes(colour = _____, size = _____), alpha = 0.2)
```

Selanjutnya, kita ingin menganalisa lebih lanjut bagaimanakah hubungan antar nilai ujian nasional tersebut jika dibagi per kelompok tahun. Karena Anda sudah menggunakan empat dimensi untuk mempresentasikan data (sumbu x, sumbu y, warna titik, ukuran titik), sekarang mungkin saatnya Anda menggunakan pendekatan berbeda yaitu menggunakan `facet`! Tambahkan baris kode `facet_wrap(~tahun, scales = "free")` pada *script* berikut

```

_____ (un_smp, aes(x = _____, y = _____)) +
_____ (aes(colour = _____, size = _____), alpha = 0.2) +
facet_wrap(~tahun, scales = "free")

```

Hasil visualisasi telah dapat memberikan analisa baru yang dapat kita buat, namun sayangnya kita masih belum puas dengan grafik tersebut dalam hal estetika. Lengkapilah script berikut untuk melakukan modifikasi estetika pada grafik tersebut, kemudian simpanlah grafik tersebut dalam obyek R dengan nama `mtk_vs_ind` dan jangan lupa cetak hasilnya pada layar.

```

_____ <- ggplot(un_smp, aes(x = _____, y = _____)) +
_____ (aes(colour = _____, size = _____), alpha = 0.2) +
_____ (~tahun, scales = "free") +
labs(
  x = "Bahasa Indonesia",
  y = "Matematika",
  colour = "Status sekolah",
  size = "# peserta",
  title = "_____",
  subtitle = "Ujian Nasional SMP di Kota Bandung 2015-2019",
  caption = "Sumber: Open Data Kota Bandung"
) +
theme_light()

mtk vs ind

```

Sekarang saatnya menyimpan grafik tersebut.

```

ggsave("_____.png", plot = _____)
#> Saving 7 x 5 in image

```