

Visualisasi data menggunakan R

(A Subtitle Would Go Here if This Were a Class)

Ujang Fahmi

Pelajaran ke-4



Salam kenal dan selamat datang.

Semoga kita semua bisa saling berbagi pengalaman dan pengetahuan. Saya adalah Ujang Fahmi, Co-founder dan mentor Sadasa Academy.

Jika anda berada dan sedang membaca tutorial ini, maka kemungkinan anda adalah orang yang sedang ingin belajar data sains, atau mungkin ditugaskan untuk mempelajari R oleh institusi atau organisasi anda. Sama seperti saya dulu, dimana tanpa latar belakang engineering saya didiharuskan untuk belajar R, demi menyelesaikan tugas akhir dan akhirnya jadilah seperti saya sekarang ini.

Satu hal yang pasti, ini adalah langkah pertama dari banyak langkah yang harus dilalui, entah melalui lembaga resmi atau belajar secara mandiri. Jadi selamat belajar!!!

Ujang Fahmi,
Yogyakarta, 2021-10-09

*Materi yang disampaikan disimpan dan dokumentasikan **disini***

Apa gg dalam ggplot?

gg merupakan singkatan dari grammar of graphic, yaitu sebuah aturan pembuatan plot yang setidaknya terdiri dari dua bagian yaitu:

1. Data yang akan diplotkan di sumbu x dan y, disebut aesthetic (`aes()`) di `ggplot2`
2. Geometry atau geom, untuk menentukan jenis plot yang akan dibuat

Menyiapkan data

Setelah kita tahu gg selanjutnya untuk bisa membuat plot kita juga harus tahu data. Data ini yang kemudian akan ditempatkan sebagai sumbu x dan y. Oleh karena itu, hal yang perlu disiapkan ketika akan membuat plot adalah data yang akan diplotkan.

Menentukan sumbu x, y dan z

Di dalam ggplot, kita tidak hanya bisa menempatkan sumbu x dan y. Tapi kita bisa menempatkan data ketiga, yang biasa disebut saja sumbu z. Sumbu z ditempatkan sesuai dengan jenis plot yang akan dibuat.

Sumbu z yang paling sering digunakan di ggplot2 ada dua yaitu:

1. `color` atau `colour` untuk plot dengan basis `line`
2. `fill` untuk plot dengan basis kotak atau lingkaran

Memilih jenis plot

Untuk bisa memilih jenis plot kita bisa menggunakan sintaks `geom_*`, dimana `*` akan dijadikan sebagai jenis plotnya. Misalnya,

1. Untuk membuat plot line kita bisa menggunakan `geom_line()`
2. Untuk membuat plot column kita bisa menggunakan `geom_col()`
3. Untuk membuat plot bar kita bisa menggunakan `geom_bar()`

Menyimpan plot

Untuk menyimpan plot, kita bisa menggunakan tombol export atau menggunakan sintaks perintah `ggsave()`.

```
ggsave("mtcars.pdf")  
ggsave("mtcars.png")
```

```
ggsave("mtcars.pdf", width = 4, height = 4)  
ggsave("mtcars.pdf", width = 20, height = 20, units = "cm")
```

Jenis Plot

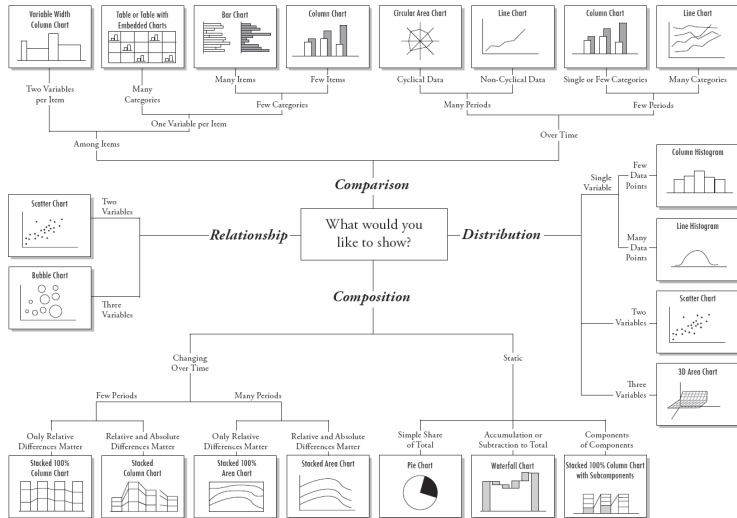


Figure 1: Jenis plot yang umum digunakan

Plot Distribusi

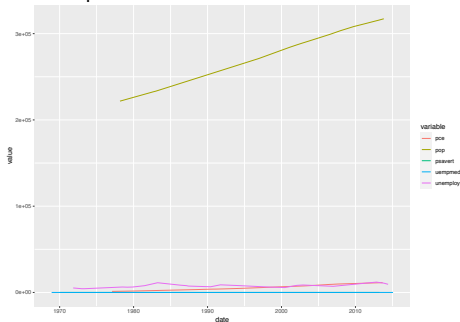
Plot distribusi biasanya dibuat menggunakan baris/line untuk menunjukkan perubahan nilai dalam sebuah rentang data.

Line Plot

```
library(tidyverse)

df1 = sample_n(economics_long, 100)
df1 %>%
  ggplot(aes(x = date,
             y = value,
             color = variable)) +
  geom_line()
```

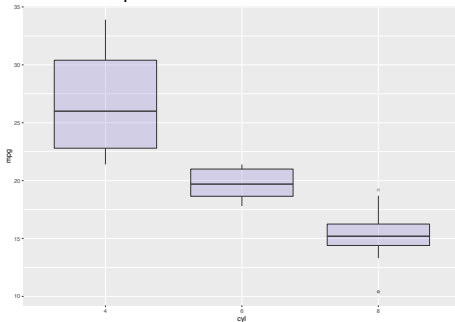
Contoh plot line:



Box-plot

```
mtcars %>%  
  ggplot(aes(x=as.factor(cyl),  
             y=mpg)) +  
  geom_boxplot(fill="slateblue",  
               alpha=0.2) +  
  xlab("cyl")
```

Contoh boxplot:



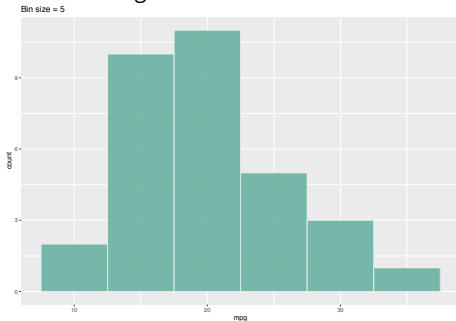
Histogram

```
library(hrbrthemes)

data <- mtcars

data %>%
  ggplot(aes(x=mpg)) +
    geom_histogram(binwidth=5,
                  fill="#69b3a2",
                  color="#e9ecef",
                  alpha=0.9) +
    ggtitle("Bin size = 5")
```

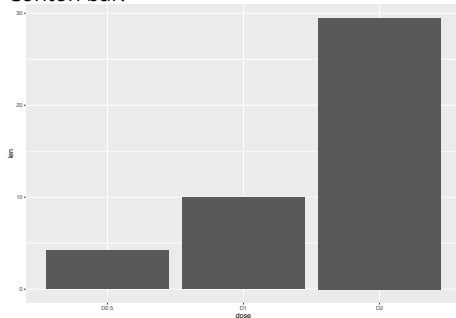
Contoh histogram:



Bar

```
df <- data.frame(  
  dose=c("D0.5", "D1", "D2"),  
  len=c(4.2, 10, 29.5))  
  
df %>%  
  ggplot(aes(x=dose, y=len)) +  
  geom_bar(stat="identity")
```

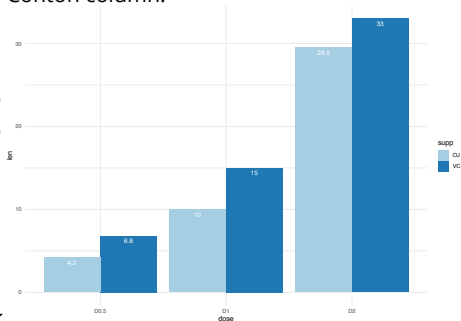
Contoh bar:



Column

```
df2 <- data.frame(  
  supp=rep(c("VC", "OJ"), each=3),  
  dose=rep(c("D0.5", "D1", "D2"), 2)  
  len=c(6.8, 15, 33, 4.2, 10, 29.5)  
  
df2 %>%  
  ggplot(aes(x=dose,  
             y=len,  
             fill=supp)) +  
  geom_col(position=position_dodge(), +  
  geom_text(aes(label=len),  
            vjust=1.6,  
            color="white",  
            position = position_dodge(0.9),  
            size=3.5)+  
  scale_fill_brewer(palette="Paired")+  
  theme_minimal()
```

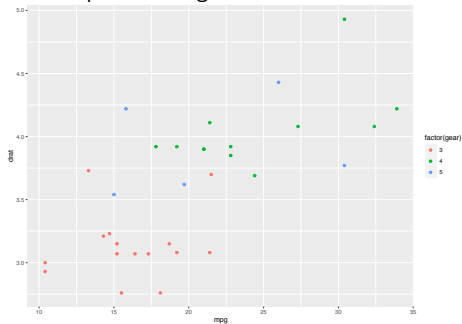
Contoh column:



Plot Hubungan

```
mtcars %>%  
  ggplot(aes(x = mpg,  
             y = drat)) +  
  geom_point(  
    aes(  
      color =  
        factor(gear)  
    )  
  )
```

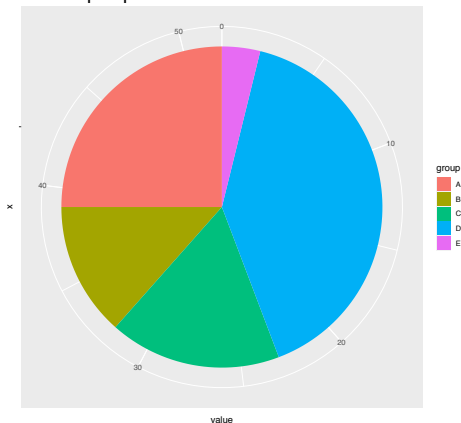
Contoh plot hubungan:



Pie chart dasar

```
data <- data.frame(  
  group=LETTERS[1:5],  
  value=c(13,7,9,21,2)  
)  
  
# Basic piechart  
data %>%  
  ggplot(aes(x="",  
             y=value,  
             fill=group)) +  
  geom_bar(stat="identity",  
           width=1) +  
  coord_polar("y", start=0)
```

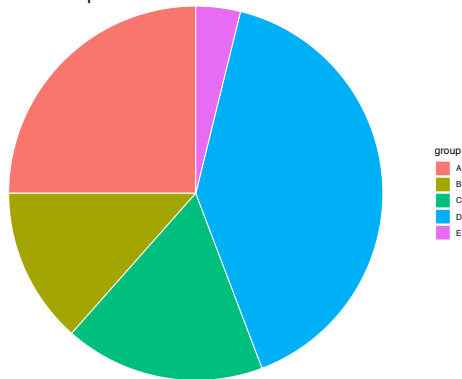
Contoh pie plot 1:



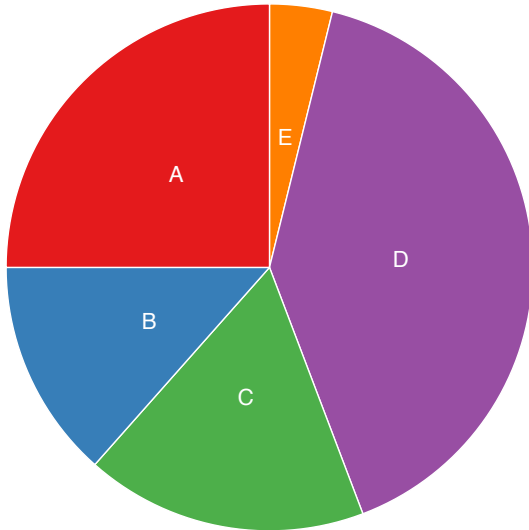
Memperbaiki pie chart 1

```
data %>%  
  ggplot(aes(x="",  
             y=value,  
             fill=group)) +  
  geom_bar(stat="identity",  
          width=1,  
          color="white") +  
  coord_polar("y", start=0) +  
  theme_void()
```

Contoh pie chart 2:



Memperbaiki pie chart 2



Your Turn

1. Membuat plot perbandingan
2. Membuat plot komposisi
3. Membuat plot hubungan
4. Membuat plot distribusi

Table of Contents

Menggunakan ggplot

- Apa gg dalam ggplot?

- Menyiapkan data

- Menentukan sumbu x, y dan z

- Memilih jenis plot

- Menyimpan plot

Jenis Plot

- Plot Distribusi

- Plot Perbandingan

- Plot Hubungan

- Plot Komposisi

Your Turn