



# KMMI 2021

## Eksplorasi dan Visualisasi Data

Pertemuan 9:  
Visualisasi Menggunakan paket ggplot2

# Sub CPMK

Mahasiswa dapat membuat visualisasi data menggunakan paket statistik seperti ggplot2.

## Pokok Bahasan

1. *install* dan *import* paket ggplot2
2. fungsi yang ada di paket ggplot2
3. membuat plot dengan fungsi qplot()
4. membuat plot dengan fungsi ggplot()
5. menginterpretasi plot

# Visualisasi dengan Paket ggplot2

Komponen visualisasi data:

- ✓ data yang akan divisualisasikan
- ✓ estetika visualisasi
- ✓ objek geometris
- ✓ transformasi statistik

Visualisasi data menggunakan paket ggplot2 memungkinkan kita untuk **menentukan komponen visualisasi** secara terpisah sehingga dapat membuat visualisasi yang kompleks dan berlapis.

Paket ggplot2 memiliki dua fungsi:

1. `qplot()` atau `quickplot()` : memiliki antarmuka yang mirip dengan `plot()`
2. `ggplot()` : menyediakan tata bahasa lengkap dari antarmuka grafis.

# Fungsi `qplot()`

- Fungsi `qplot()` dapat digunakan **sebagai pintasan** bagi yang sudah terbiasa menggunakan plot dasar seperti fungsi `plot()`.
- Untuk membuat jenis plot berbeda, **cara pemanggilan dapat dilakukan secara konsisten**. Hal ini berbeda jika menggunakan plot dasar dari R, misal menggunakan fungsi `hist()` untuk membuat histogram.

## Cara pemanggilan fungsi `qplot()`

```
qplot(  
  x,  
  y,  
  ...,  
  data,  
  facets = NULL,  
  margins = FALSE,  
  geom = "auto",  
  xlim = c(NA, NA),  
  ylim = c(NA, NA),  
  log="",  
  main = NULL,  
  xlab = NULL,  
  ylab = NULL,  
  asp = NA,  
  stat = NULL,  
  position = NULL  
)
```

**facets** : menghasilkan beberapa plot kecil yang mewakili subset data;

**margins** : faset tambahan (TRUE atau FALSE));

**geom** : menentukan geom yang akan digambar, jika x dan y ditentukan akan terbentuk titik-titik (scatter plot) dan jika hanya x yang ditentukan akan terbentuk histogram;

**xlim,ylim** : batas sumbu x dan y;

**log** : variabel mana yang akan diubah lognya;

**main** : memberikan judul plot;

**xlab,ylab** : label sumbu x dan label sumbu y;

**asp** : rasio aspek y/x;

**stat,position** sudah tidak digunakan lagi.

# Fungsi ggplot()

- Pemanggilan fungsi ggplot() dilakukan untuk menyediakan dataset dan pemetaan estetika (seperti warna, ukuran, dan lokasi x dan y) dengan aes().

```
a = ggplot(data = NULL, mapping = aes())
```

menambahkan lapisan-lapisan lain dengan menambahkan geom, facet, coordinate, scale, dan theme.

```
a + geom() + facet() + coordinate() + scale() + theme()
```

- Fungsi geom() untuk mewakili titik data serta properti estetika fungsi geom untuk mewakili variabel, misal geom\_point(aes(alpha, color, size))
- Fungsi facet() membagi plot menjadi subplot
- Fungsi coordinate() mengatur sistem koordinat dari plot.
- Fungsi scale() memetakan nilai data ke nilai visual sebuah estetika.
- Tema (*theme*) dapat digunakan untuk mengatur komponen plot selain data, seperti judul, latar belakang, garis kisi, font, label, dan legenda.

# Instalasi dan *Import* ggplot2

```
install.packages("tidyverse")  
library(ggplot2)
```

1

```
install.packages("ggplot2")  
library(ggplot2)
```

2

```
install.packages("devtools")  
devtools::install_github("tidyverse/ggplot2")  
library(ggplot2)
```

3

# Membuat dan Menginterpretasi Plot

a. *Scatter Plot*

b. Histogram

c. Dot Plot

d. *Boxplot*

e. Plot Densitas

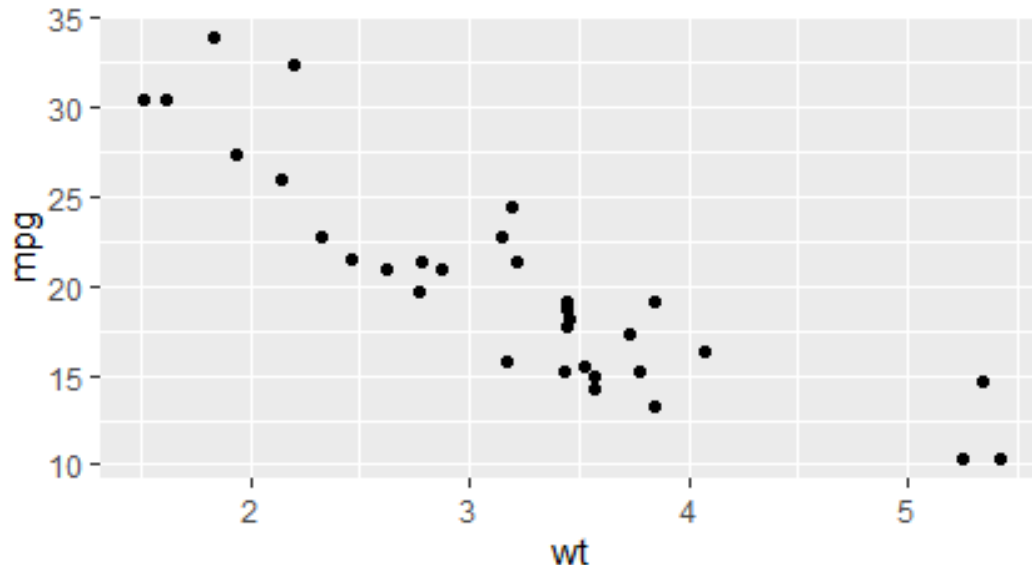
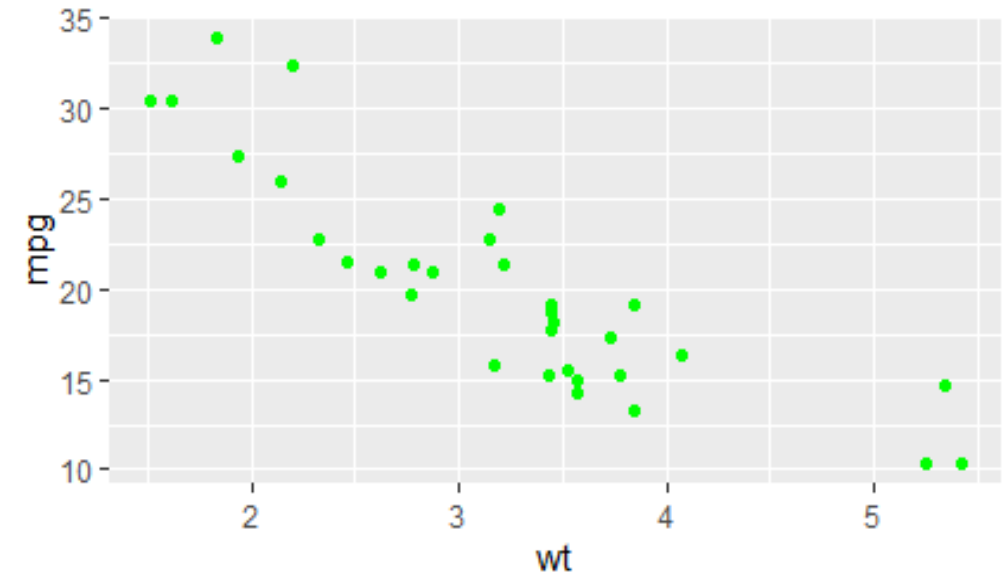
f. Diagram Batang

**a.**

Scatter plot antara variabel **mpg** (konsumsi bahan bakar) dan variabel **wt** (*weight*) dari data **mtcars** (*Motor Trend Cars Road Test*)

```
library(ggplot2)
qplot(wt, mpg, data = mtcars)
```

```
# To set aesthetics, wrap in I()
qplot(mpg, wt, data = mtcars, colour = I("green"))
```

**Gambar 1****Gambar 2**

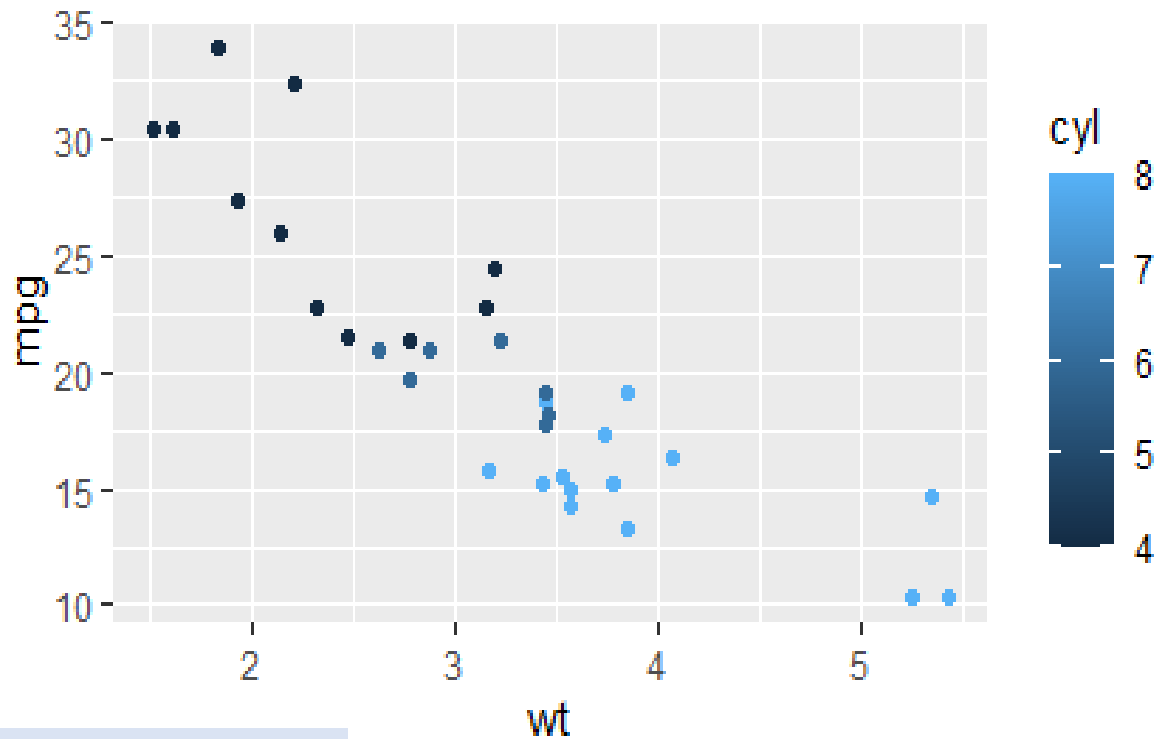
Ada korelasi negatif antara variabel mpg dan wt. *Motor Trend* yang semakin berat cenderung menghabiskan bahan bakar yang semakin sedikit.



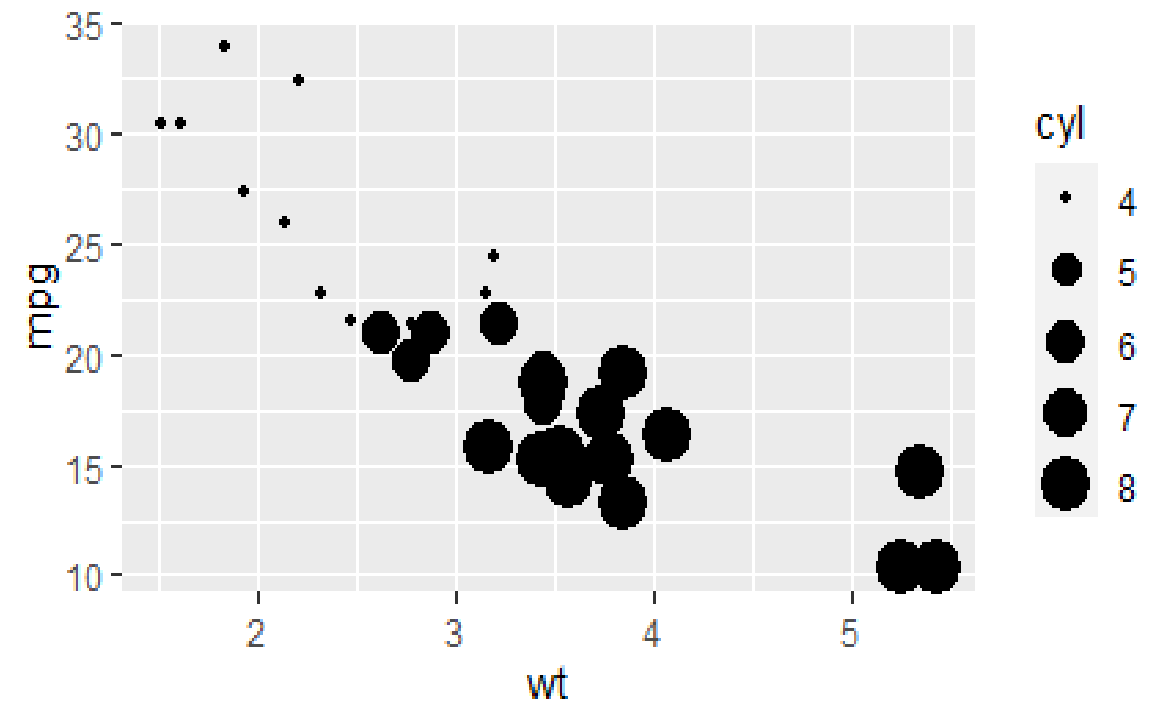
**a.**

Mobil yang **lebih berat** cenderung memiliki silinder yang **lebih banyak** pula.  
Konsumsi bahan bakar mobil **semakin menurun** seiring bertambahnya berat *Motor Trend*

```
library(ggplot2)
qplot(wt, mpg, data = mtcars, colour = cyl)
```

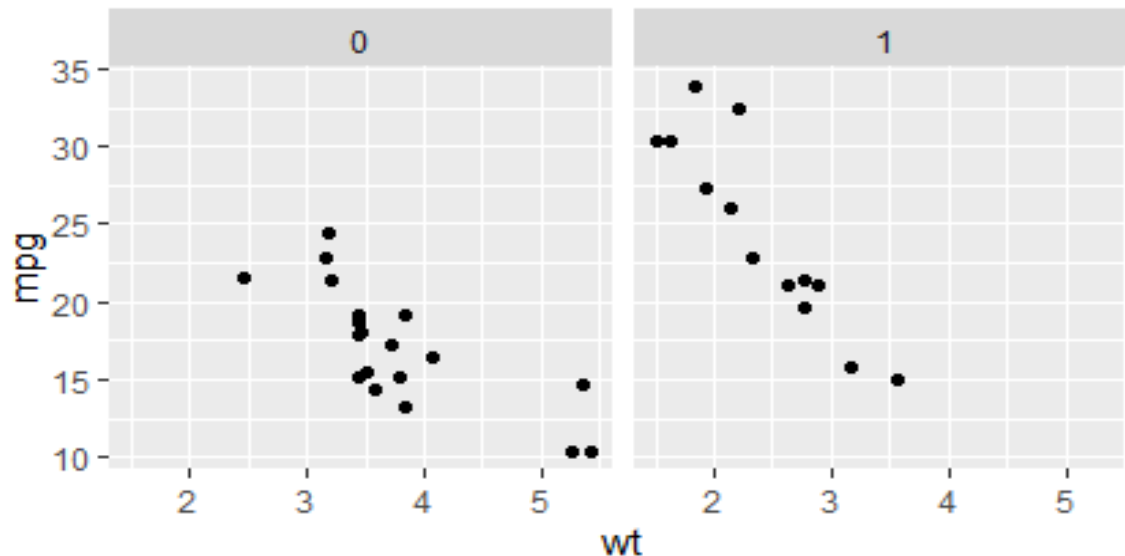
**Gambar 3**

```
library(ggplot2)
qplot(wt, mpg, data = mtcars, size = cyl)
```

**Gambar 4**

## a. *Scatter Plot*

```
library(ggplot2)
qplot(wt, mpg, data = mtcars, facets = . ~ am)
```



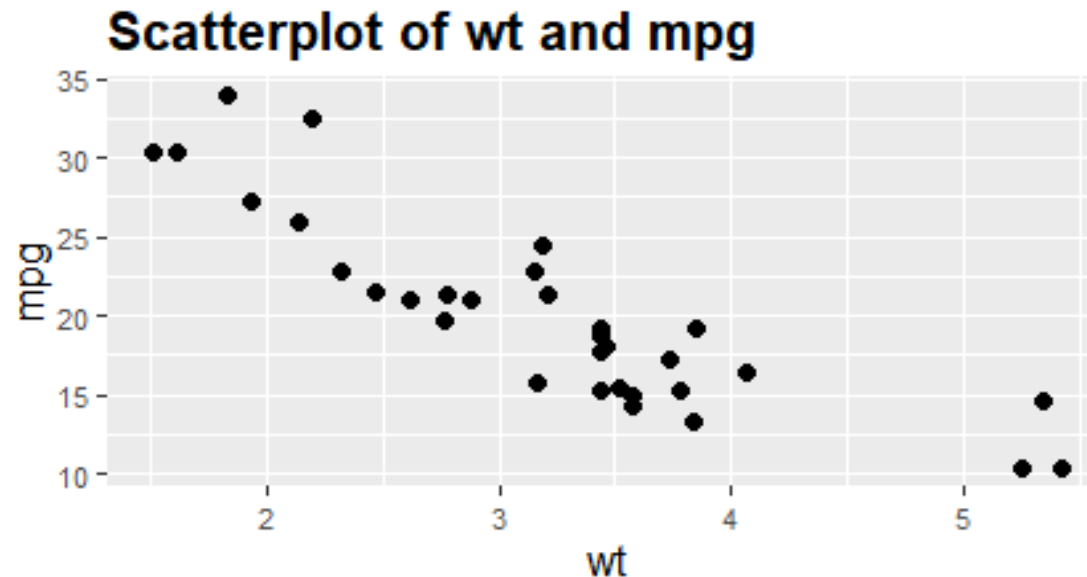
Gambar 5

- ✓ Dari 32 observasi *Motor Trend*, **transmisi otomatis (kode 0)** memiliki berat sekitar 2,5 sampai 5,5 dalam 1000 lbs dengan konsumsi bahan bakar dari 10 hingga 25 mpg.
- ✓ Untuk **transmisi manual (kode 1)**, berat *Motor Trend* sekitar 1,5 sampai 3,5 dalam 1000 lbs lebih rendah dibandingkan dengan transmisi otomatis.
- ✓ Jadi, konsumsi bahan bakar mobil dengan transmisi manual **cenderung lebih banyak** karena berat dan konsumsi bahan bakar berbanding terbalik.

**a.**

*Scatter plot* dengan fungsi `ggplot()` memuat fungsi `geom_point()` untuk mengatur objek geomteris, fungsi `labs()` untuk mengatur label, dan fungsi `theme()` untuk mengubah tema.

```
ggplot(data=mtcars, aes(x=wt, y=mpg)) +  
  geom_point(size=2) + labs(title = "Scatterplot of wt and mpg") +  
  theme(axis.text = element_text(size=8),  
        axis.title = element_text(size=12),  
        plot.title = element_text(size=15, face="bold"))
```

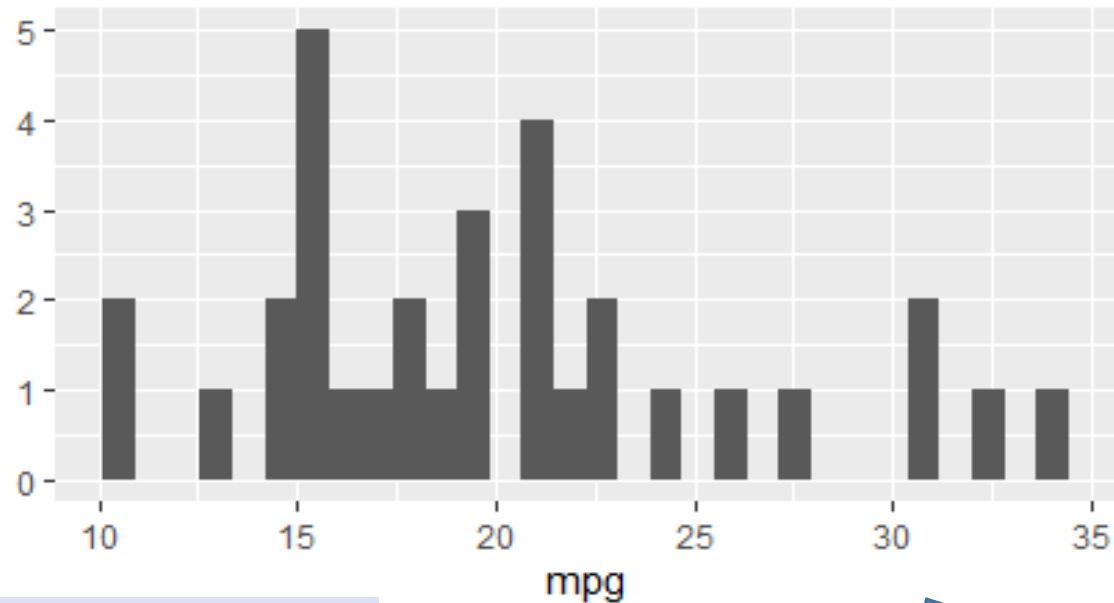


**Gambar 6**

## b. Histogram

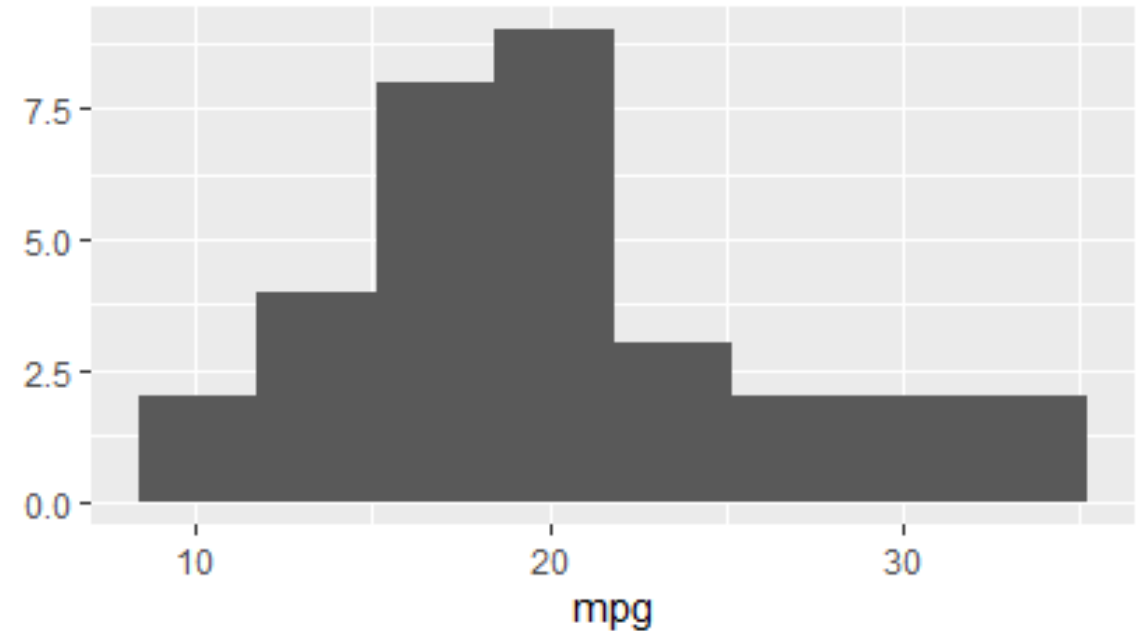
```
library(ggplot2)  
qplot(mpg, data = mtcars)
```

```
# atau dengan menentukan geom  
qplot(mpg, data = mtcars, geom = "histogram")
```



Gambar 7

Gambar 8

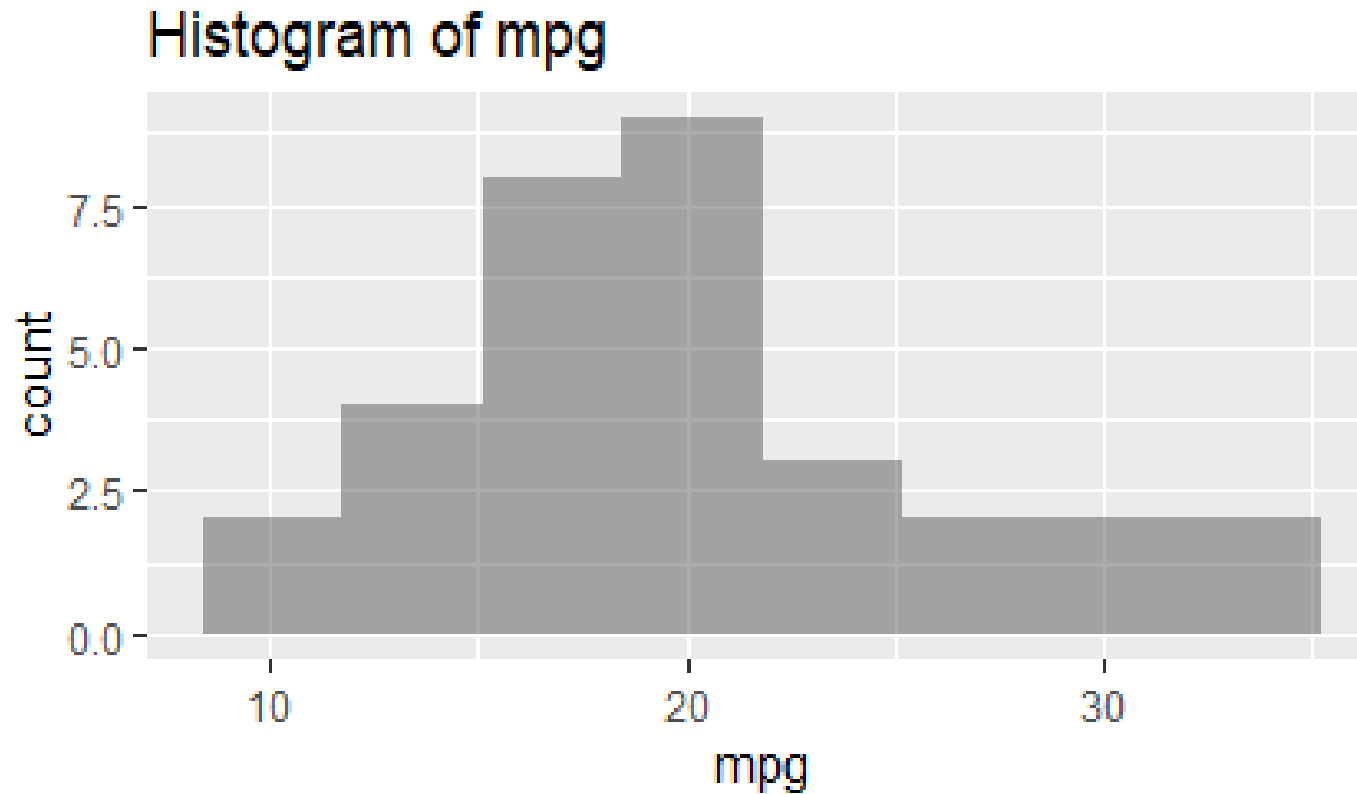


```
library(ggplot2)  
qplot(mpg, data = mtcars, geom = "histogram", bins = 8)
```

Batang yang cukup banyak membuat **distribusi data kurang terlihat** sehingga kode ditambahkan **bins** untuk **mengatur banyaknya batang**.

## b. Histogram

```
ggplot(data=mtcars, aes(x=mpg)) +  
  geom_histogram(alpha=0.5, bins = 8) +  
  ggtitle ("Histogram of mpg")
```

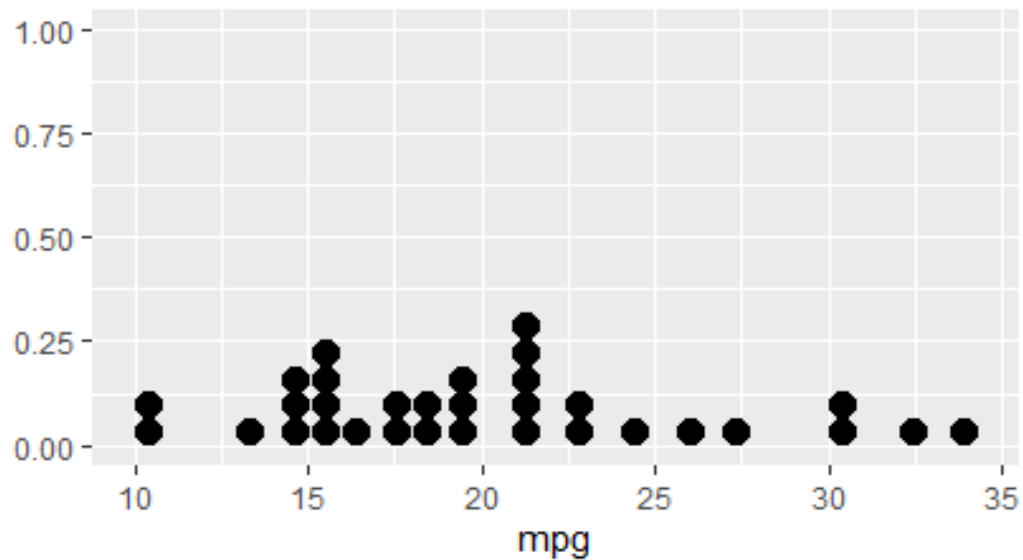


Gambar 9

## c. Dot Plot

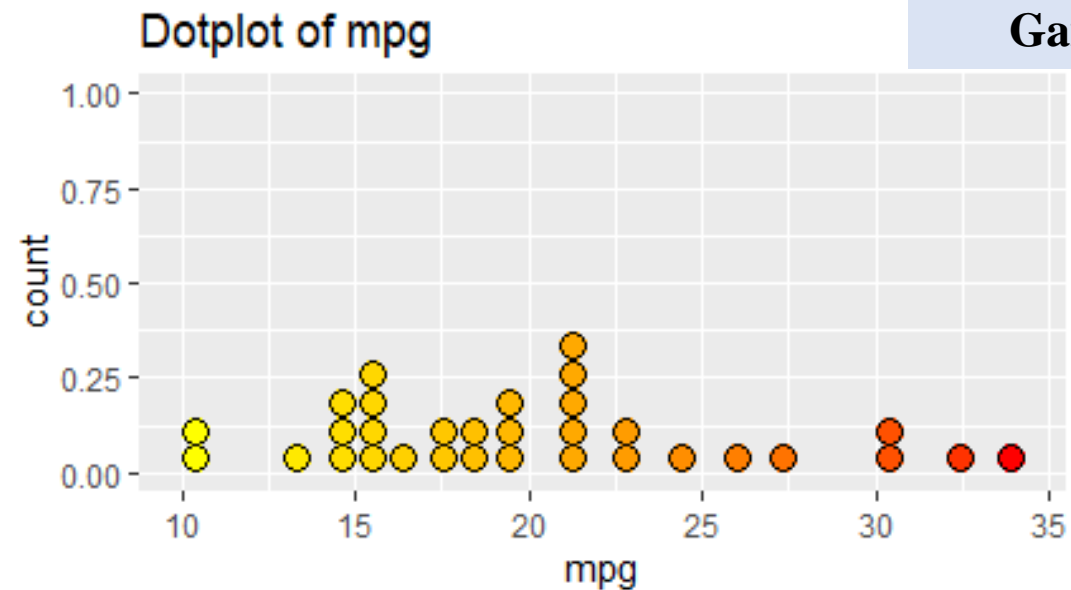
Nilai **median sekitar 19 mpg**, artinya 50% data berada di bawah 19 mpg dan 50% lainnya berada di atas nilai median.

```
library(ggplot2)
qplot(mpg, data = mtcars, geom = "dotplot")
```



Gambar 10

```
ggplot(data=mtcars, aes(x=mpg)) +
  geom_dotplot(aes(fill=..x..)) +
  ggtitle("Dotplot of mpg")+
  scale_fill_gradient(low="yellow", high="red") +
  guides(fill="none")
```

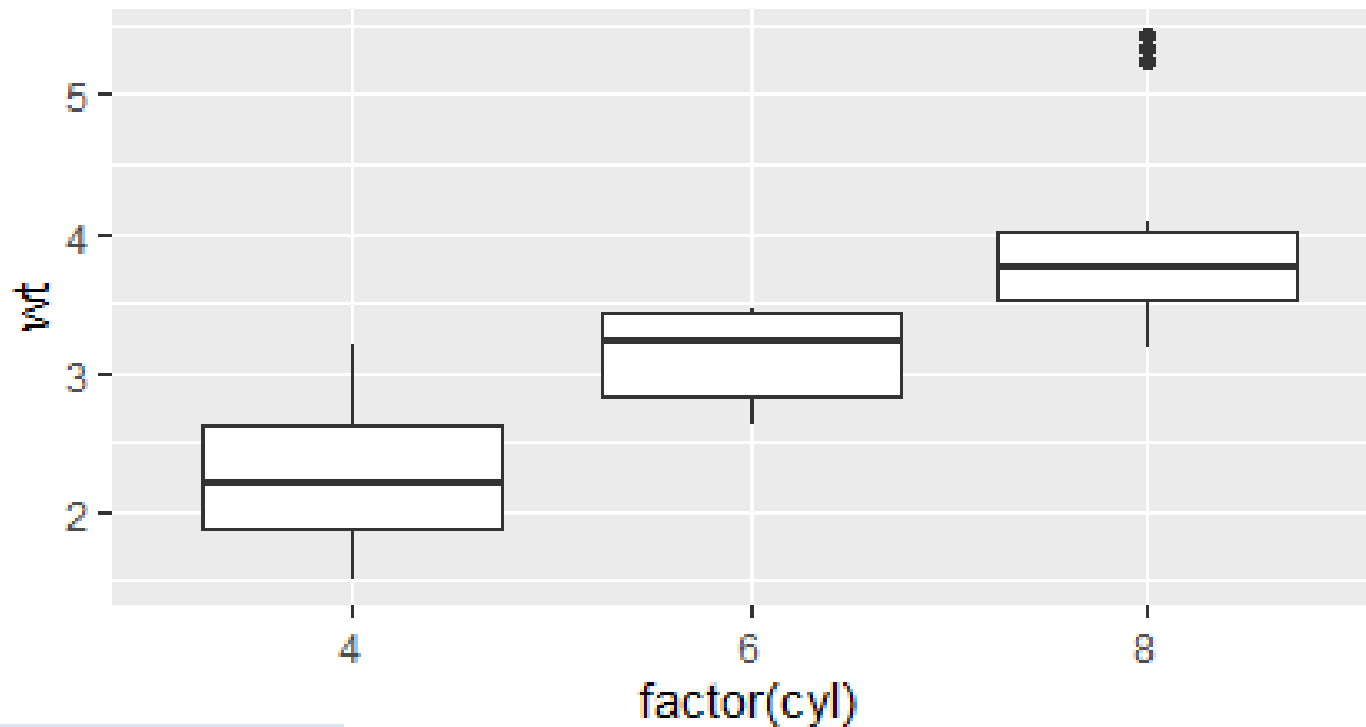


Gambar 11

Bandingkan sebaran data di bawah nilai median dan di atas nilai median, apa yang dapat disimpulkan?

## d. *Boxplot*

```
library(ggplot2)
qplot(factor(cyl), wt, data = mtcars, geom = ("boxplot"))
```



Gambar 12

Hasil boxplot selaras dengan *scatter plot* dalam Gambar 3 dan Gambar 4.

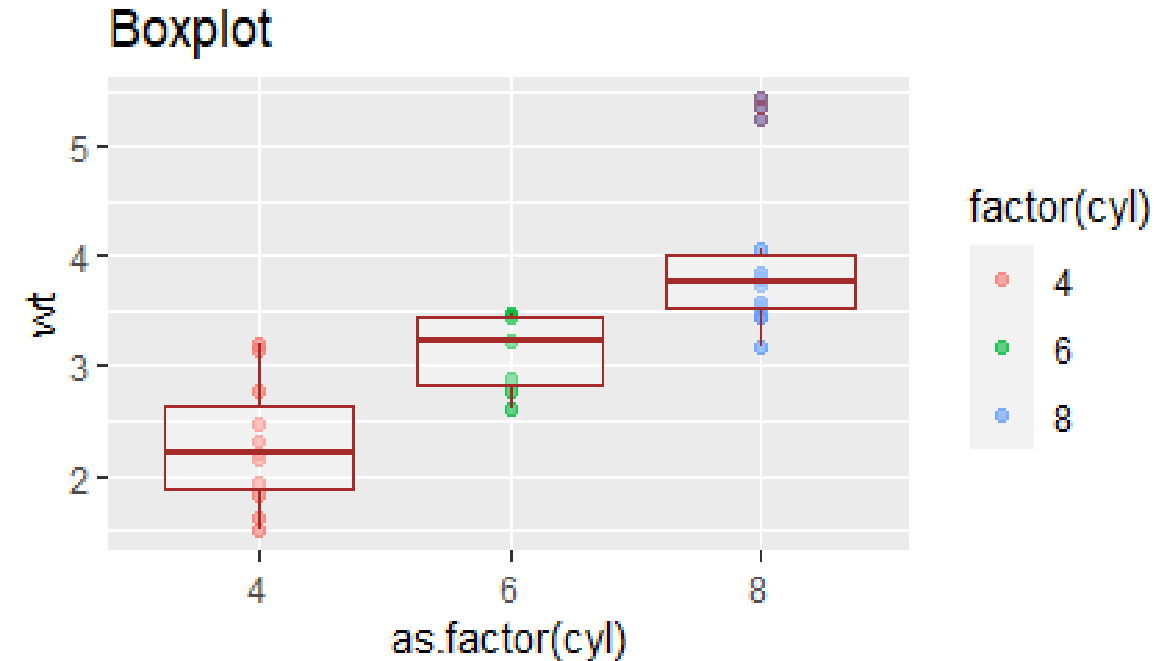
Nilai tengah berat mobil untuk jumlah silinder sebanyak 4 lebih rendah daripada silinder sebanyak 6 dan 8.

Untuk silinder sebanyak 6, *boxplot* menunjukkan adanya *data outlier*.

## d. *Boxplot*

```
library(ggplot2)
ggplot(data=mtcars, aes(x=as.factor(cyl), y=wt)) +
  geom_point(aes(color=factor(cyl)), alpha=0.6) +
  geom_boxplot(alpha=0.3, colour = "brown" ) +
  ggtitle("Boxplot")
# + guides(colour=FALSE) agar legenda tidak muncul
```

- Titik data untuk setiap kategori jumlah silinder berbeda-beda warnanya yang ditunjukkan oleh tanda bulatan di setiap *boxplot*.
- Observasi paling sedikit adalah observasi dengan jumlah silinder sebanyak 6.
- Data *outlier* ditunjukkan dengan titik data yang memiliki *outline* warna coklat (ada di silinder 8).



**Gambar 13**



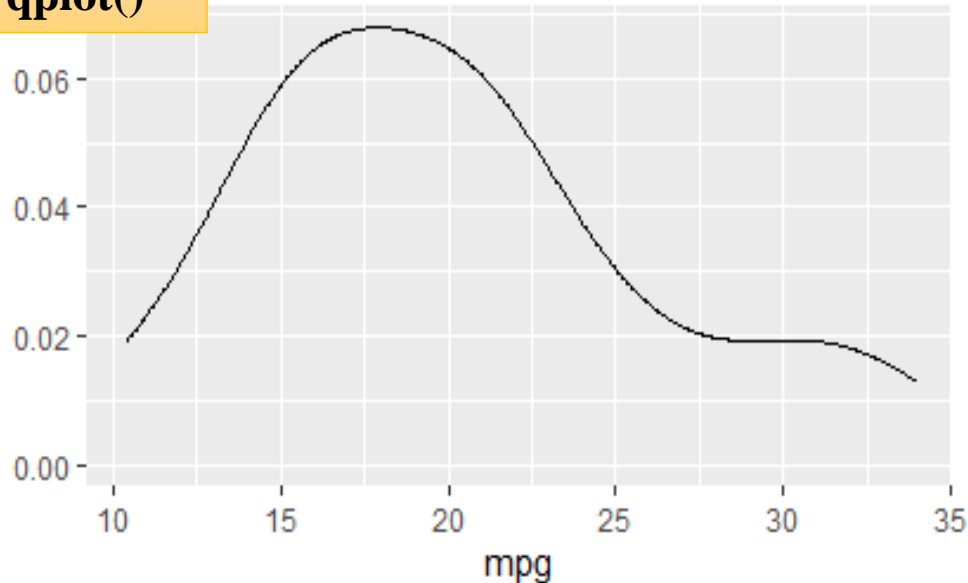
## e. Plot Densitas

```
library(ggplot2)
qplot(mpg, data = mtcars, geom = "density") # Gambar 14(a)

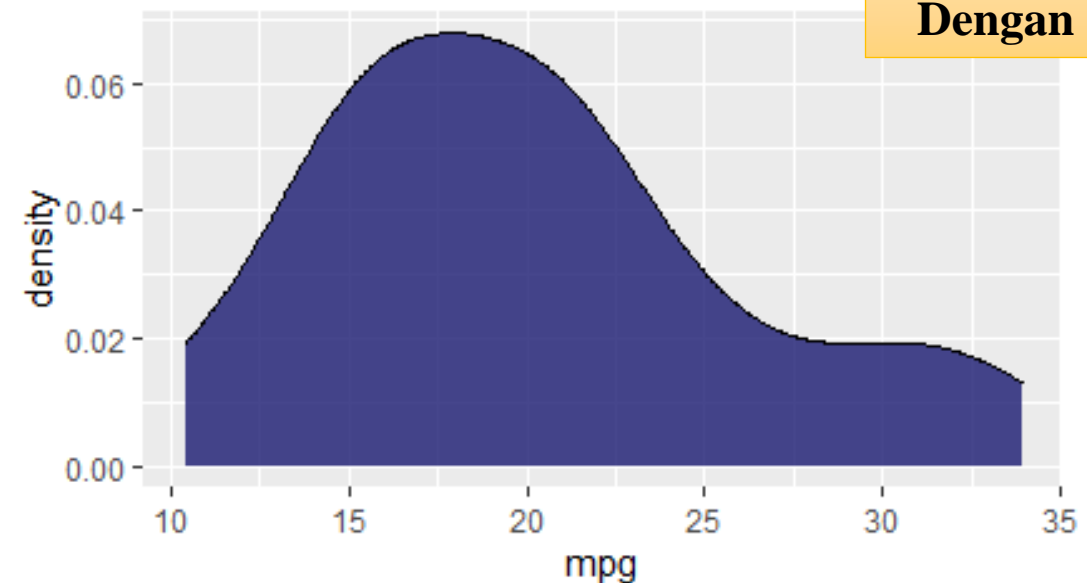
ggplot(data = mtcars, aes(x = mpg)) +
  geom_density(alpha= 0.8, fill = "midnightblue",
              color = "black") # Gambar 14(b)
```

Kurva densitas sesuai dengan histogram pada Gambar 8, yaitu distribusi variabel mpg **miring ke kanan** (memiliki kemiringan positif) yang terlihat dengan adanya ekor pada bagian kanan. Hal ini disebabkan karena adanya dugaan **data outlier**.

Dengan qplot()



Dengan ggplot()



Gambar 14

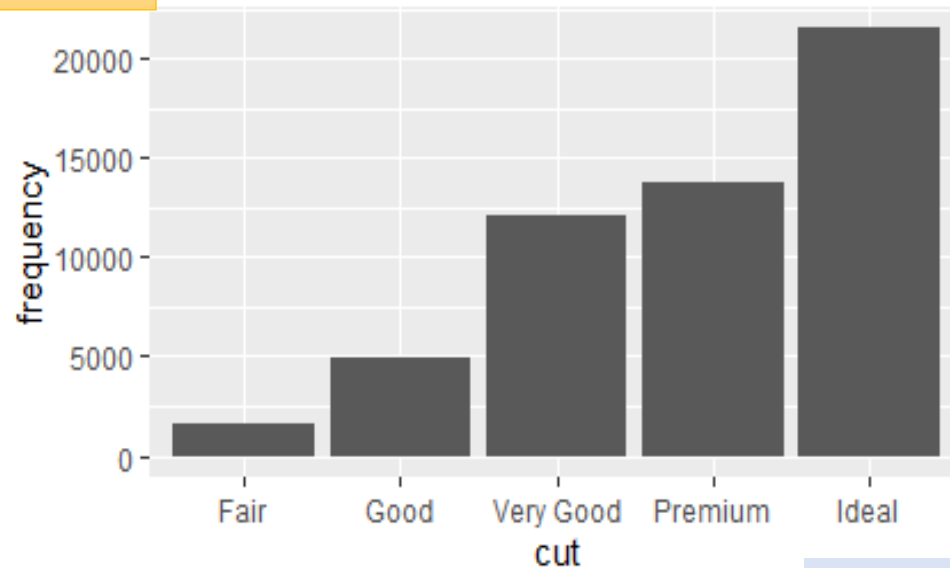
## f. Diagram Batang

Diagram batang **kualitas pemotongan berlian (cut)** yang diambil dari data diamonds, terdiri dari 53.940 observasi, terdapat 5 kategori, yaitu **fair**, **good**, **very good**, **premium**, dan **ideal**.

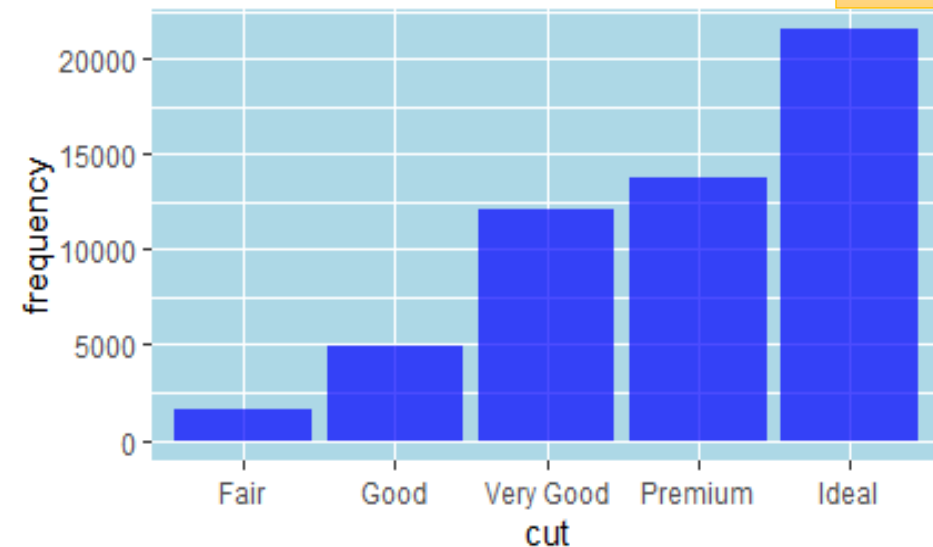
```
# membuat bar chart dengan qplot
qplot(cut, data = diamonds, geom = "bar", ylab = "frequency")
# membuat bar chart dengan ggplot
ggplot(diamonds, aes(cut)) +
  geom_bar(alpha=0.7, fill="blue") + labs(y="frequency") +
  theme(panel.background = element_rect(fill="light blue"))
```

Kualitas pemotongan berlian secara **ideal** memiliki frekuensi terbanyak, yaitu sekitar 22.000 berlian atau **sekitar 40% dari total observasi**.

Dengan qplot()



Dengan ggplot()

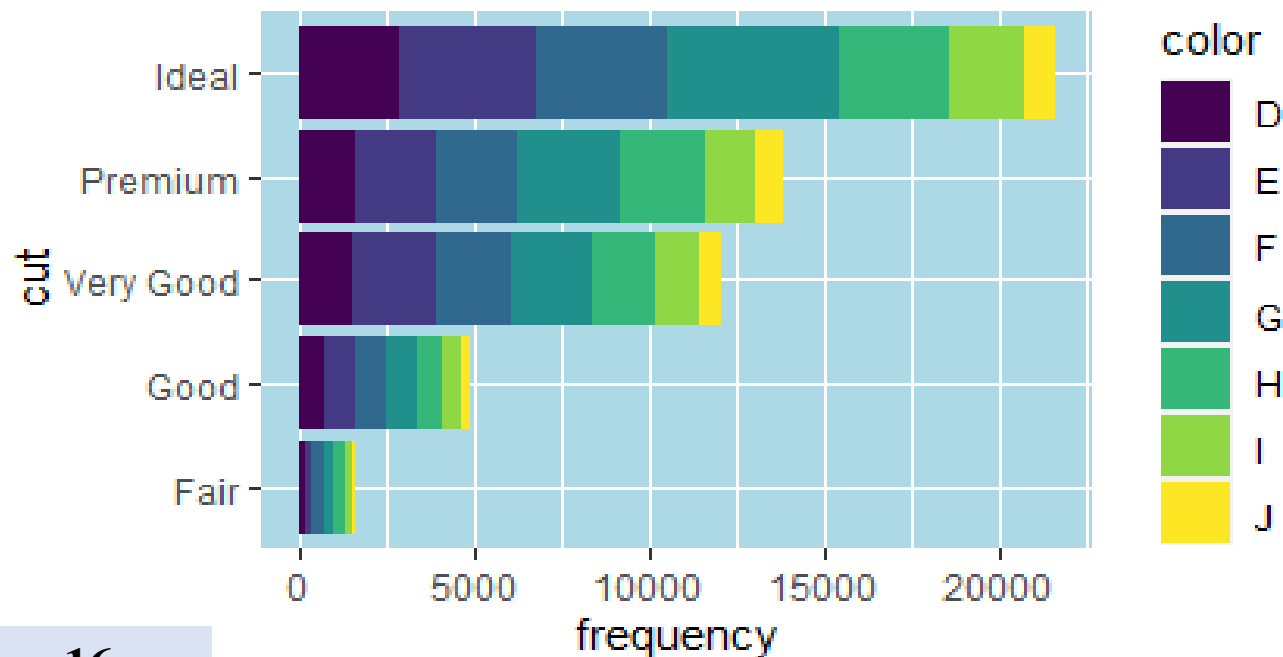


Gambar 15

## f. Diagram Batang

Posisi batang diubah menjadi horizontal dengan `parameter_stack` dan menambahkan variabel warna berlian (`color`) ke dalam plot yang sama.

```
ggplot(diamonds, aes(y = cut)) +  
  geom_bar(position = position_stack(reverse = TRUE), aes(fill=color)) +  
  labs(x="frequency", y="cut") +  
  theme(panel.background = element_rect(fill="light blue"))
```



Warna yang terdapat di setiap batang menunjukkan kategori warna berlian, mulai dari yang terbaik (label D) hingga level terburuk (label J).

Gambar 16

# Percobaan Praktikum & Tugas



# Referensi

Pathak, M. A., (2014), *Beginning Data Science with R*, Springer International Publishing, Switzerland.

Dietrich, D., Barry H., & Beibei Y., (2015), *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, John Wiley & Sons, Inc., Indianapolis, Indiana.

[www.rdocumentation.org/packages/ggplot2/versions/3.3.5](http://www.rdocumentation.org/packages/ggplot2/versions/3.3.5)

[www.rdocumentation.org/packages/ggplot2/versions/3.3.5/topics/qplot](http://www.rdocumentation.org/packages/ggplot2/versions/3.3.5/topics/qplot)

[www.rdocumentation.org/packages/ggplot2/versions/3.3.5/topics/ggplot](http://www.rdocumentation.org/packages/ggplot2/versions/3.3.5/topics/ggplot)

[www.tutorialgateway.org/r-ggplot2-density-plot/](http://www.tutorialgateway.org/r-ggplot2-density-plot/)

<https://github.com/rstudio/cheatsheets/blob/master/data-visualization-2.1.pdf>



# KMMI 2021

## Eksplorasi dan Visualisasi Data

Pertemuan 9:  
Visualisasi Menggunakan paket ggplot2