Visualisasi data menggunakan `ggplot2`

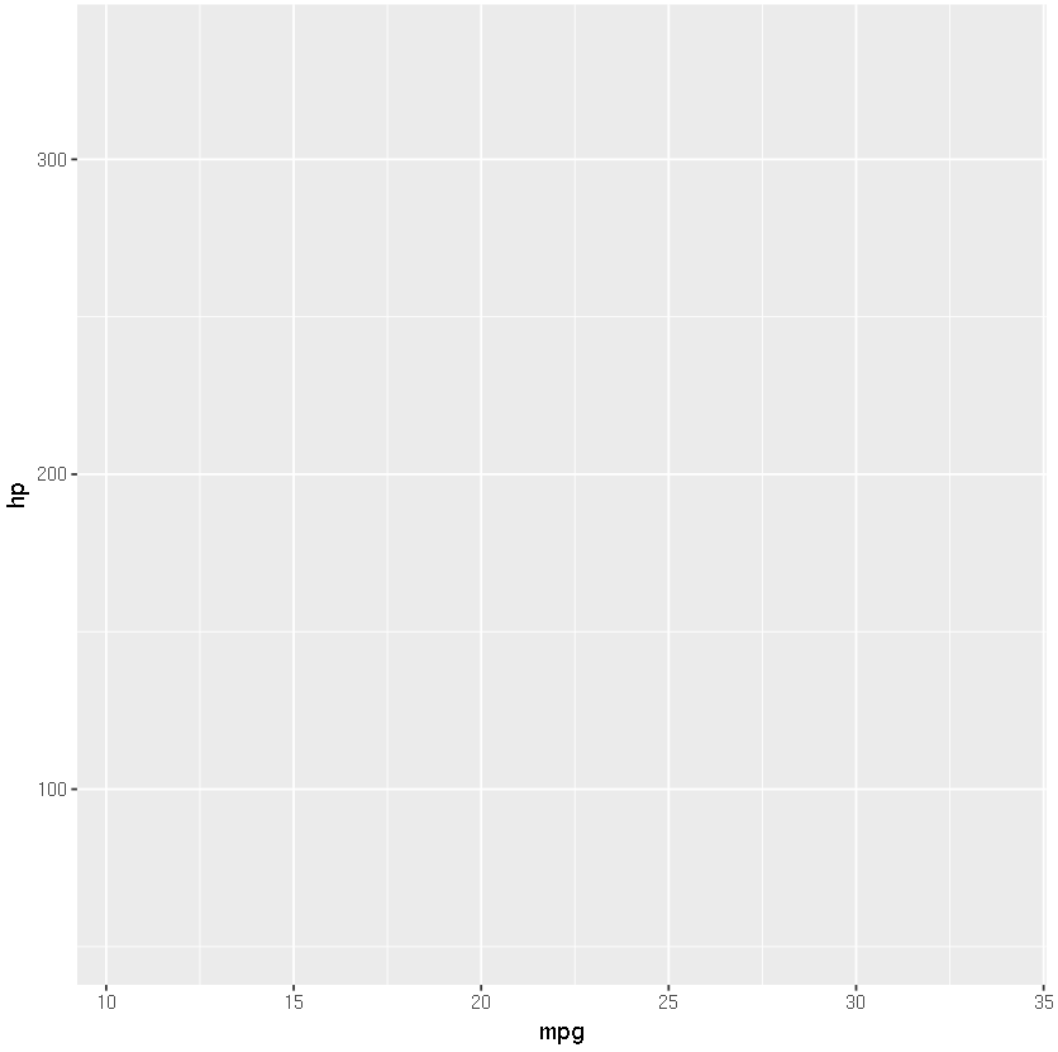
Pendahuluan

- `ggplot2` merupakan pustaka visualisasi pada bahasa pemrograman R.
- Dibangun berdasarkan konsep penambahan lapisan (*layer*) dalam visualisasi.
- Terdapat 7 lapisan: Data, Aesthetics, Geometries, Facets, Statistics, Coordinates, Themes.
- 4 lapisan terakhir tidak wajib, namun dpt digunakan untuk kostumisasi.

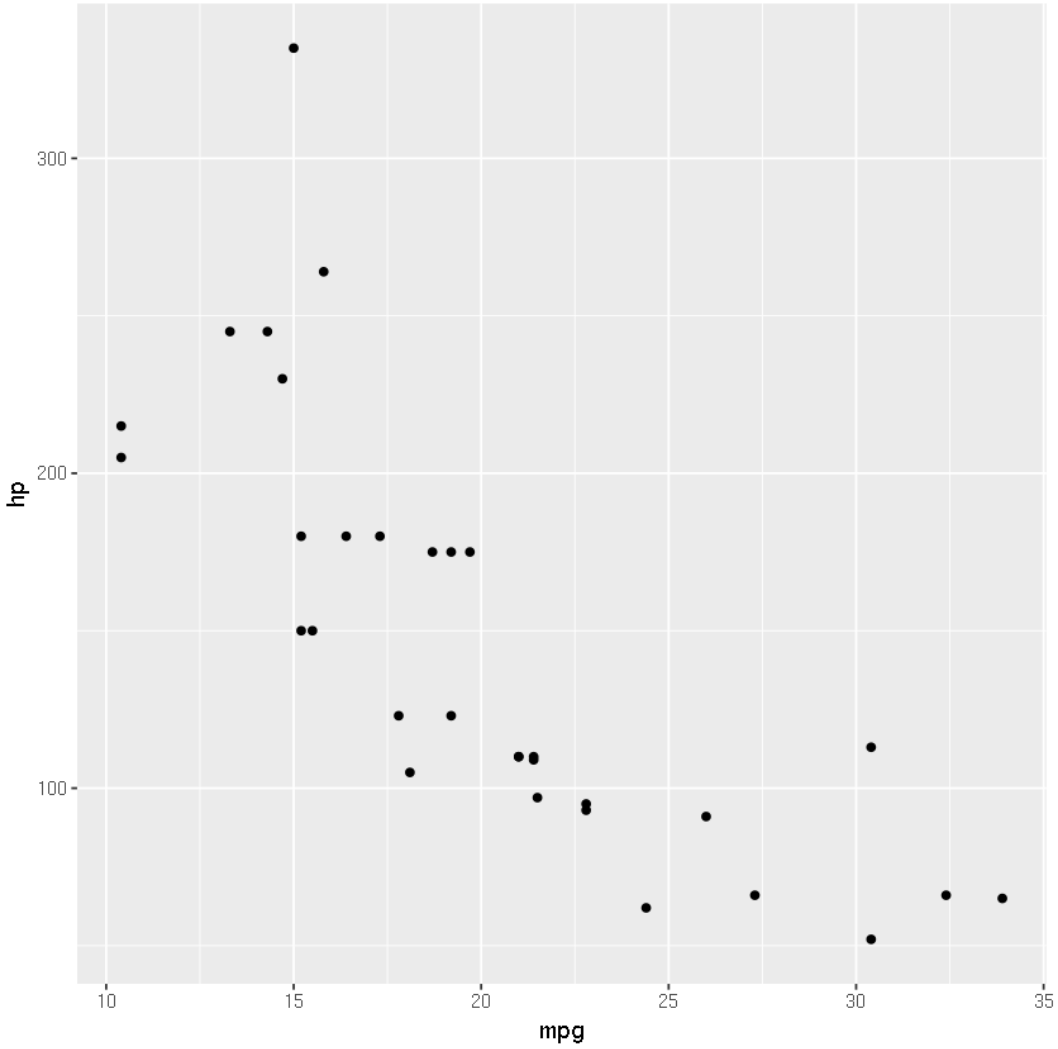
```
library(ggplot2)
```

```
Registered S3 methods overwritten by 'ggplot2':
  method         from
  [.quosures     rlang
  c.quosures     rlang
  print.quosures rlang
```

```
ggplot(data = mtcars, # 1) Lapisan 1: Data
    aes(x = mpg, y = hp)) # 2) Lapisan 2: Aesthetics
```




```
# Lapisan 3): Geometrics
pl <- ggplot(data = mtcars,
             aes(x = mpg, y = hp))
pl + geom_point()
```




```
# 4) Layer 4: Facets
## Membuat kita dapat memplot banyak grafik di dalam satu kanvas

# 5) Layer 5: Statistics

# 6) Layer 6: Coordinates
## Membatasi limit sumbu-x dan y

# 7) Lapisan 7: Theme
## Menambahkan tema ke dalam suatu plot
```

Histogram

```
library(ggplot2movies)
```

```
head(movies)
```

title	year	length	budget	rating	votes	r1	r2	r3	r4	...	r9	r10	mpaa	Action	Ar
$	1971	121	NA	6.4	348	4.5	4.5	4.5	4.5	...	4.5	4.5		0	0
$1000 a Touchdown	1939	71	NA	6.0	20	0.0	14.5	4.5	24.5	...	4.5	14.5		0	0
$21 a Day Once a Month	1941	7	NA	8.2	5	0.0	0.0	0.0	0.0	...	24.5	24.5		0	1
$40,000	1996	70	NA	8.2	6	14.5	0.0	0.0	0.0	...	34.5	45.5		0	0
$50,000 Climax Show, The	1975	71	NA	3.4	17	24.5	4.5	0.0	14.5	...	0.0	24.5		0	0
$pent	2000	91	NA	4.3	45	4.5	4.5	4.5	14.5	...	14.5	14.5		0	0

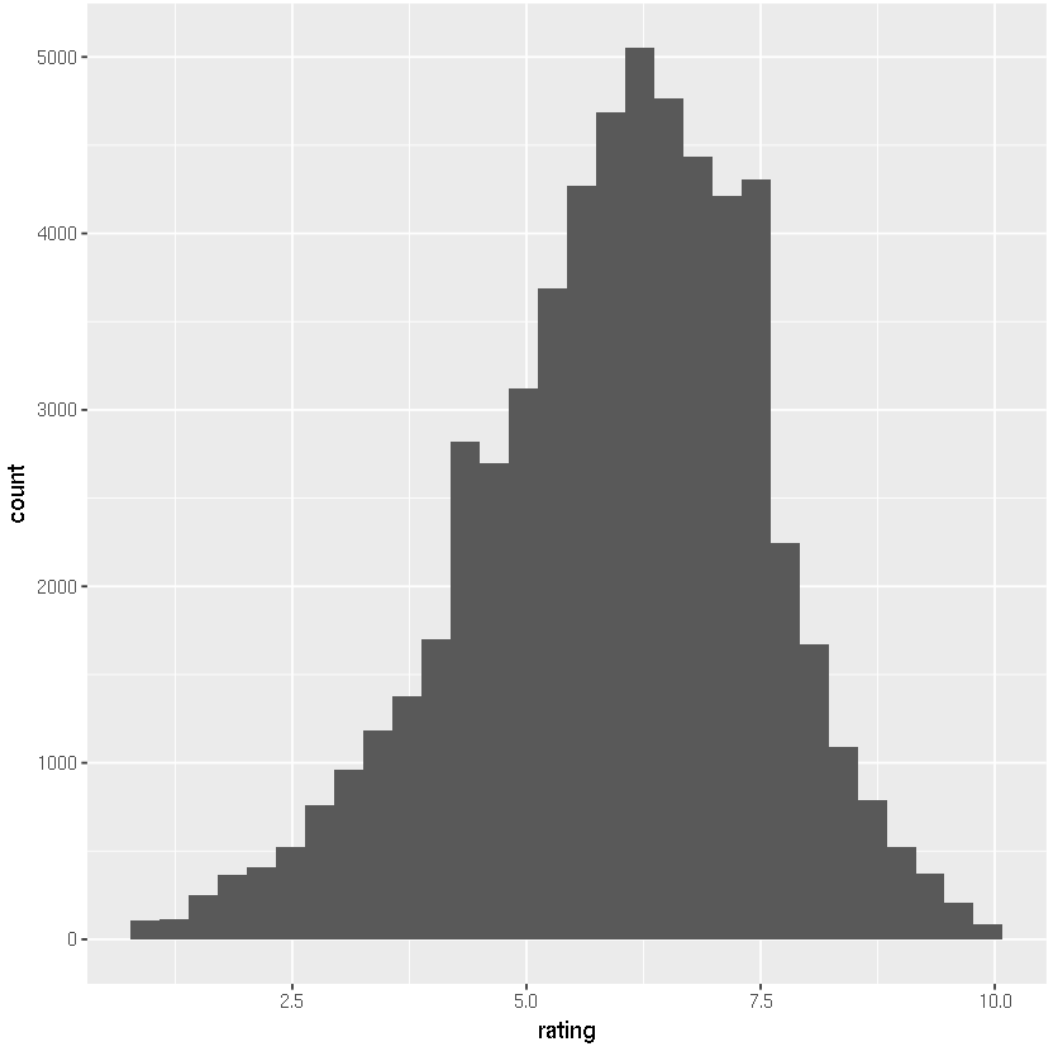
```
colnames(movies)
```

1. 'title'
2. 'year'
3. 'length'
4. 'budget'
5. 'rating'
6. 'votes'
7. 'r1'
8. 'r2'
9. 'r3'
10. 'r4'
11. 'r5'
12. 'r6'
13. 'r7'
14. 'r8'
15. 'r9'
16. 'r10'
17. 'mpaa'
18. 'Action'
19. 'Animation'
20. 'Comedy'
21. 'Drama'
22. 'Documentary'
23. 'Romance'
24. 'Short'

Cheatsheet : https://rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf

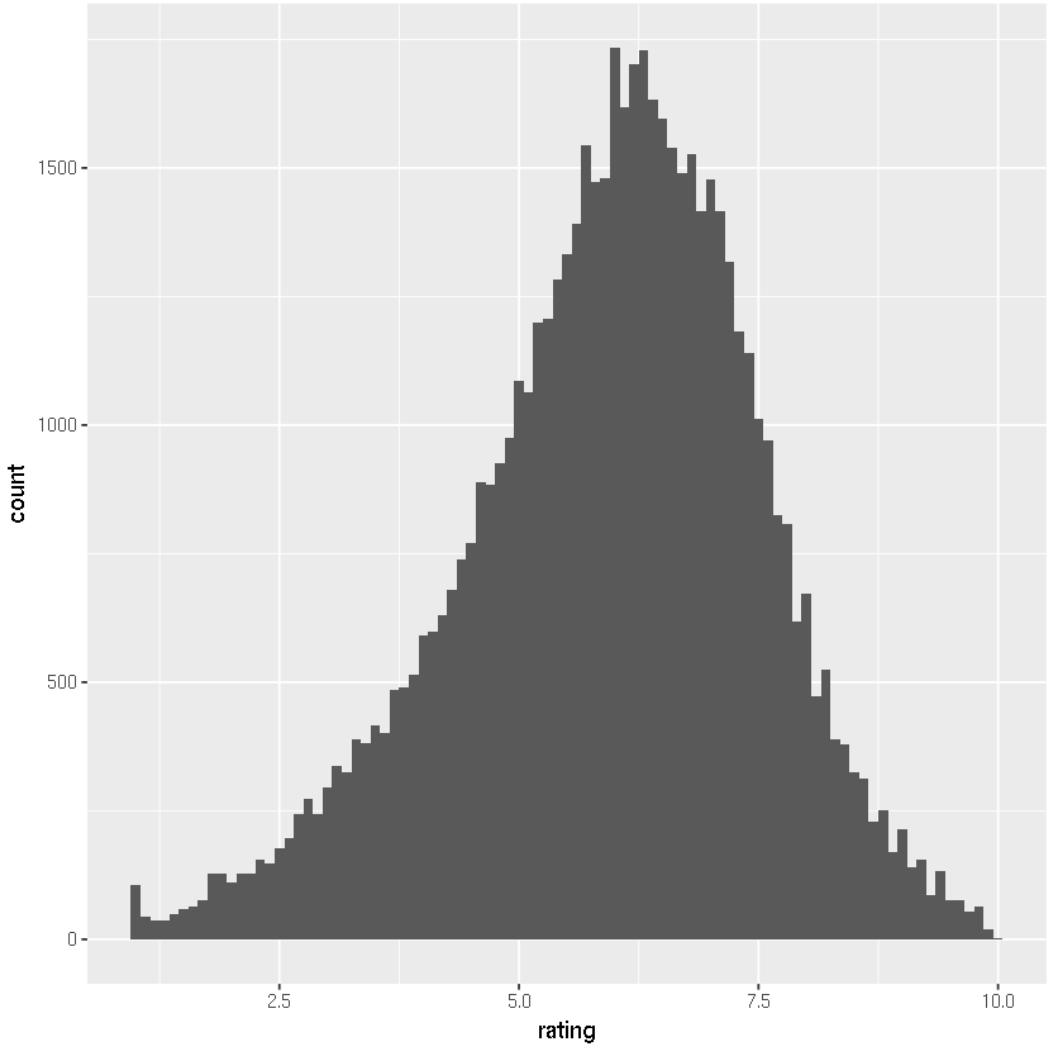
```
pl <- ggplot(data = movies,
             aes(x = rating))
pl + geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

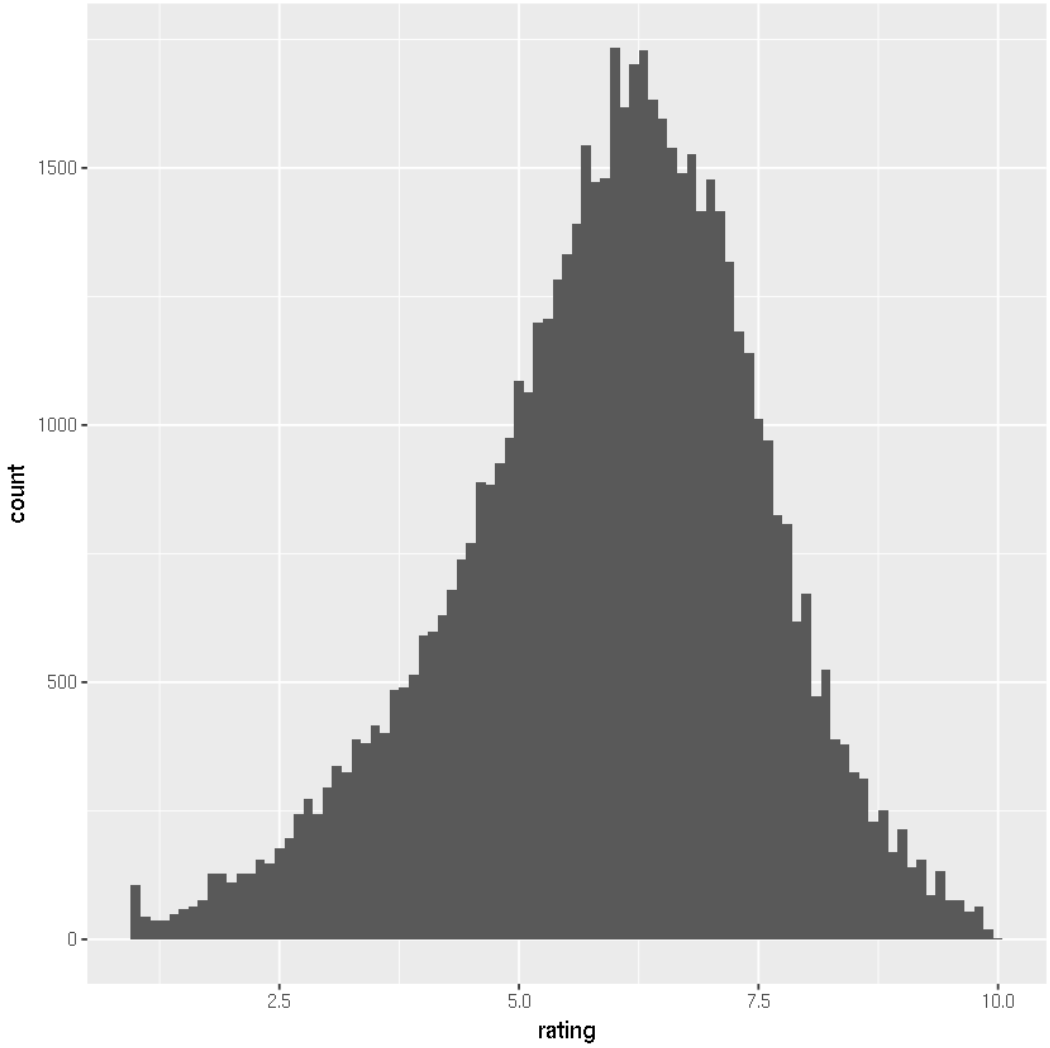



```
# Untuk mengetahui lebih lanjut, perintahkan:
# help("geom_histogram")
```

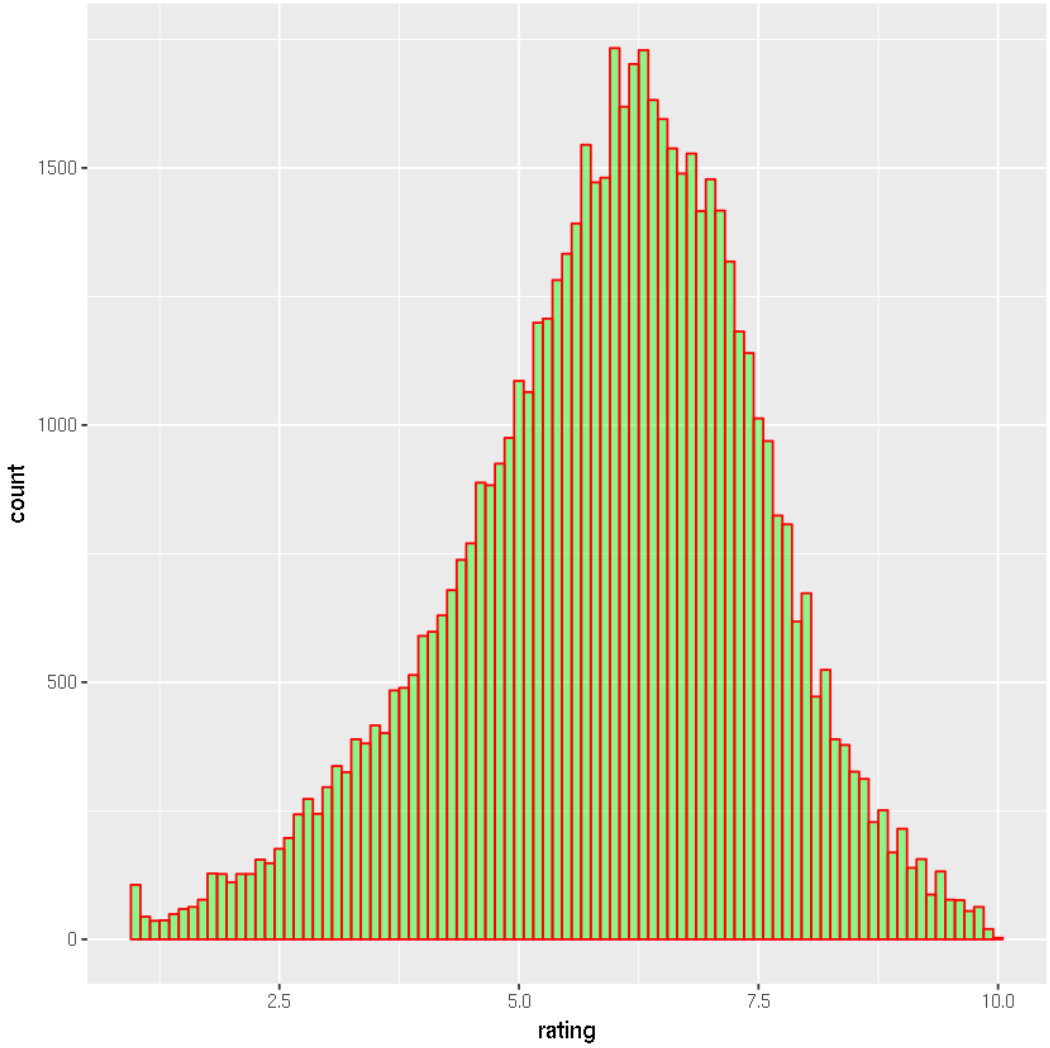
```
pl + geom_histogram(binwidth=0.1) # binwidth = 0.1
```




```
pl + geom_histogram(binwidth=0.1, bins=100) # defaultnya bins = 30
```

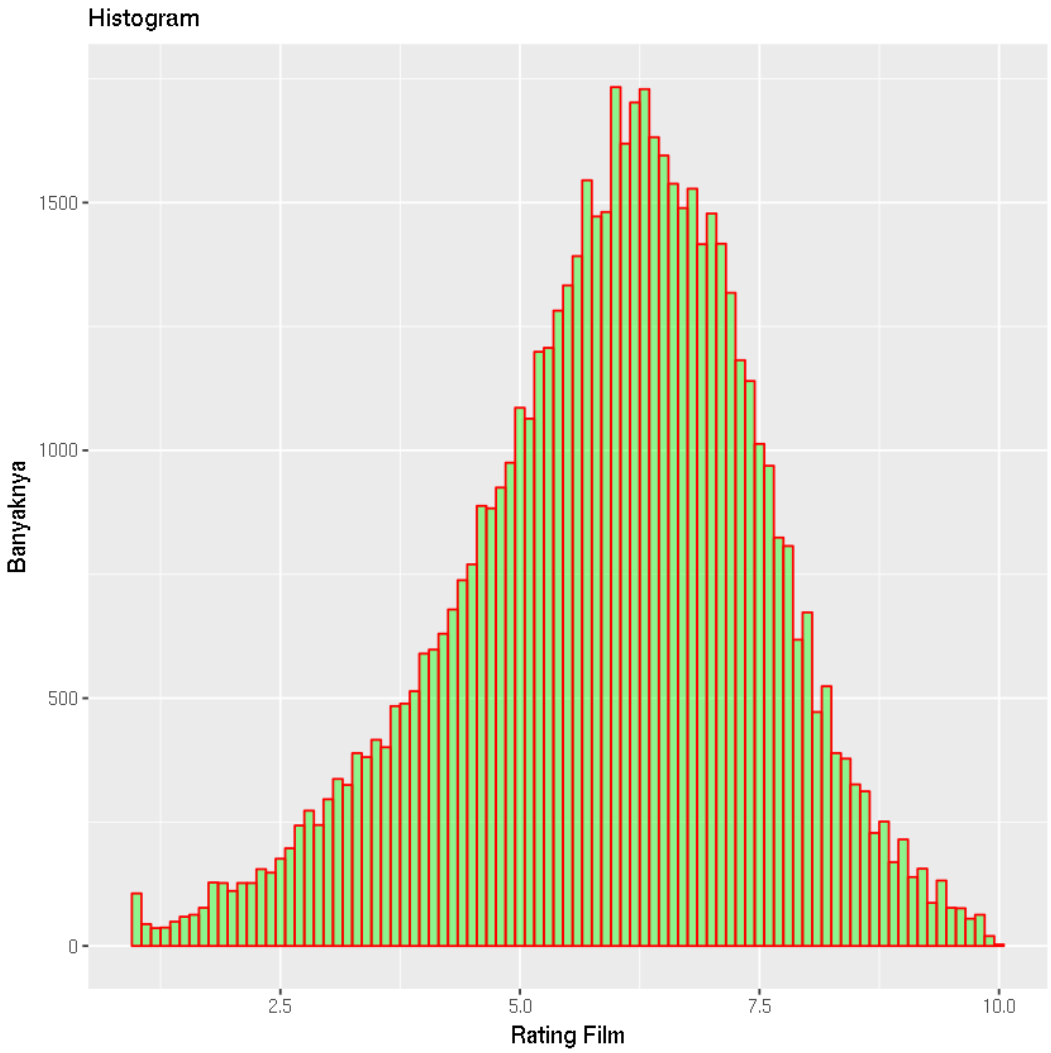



```
pl + geom_histogram(binwidth=0.1, bins=100,
                    color = 'red', fill='green',
                   alpha = 0.4)
```



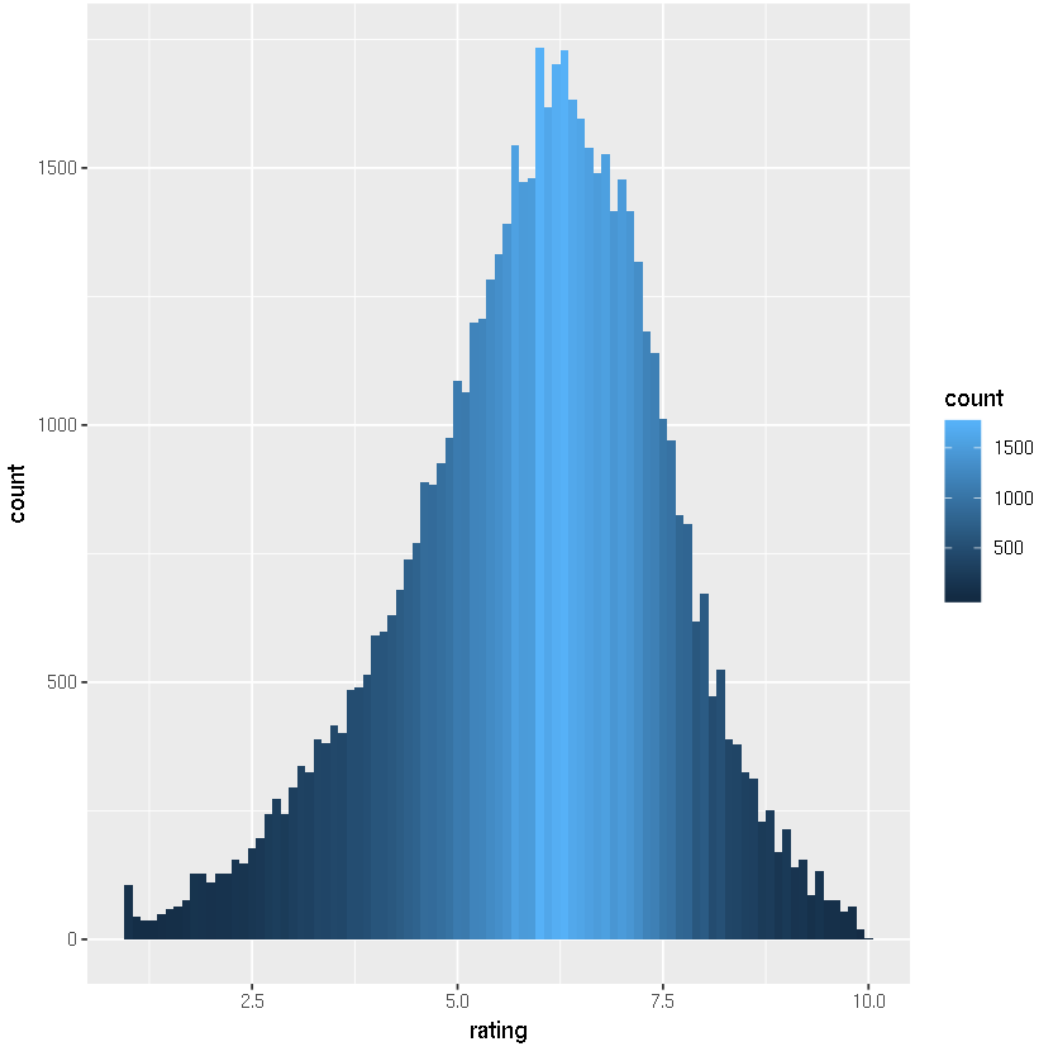
Menambahkan label

```
pl2 <- pl + geom_histogram(binwidth=0.1, bins=100,
                    color = 'red', fill='green',
                   alpha = 0.4)
pl2 + xlab('Rating Film') + ylab('Banyaknya') + ggtitle("Histogram")
```



Teknik aesthetics lanjutan

```
pl + geom_histogram(binwidth=0.1, aes(fill= ..count..))
```

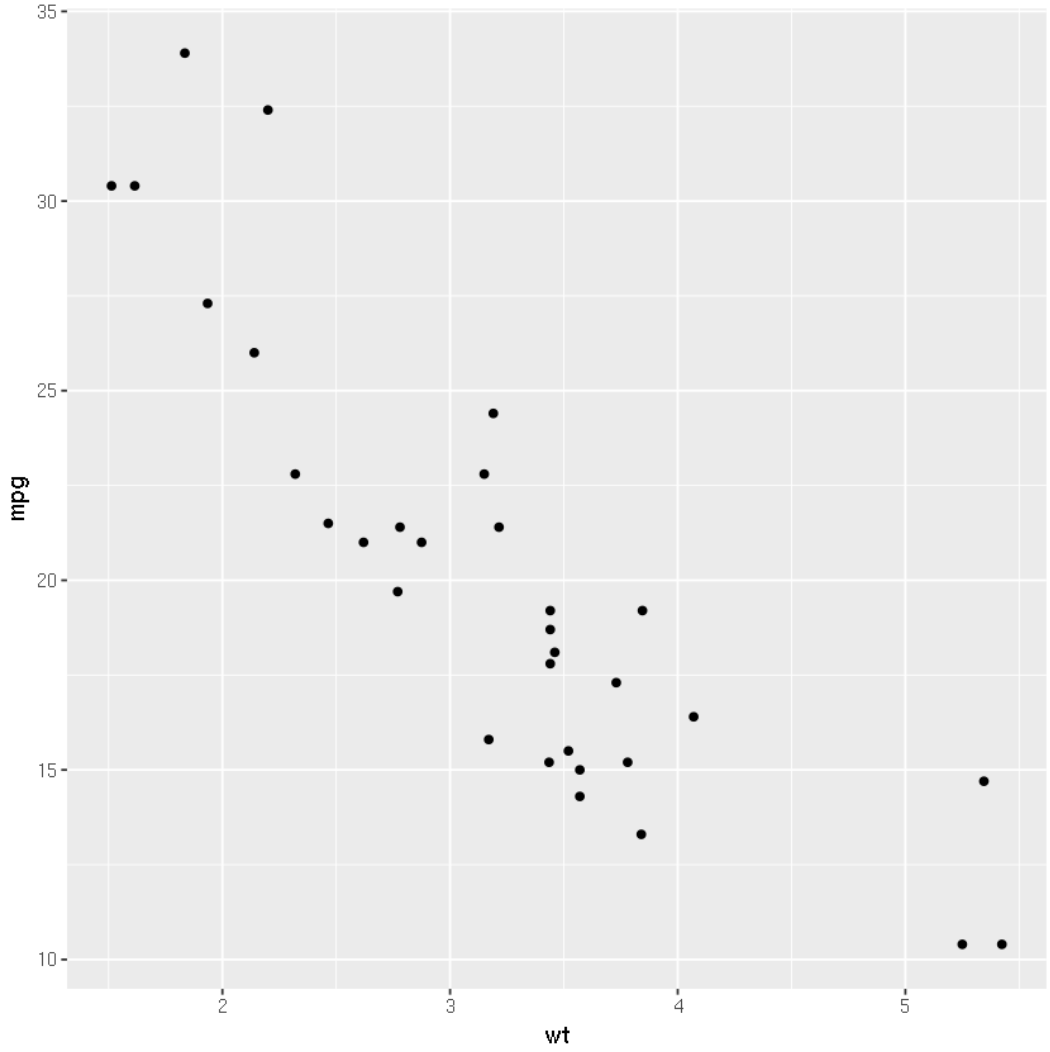


Scatterplot

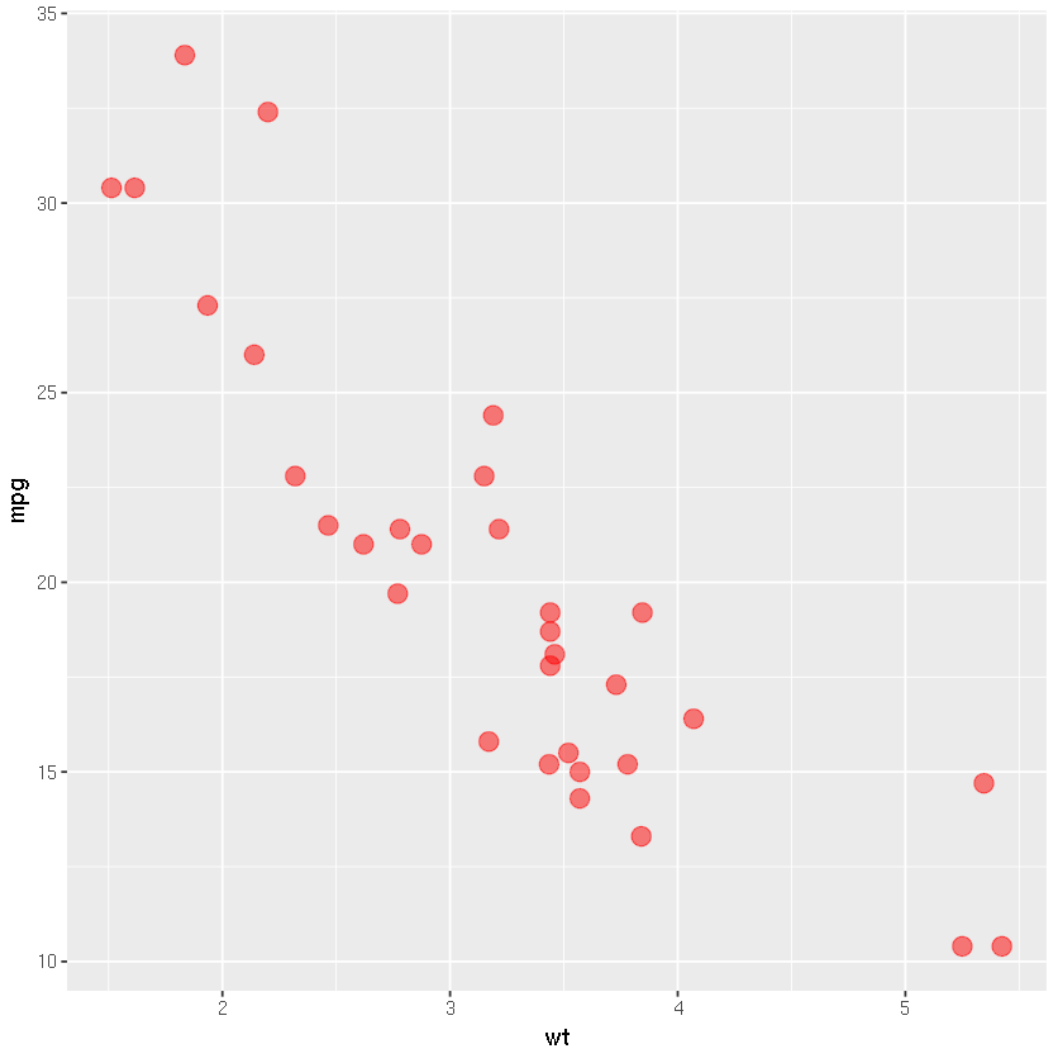
```
df <- mtcars
head(df)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

```
pl <- ggplot(data = df, aes(x=wt, y=mpg))
pl + geom_point()
```



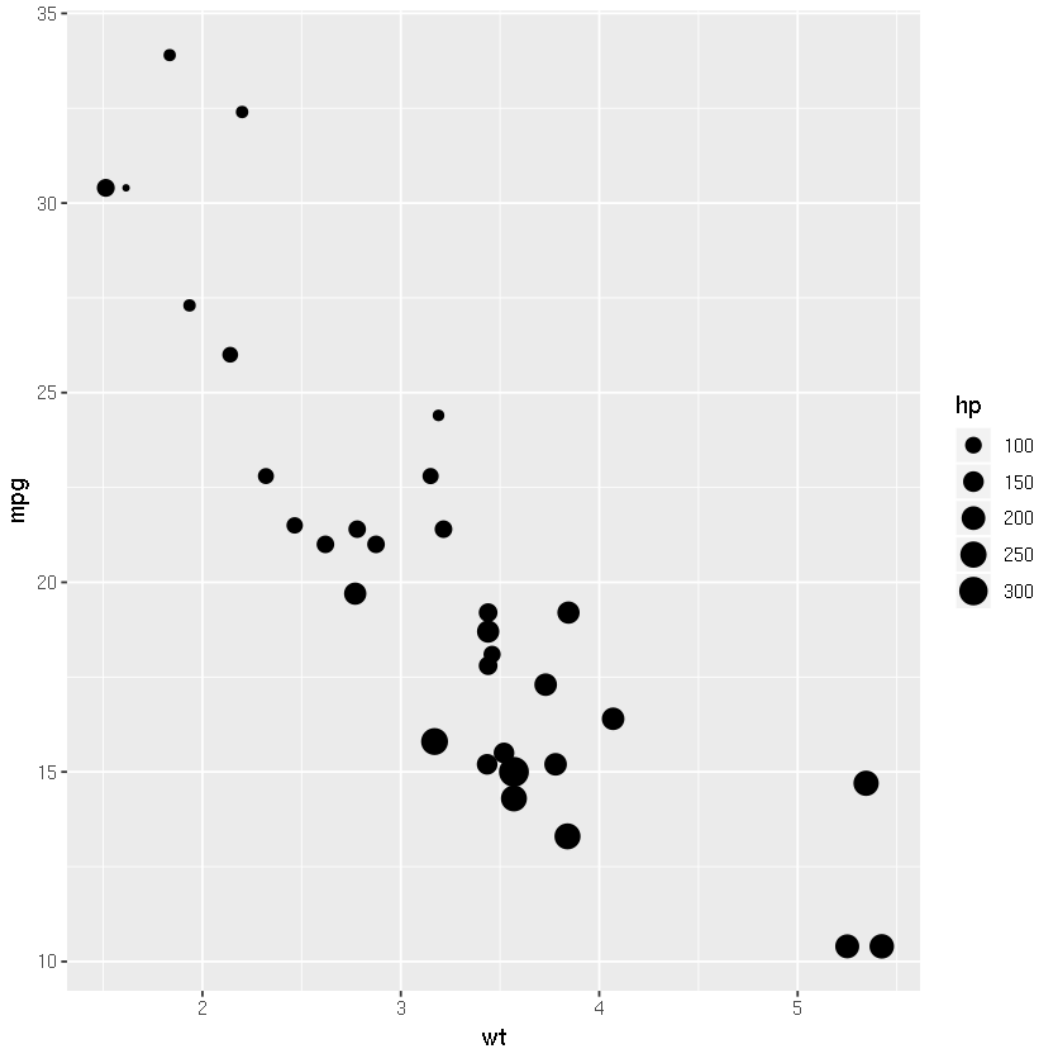

```
pl + geom_point(color = 'red', size=4, alpha = 0.5)
```



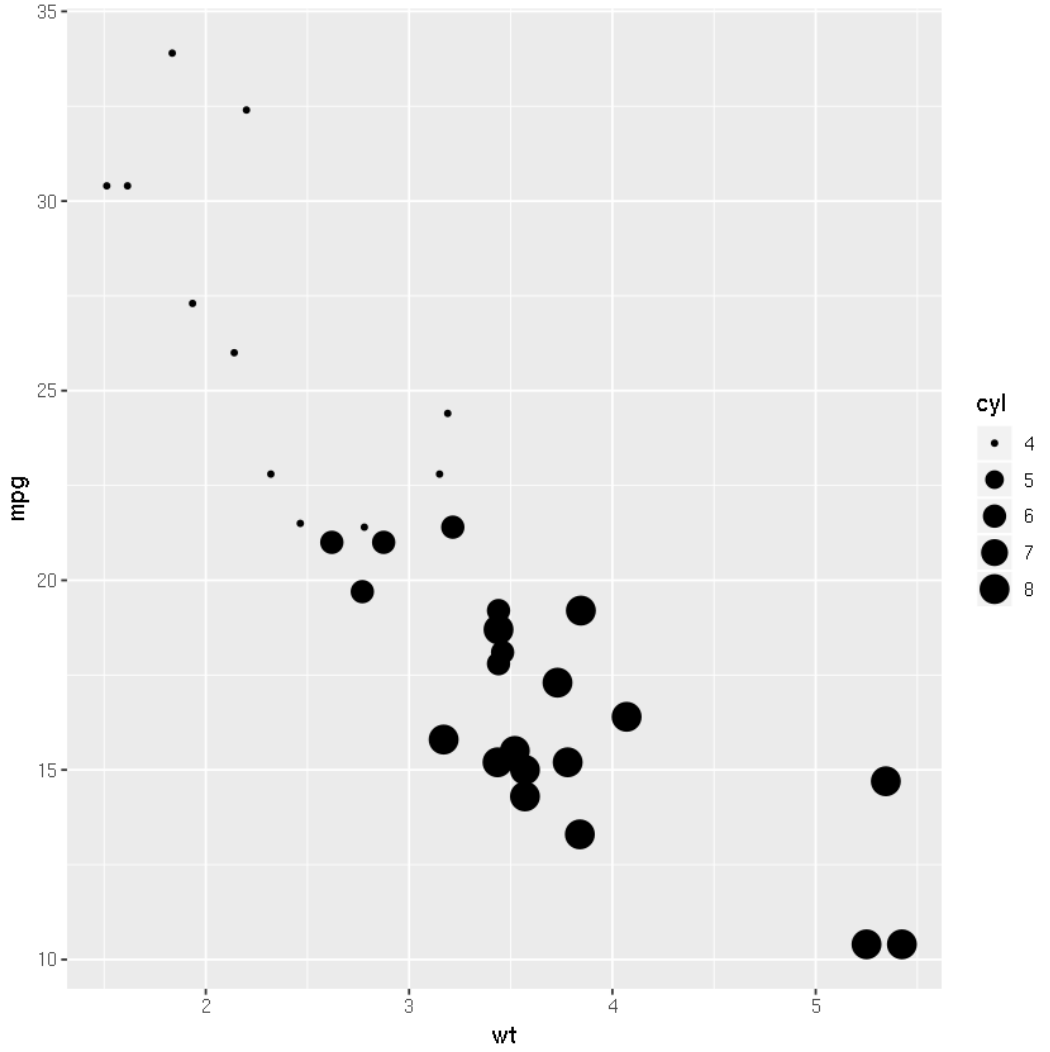

```
# Untuk mengetahui secara lebih lanjut, perintahkan:
# help("geom_point")
```

Menambahkan pemetaan aesthetics

```
pl + geom_point(aes(size = hp)) # ukuran titik berdasarkan besaran hp
```




```
pl + geom_point(aes(size = cyl))  # ukuran titik berdasarkan besaran cyl
```

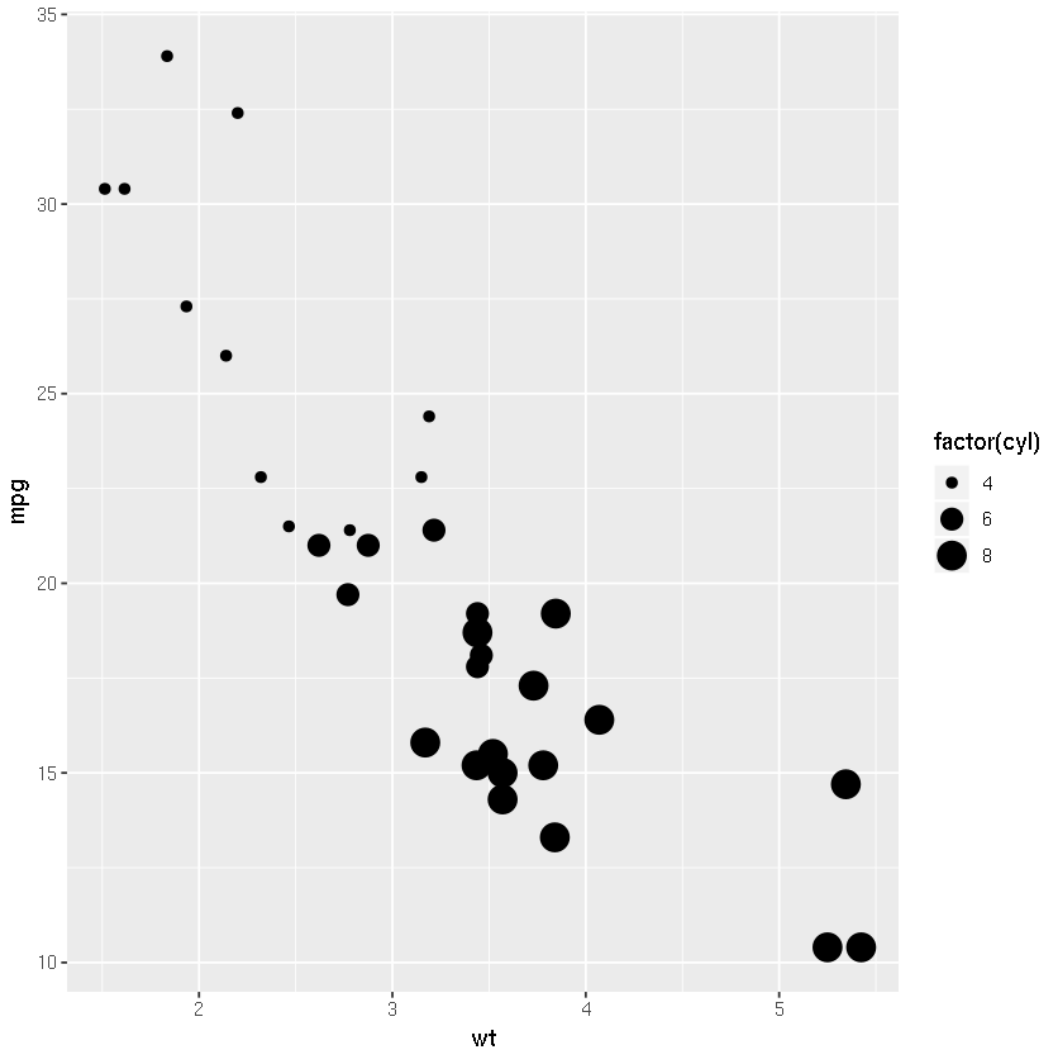



```
df$cyl # bersifat kategorikal, maka kita harus menggunakan fungsi factor()
```

1. 6
2. 6
3. 4
4. 6
5. 8
6. 6
7. 8
8. 4
9. 4
10. 6
11. 6
12. 8
13. 8
14. 8
15. 8
16. 8
17. 8
18. 4
19. 4
20. 4
21. 4
22. 8
23. 8
24. 8
25. 8
26. 4
27. 4
28. 4
29. 8
30. 6
31. 8
32. 4

```
pl + geom_point(aes( size = factor(cyl)))
```

```
Warning message:
“Using size for a discrete variable is not advised.”
```

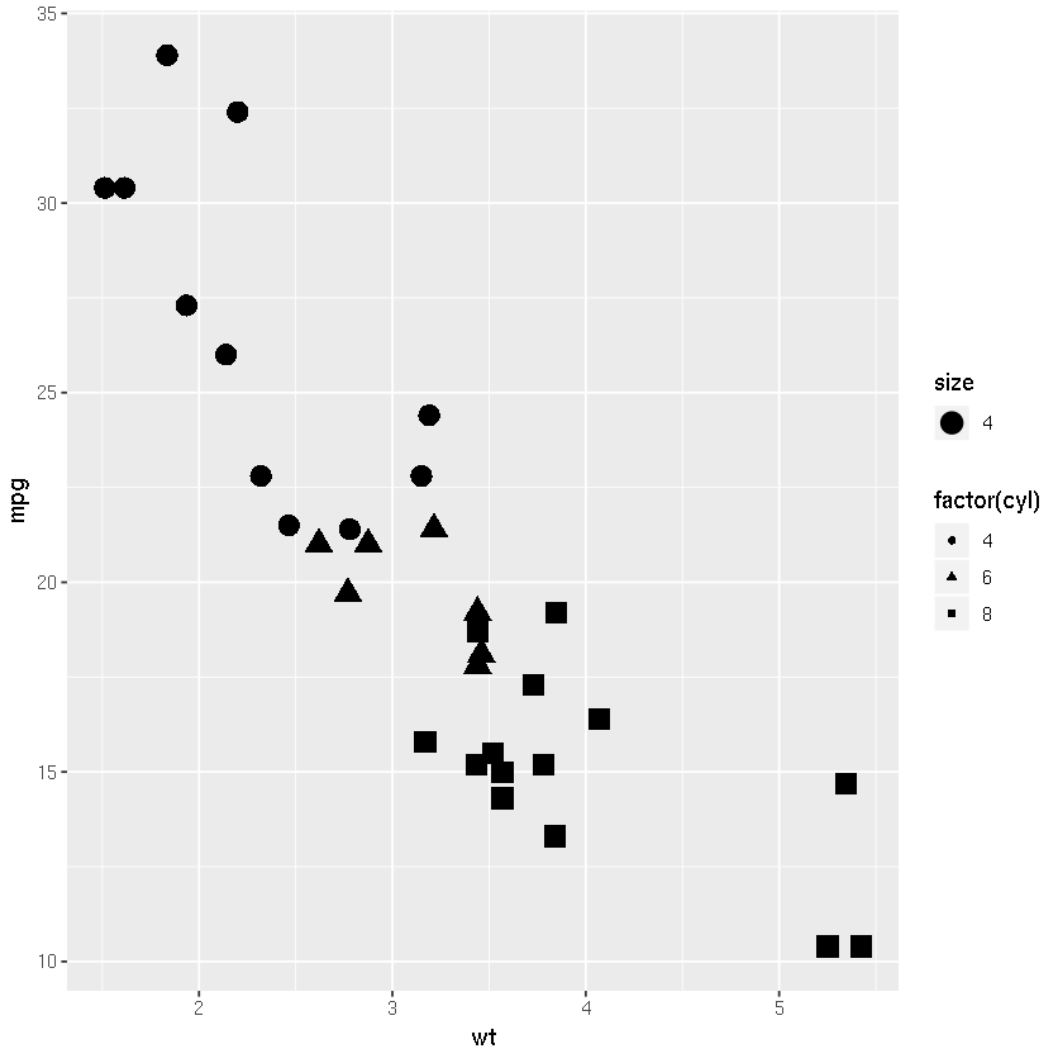


Terdapat pesan:

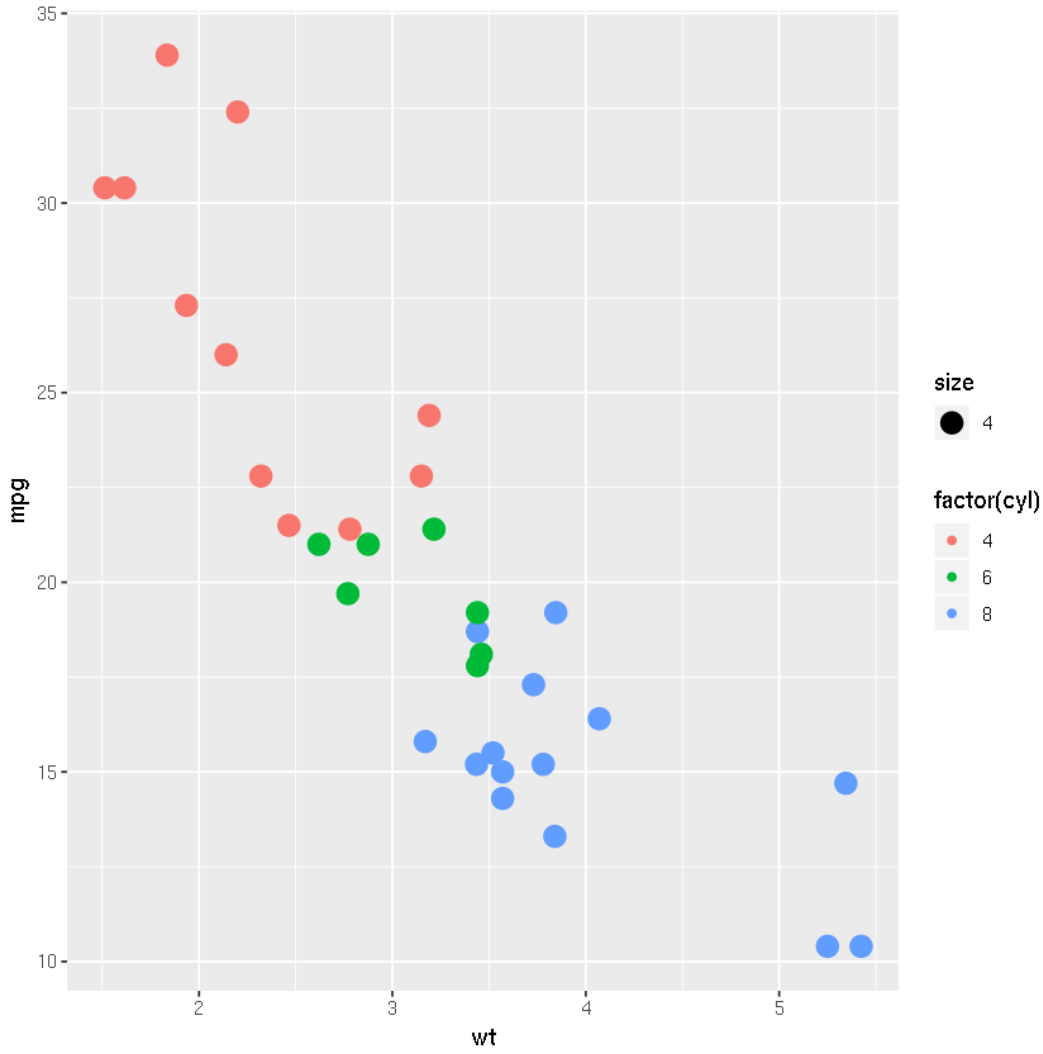
Warning message:
“Using size for a discrete variable is not advised.”

Maka, lebih baik tidak usah digunakan

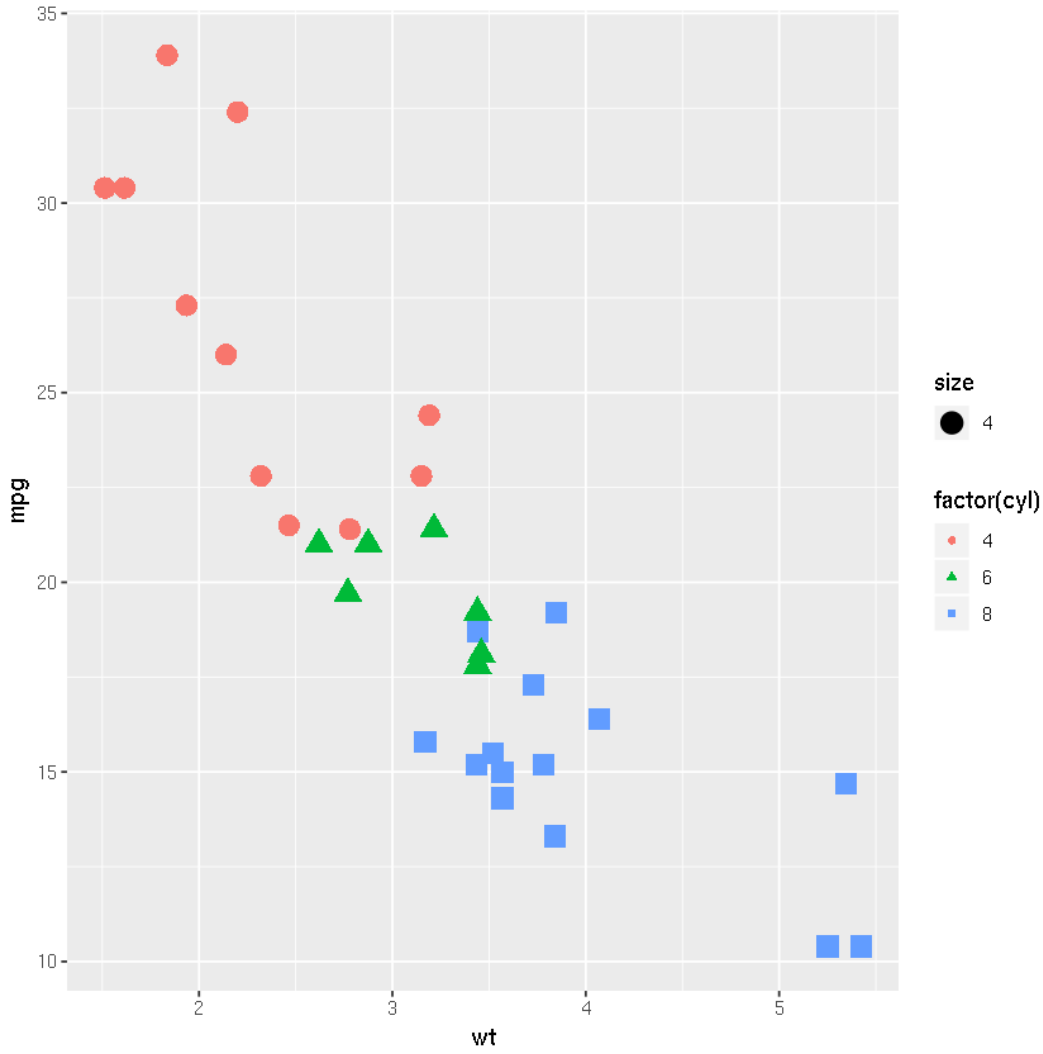
```
pl + geom_point(aes(shape=factor(cyl), size=4)) # pakai ini lebih baik
```



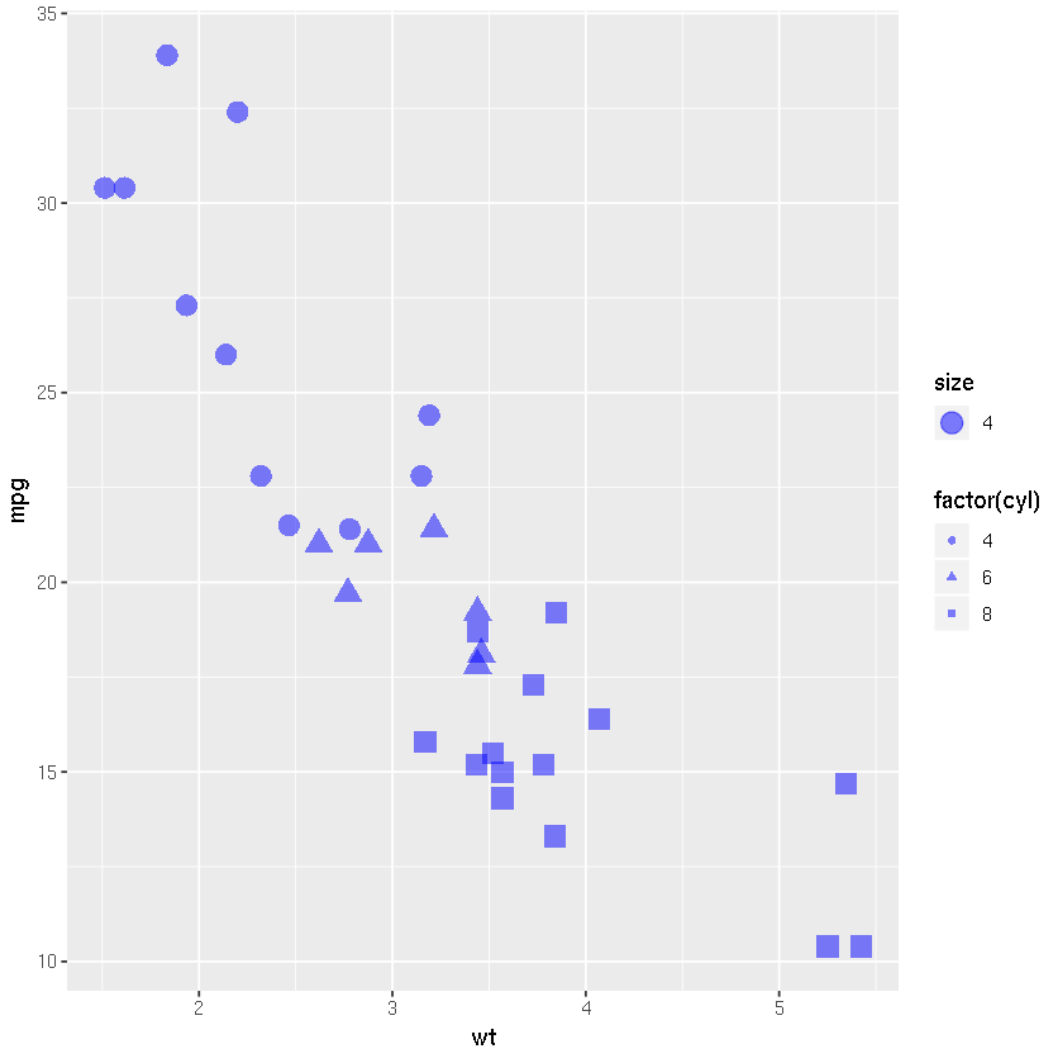

```
# Kita juga dapat membedakan dengan warna
pl + geom_point(aes(color=factor(cyl), size=4))
```




```
# Sintaks lengkap
pl + geom_point(aes(color=factor(cyl), shape = factor(cyl), size=4))
```

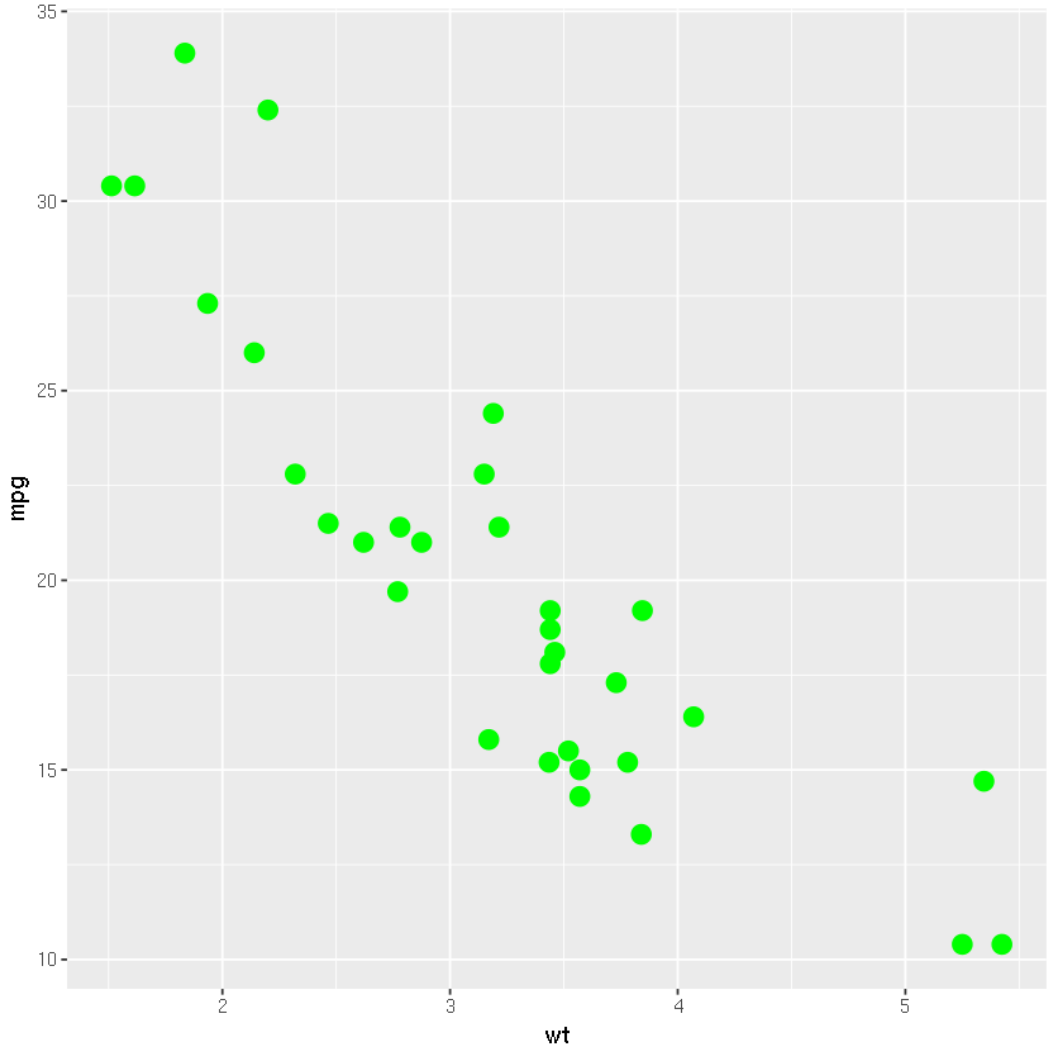



```
pl + geom_point(aes(shape = factor(cyl), size=4), color='blue', alpha=0.5)
# menambahkan warna di luar aes
```



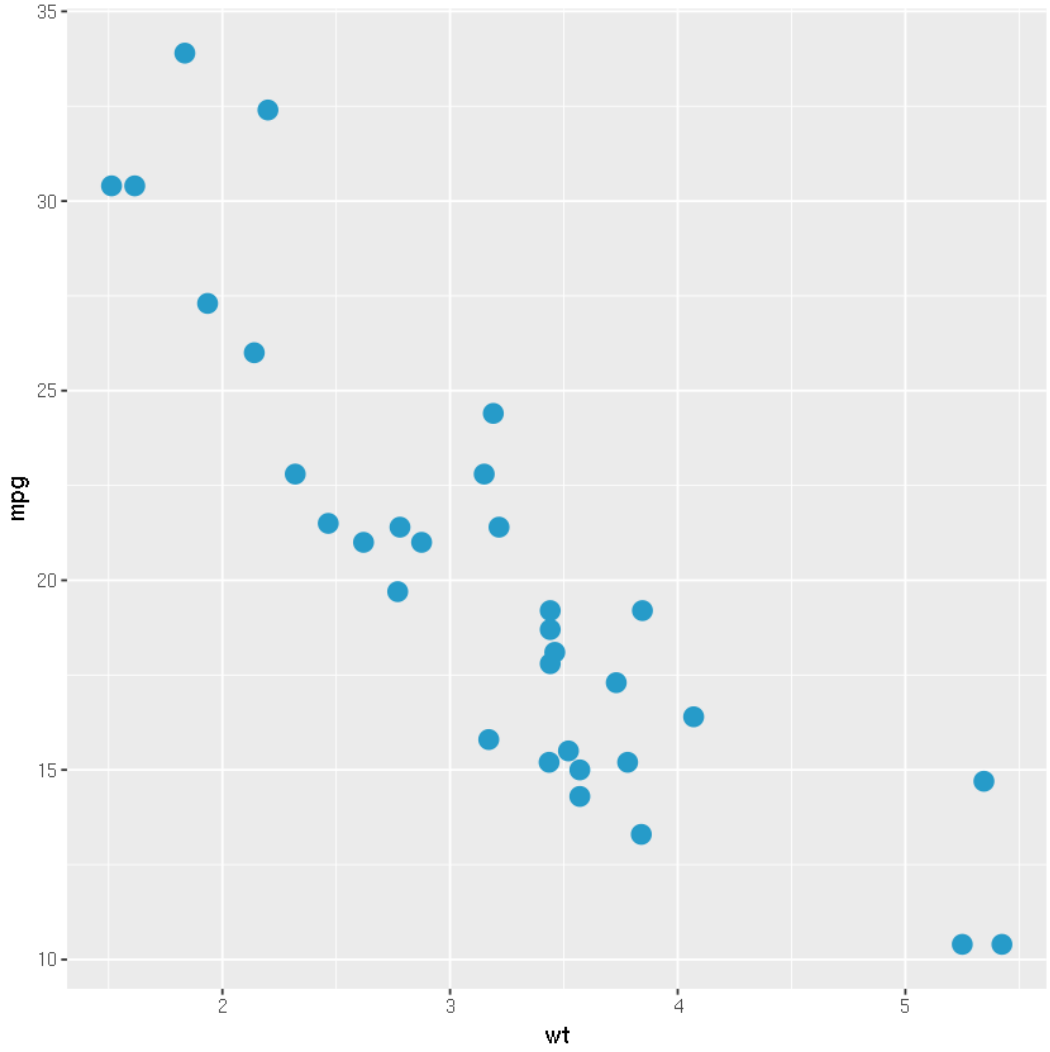
Hex color coding

```
pl + geom_point(size=4, color='green')
```

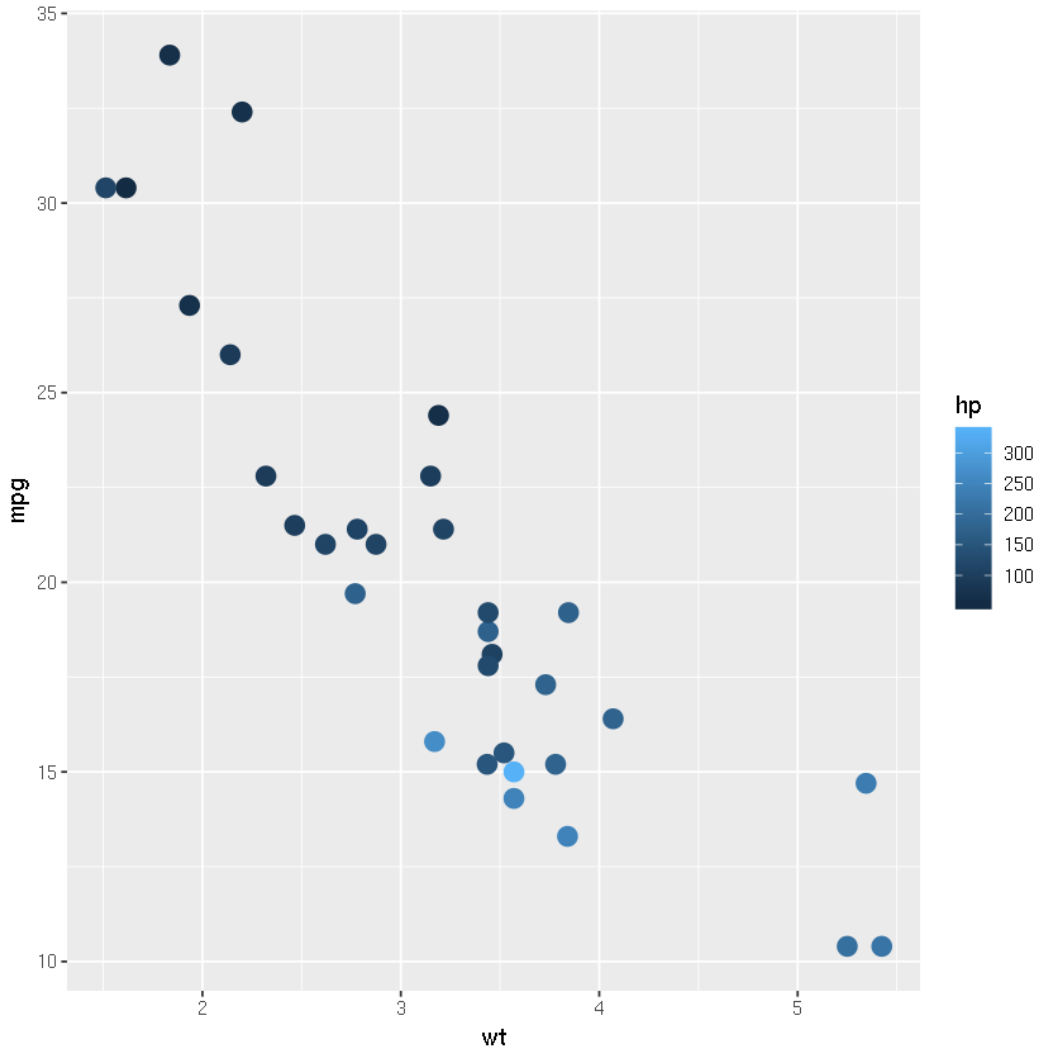


Cari di mesin pencari: hex color code

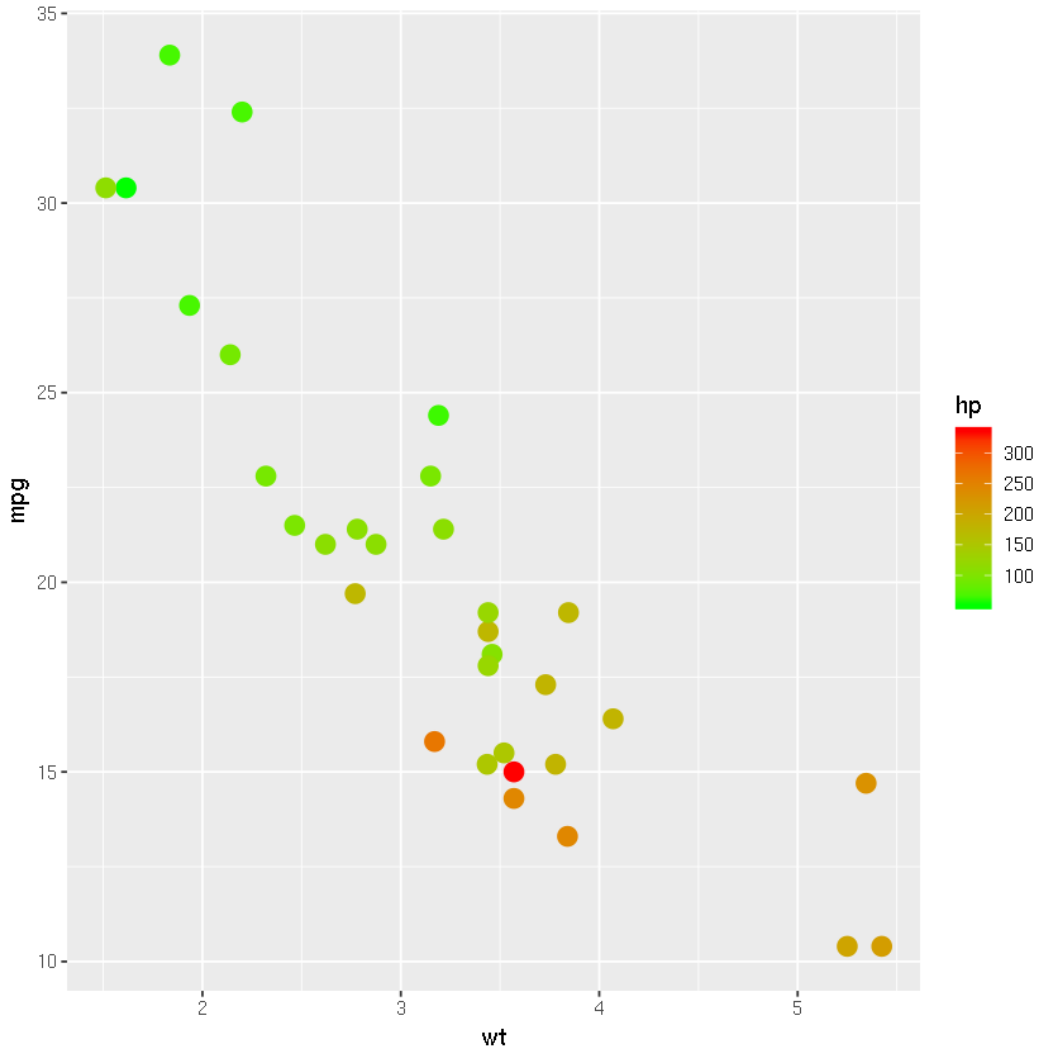
```
pl + geom_point(size=4, color='#269BC9')
```




```
p <- ggplot(df, aes(x=wt, y=mpg))
pl2 <- p + geom_point(aes(color=hp), size=4)
pl2
```

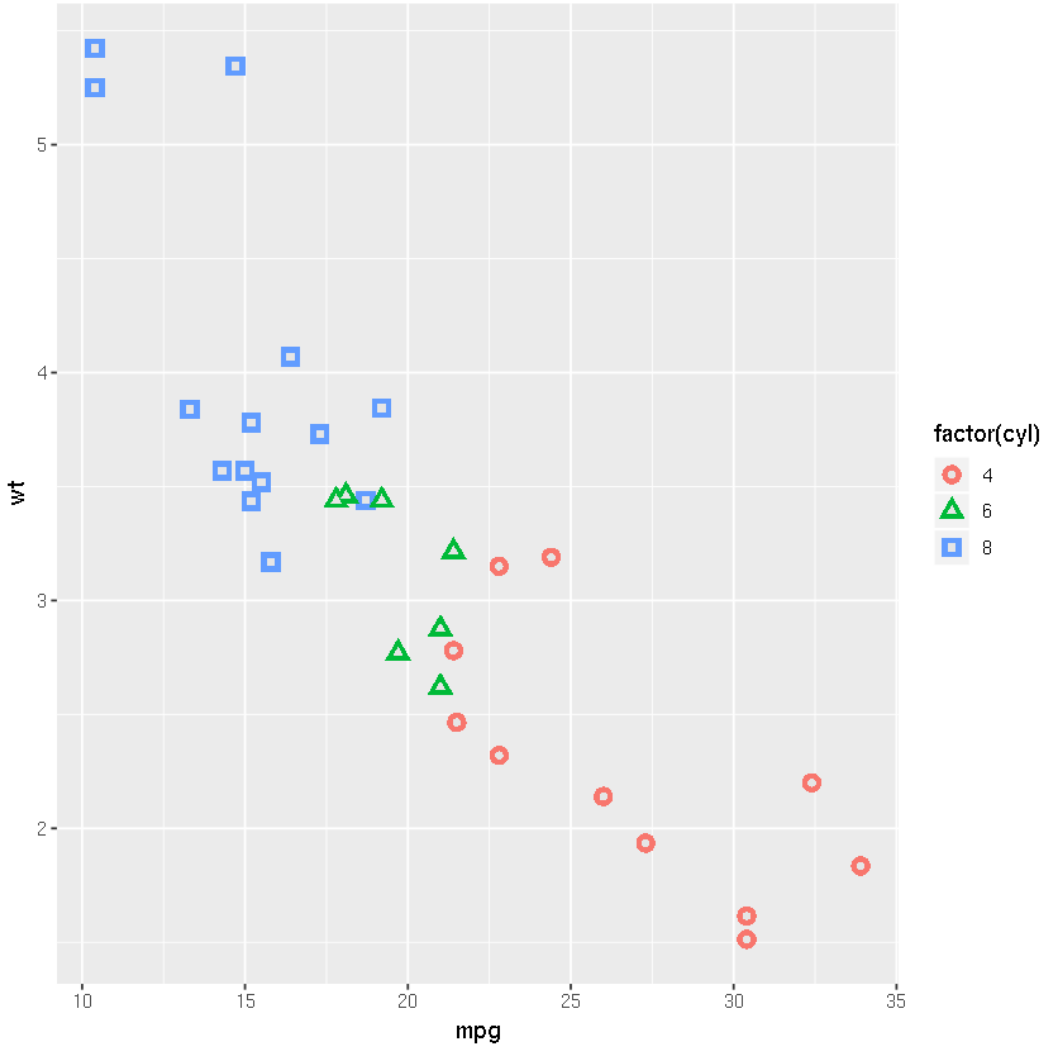



```
pl2 + scale_color_gradient(low='green', high='red')
```




```
# help("geom_point")
```

```
b <- ggplot(mtcars, aes(mpg, wt, shape = factor(cyl)))
b + geom_point(aes(colour = factor(cyl)), size = 4) +
  geom_point(colour = "grey90", size = 1.5)
b + geom_point(colour = "black", size = 4.5) +
  geom_point(colour = "pink", size = 4) +
  geom_point(aes(shape = factor(cyl)))
```


wt
5
4
3
2
10
15
20
25
30
35
mpg
factor(cyl)
4
6
8

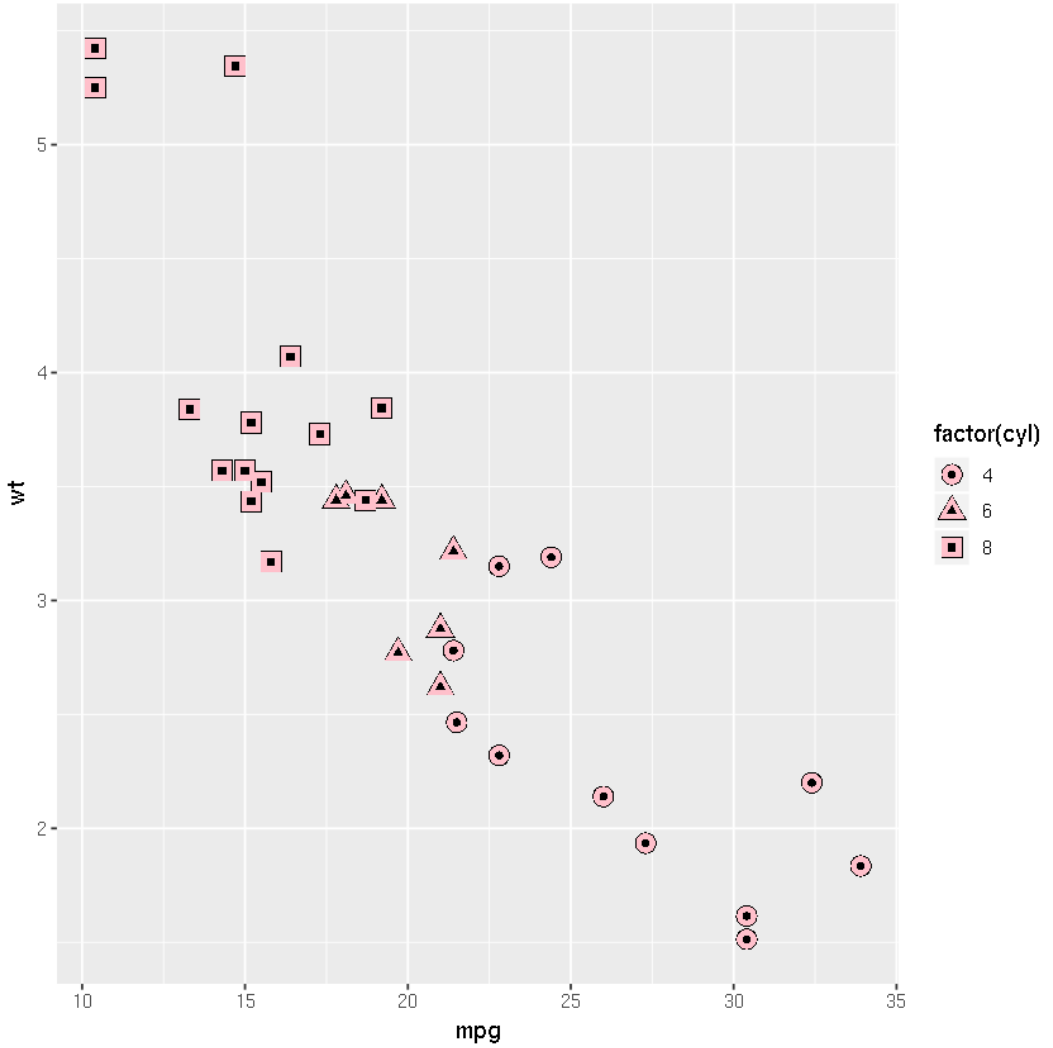


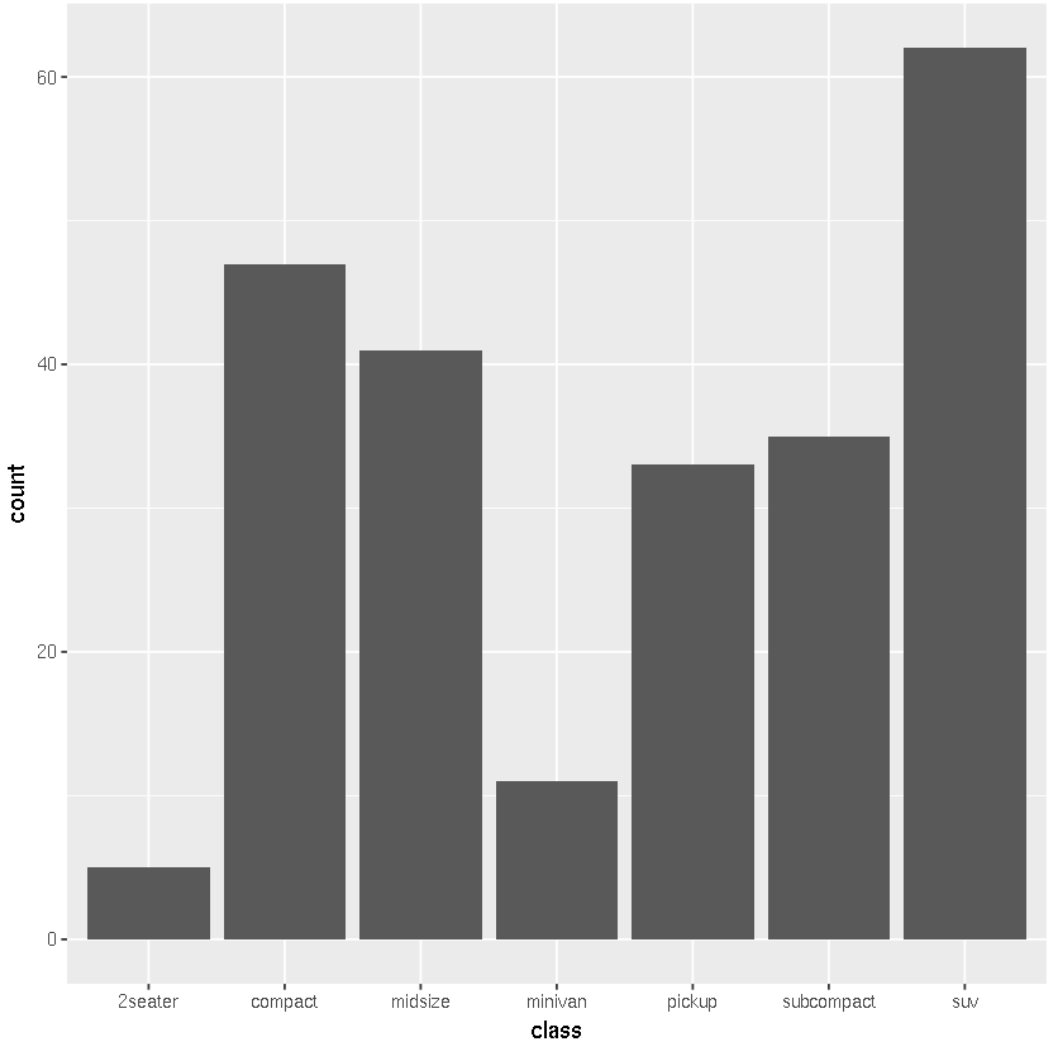
Diagram batang

Umum digunakan untuk menangani data kategorikal

```
df <- mpg
head(mpg)
```

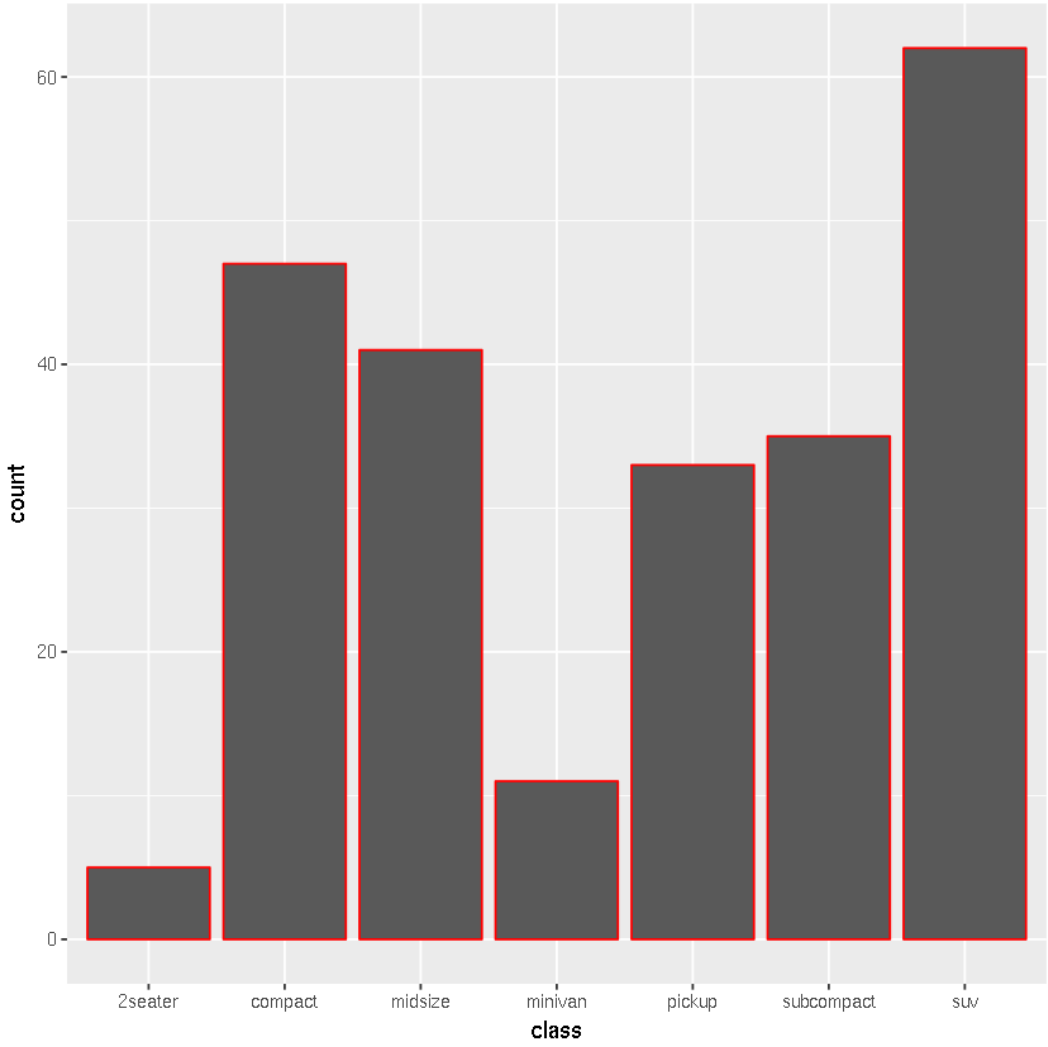
manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact

```
pl <- ggplot(df, aes(x=class)) # class : data kategorikal
pl + geom_bar()
```

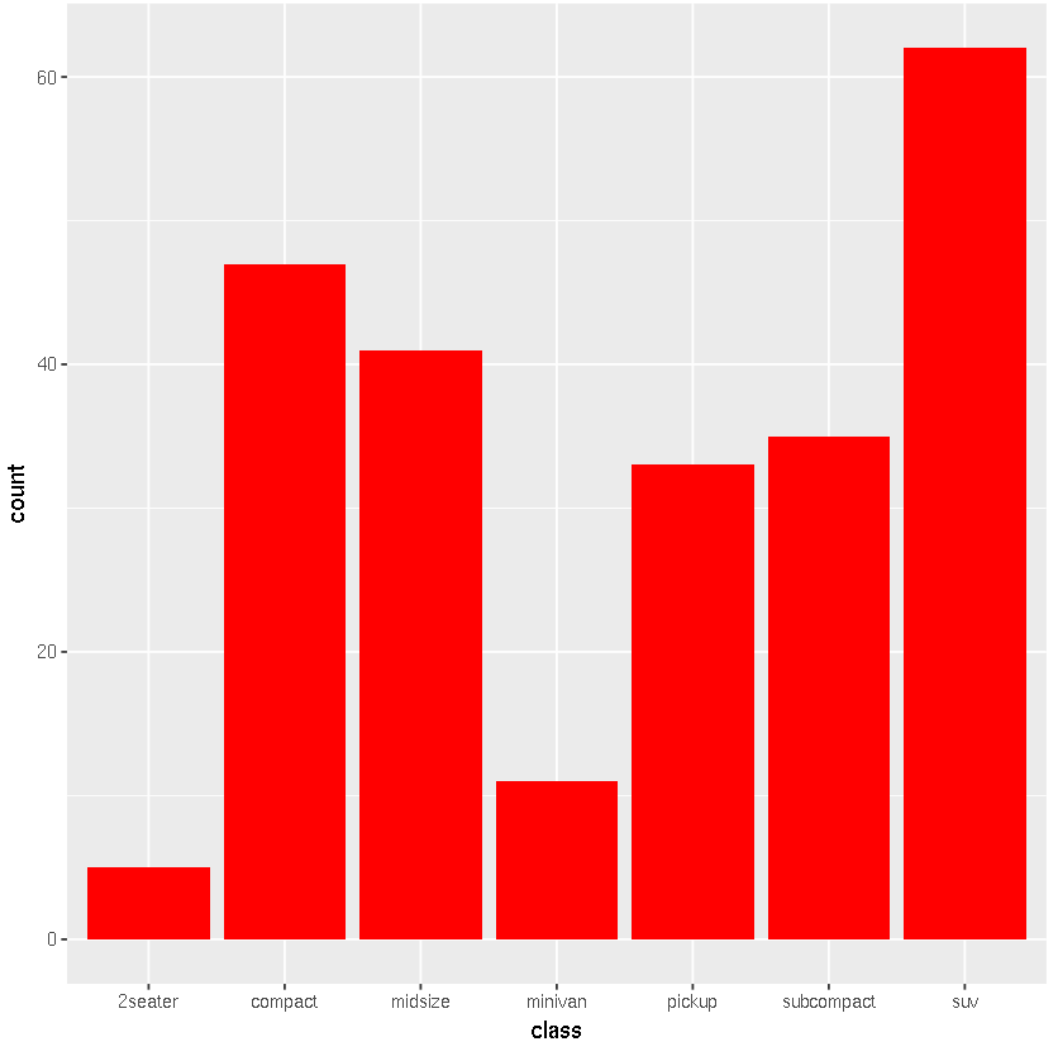



```
# Untuk mengetahui secara lebih lanjut, perintahkan:
# help("geom_bar")
```

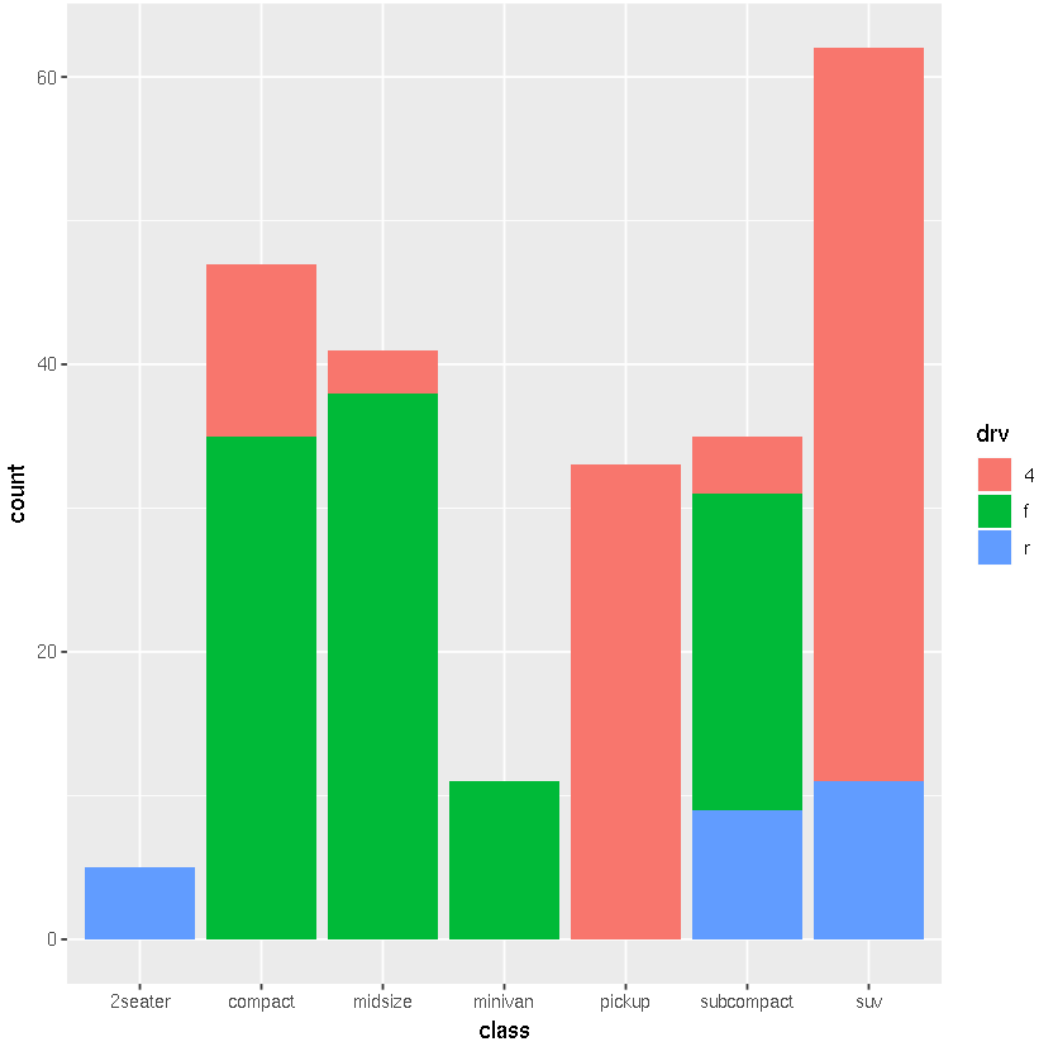
```
pl + geom_bar(color='red')
```



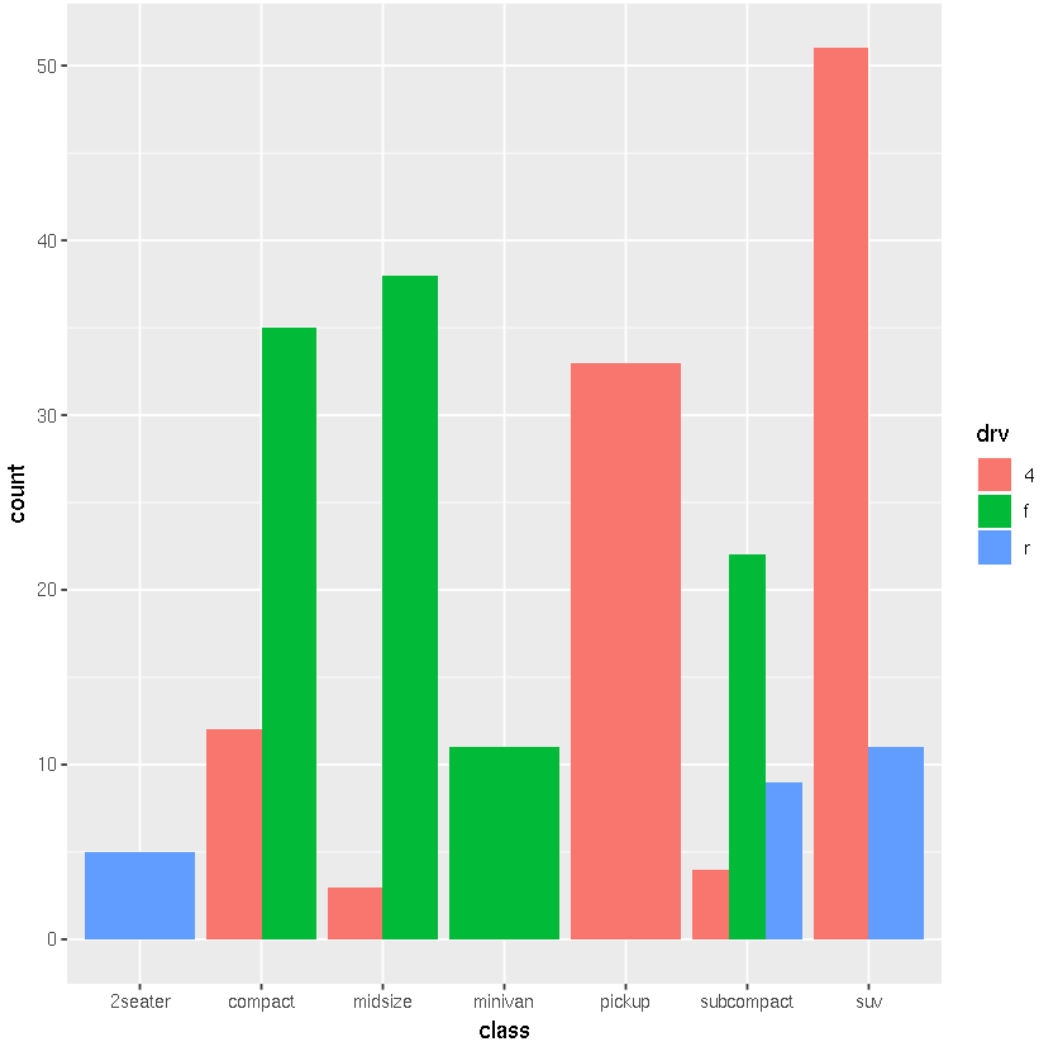

```
pl + geom_bar(fill='red')
```



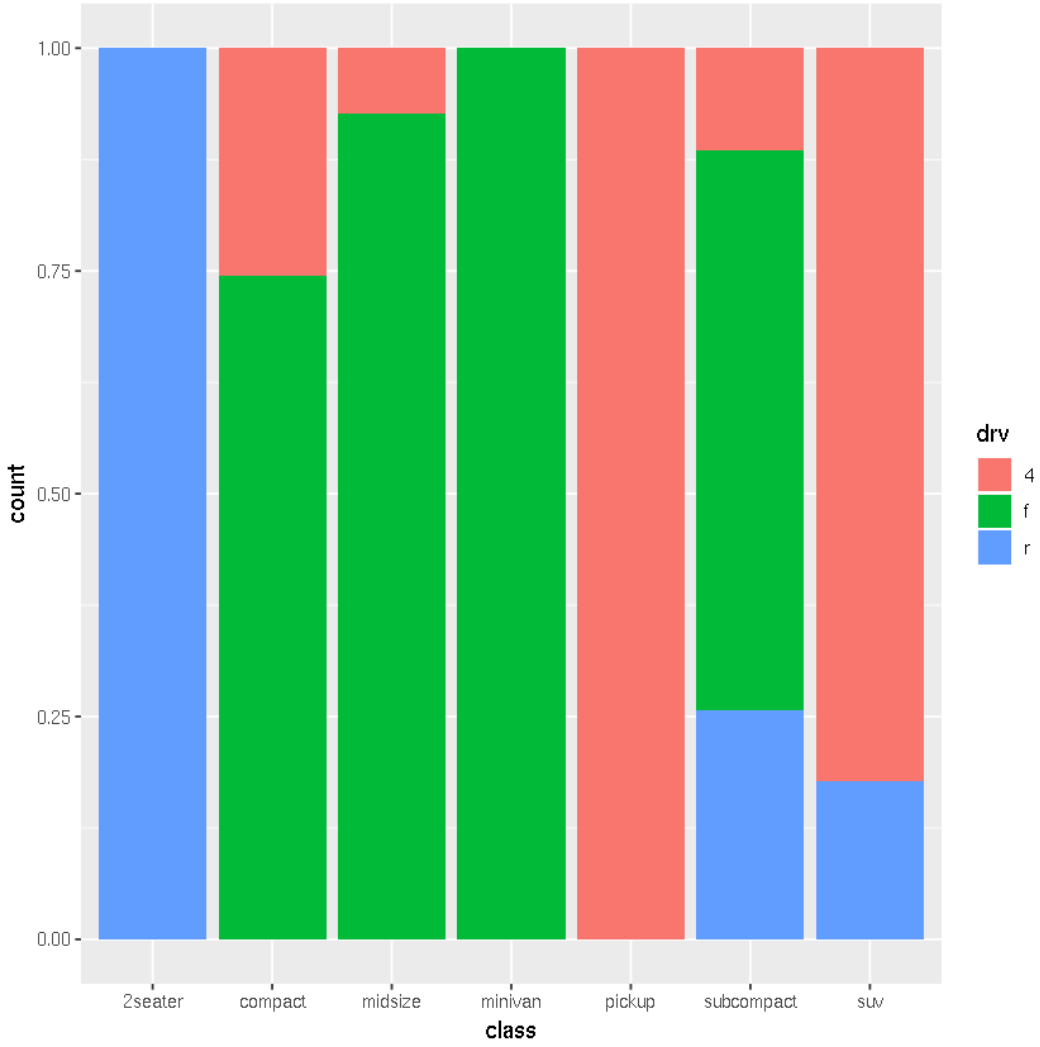

```
pl + geom_bar(aes(fill=drv)) # fill di dasarkan pada jumlah drv
```




```
pl + geom_bar(aes(fill=drv), position='dodge') # dipisahkan
```

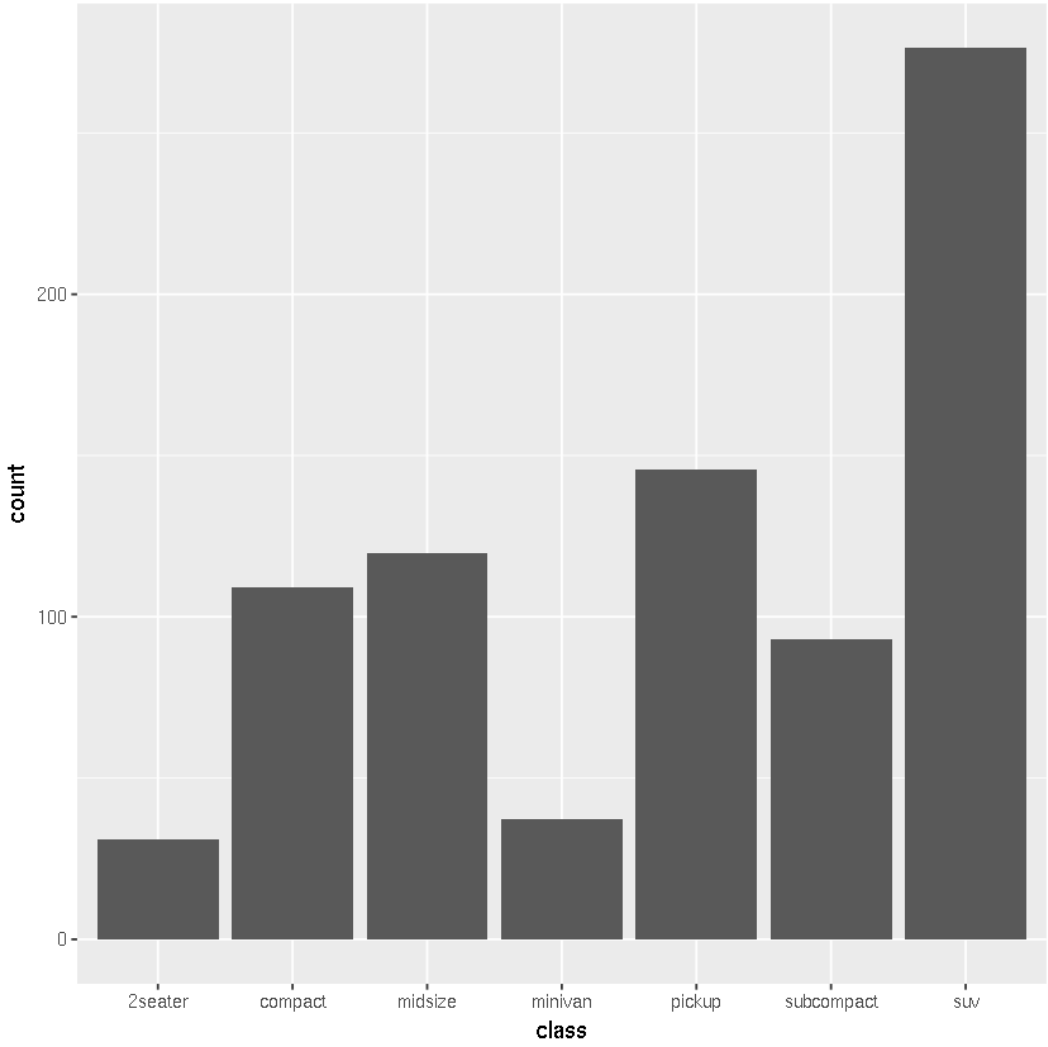



```
pl + geom_bar(aes(fill=drv), position='fill') # dihitung berdasarkan persentase
```



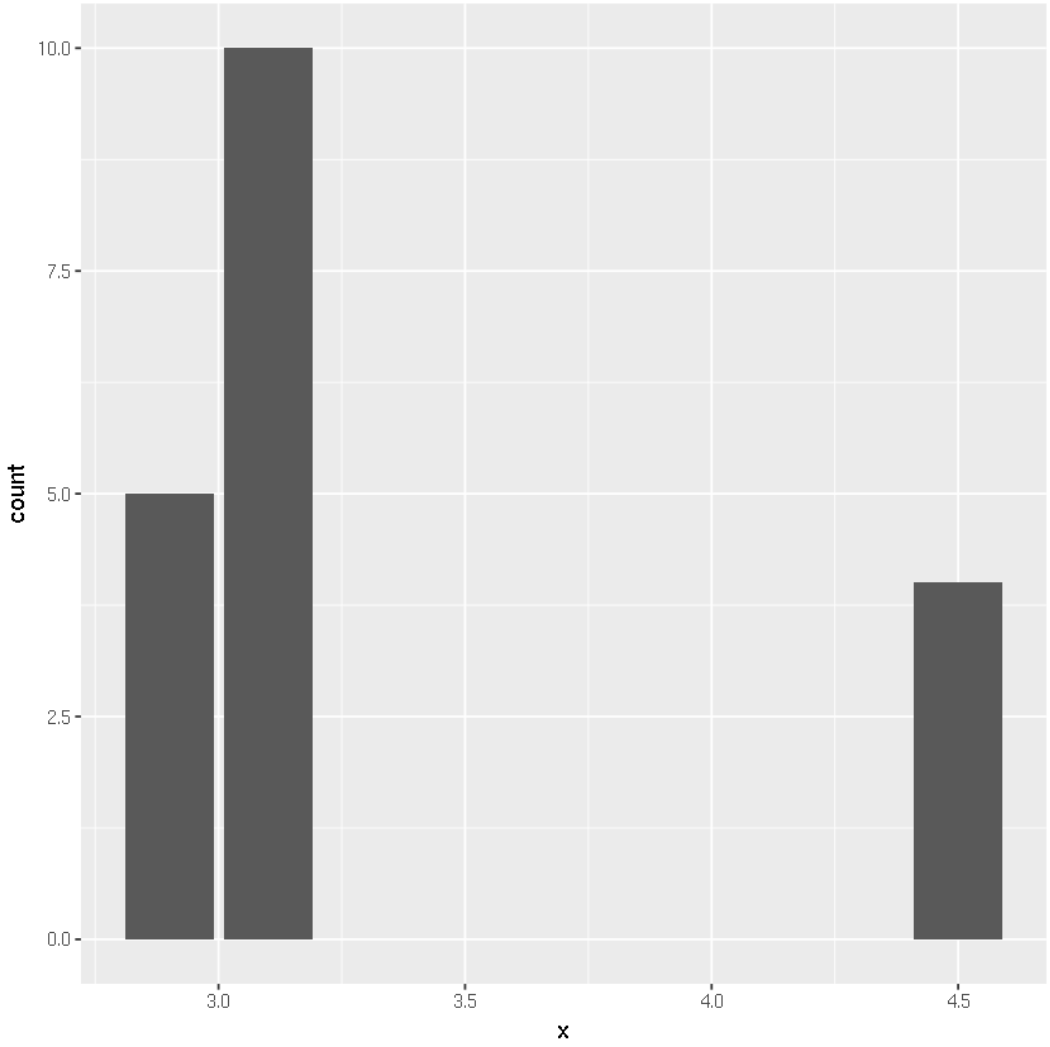

```
# help("geom_bar")
```

```
# Total engine displacement of each class
pl + geom_bar(aes(weight = displ))
```




```
# help("geom_bar")
```

```
# You can also use geom_bar() with continuous data, in which case
# it will show counts at unique locations
df <- data.frame(x = rep(c(2.9, 3.1, 4.5), c(5, 10, 4)))
ggplot(df, aes(x)) + geom_bar()
```



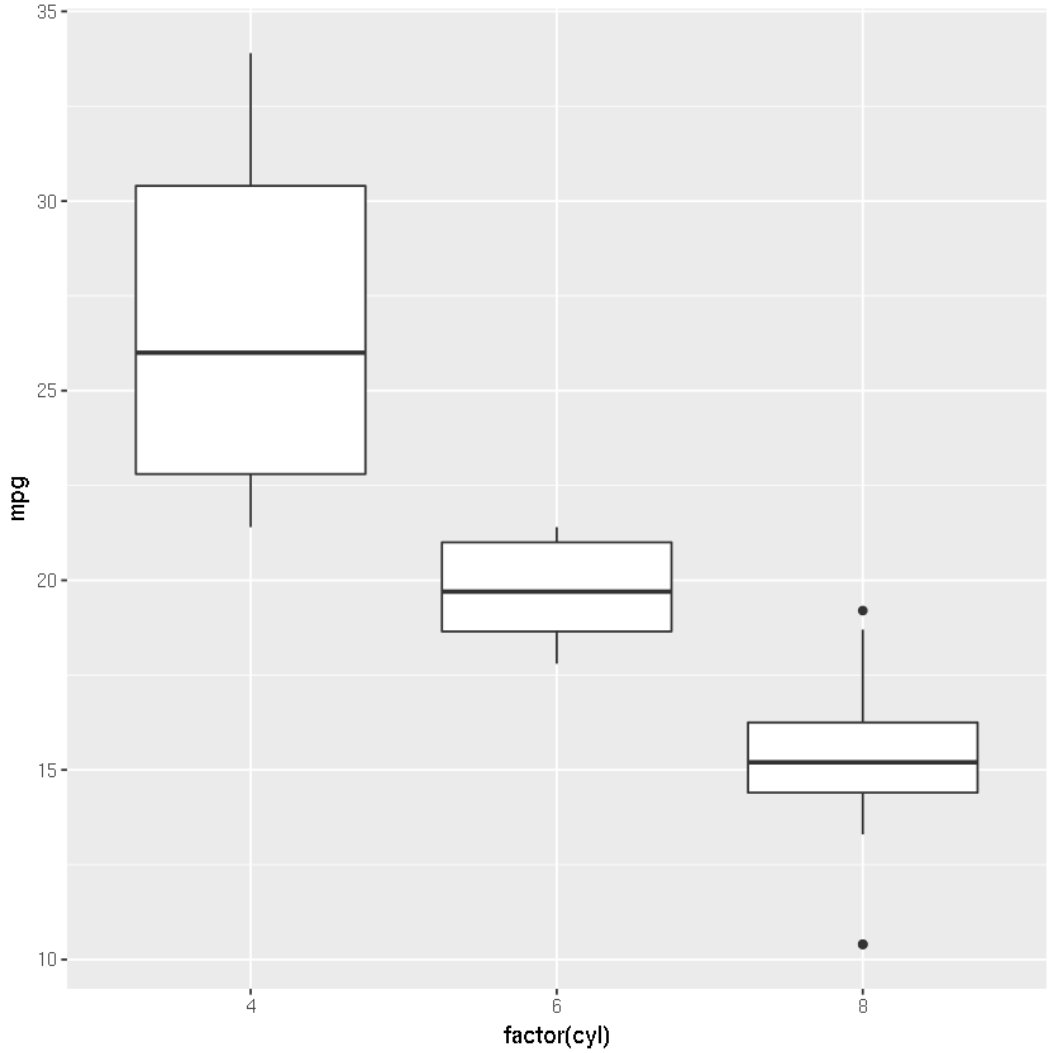
Boxplots

Digunakan untuk menampilkan sari statistik

```
df <- mtcars
head(df)
```

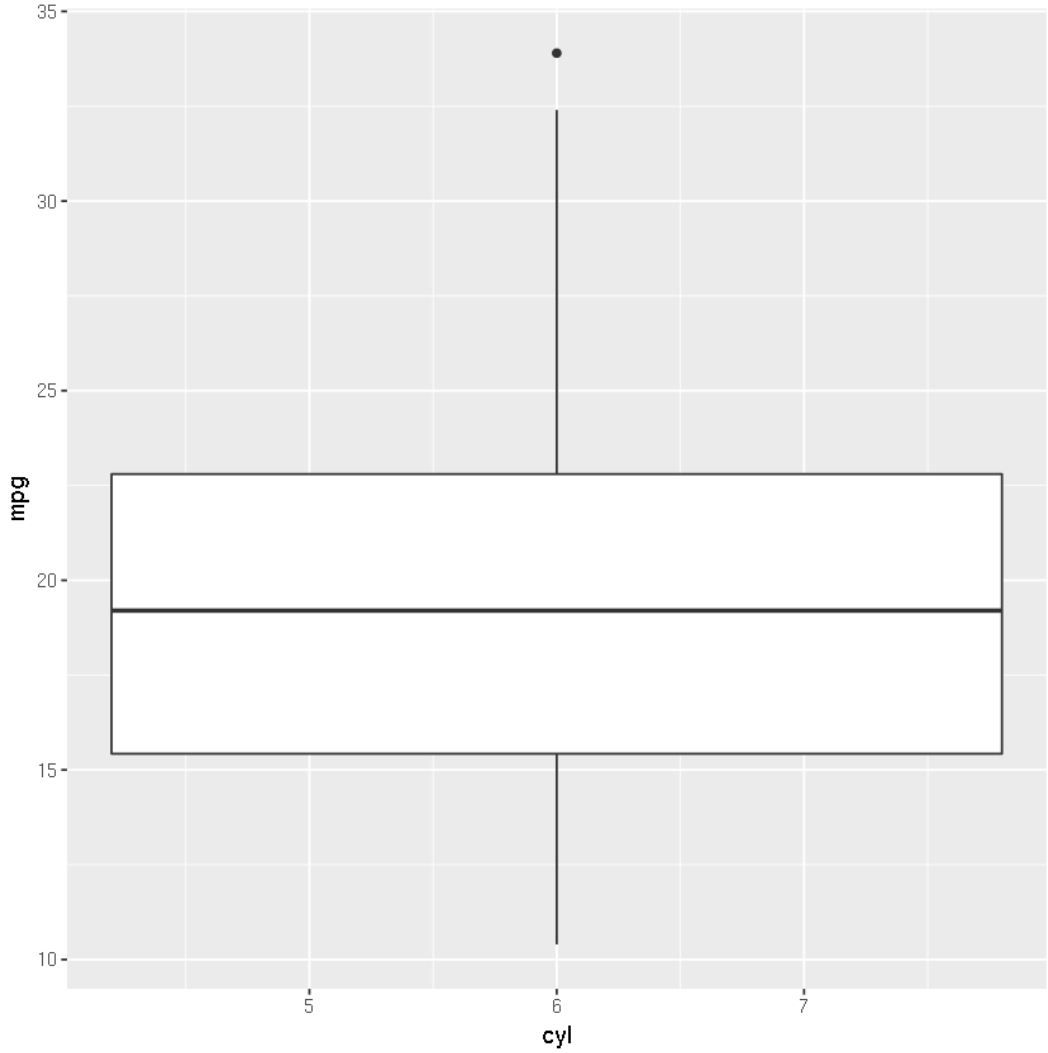
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

```
pl <- ggplot(df, aes(x = factor(cyl), y = mpg))
pl + geom_boxplot()
```



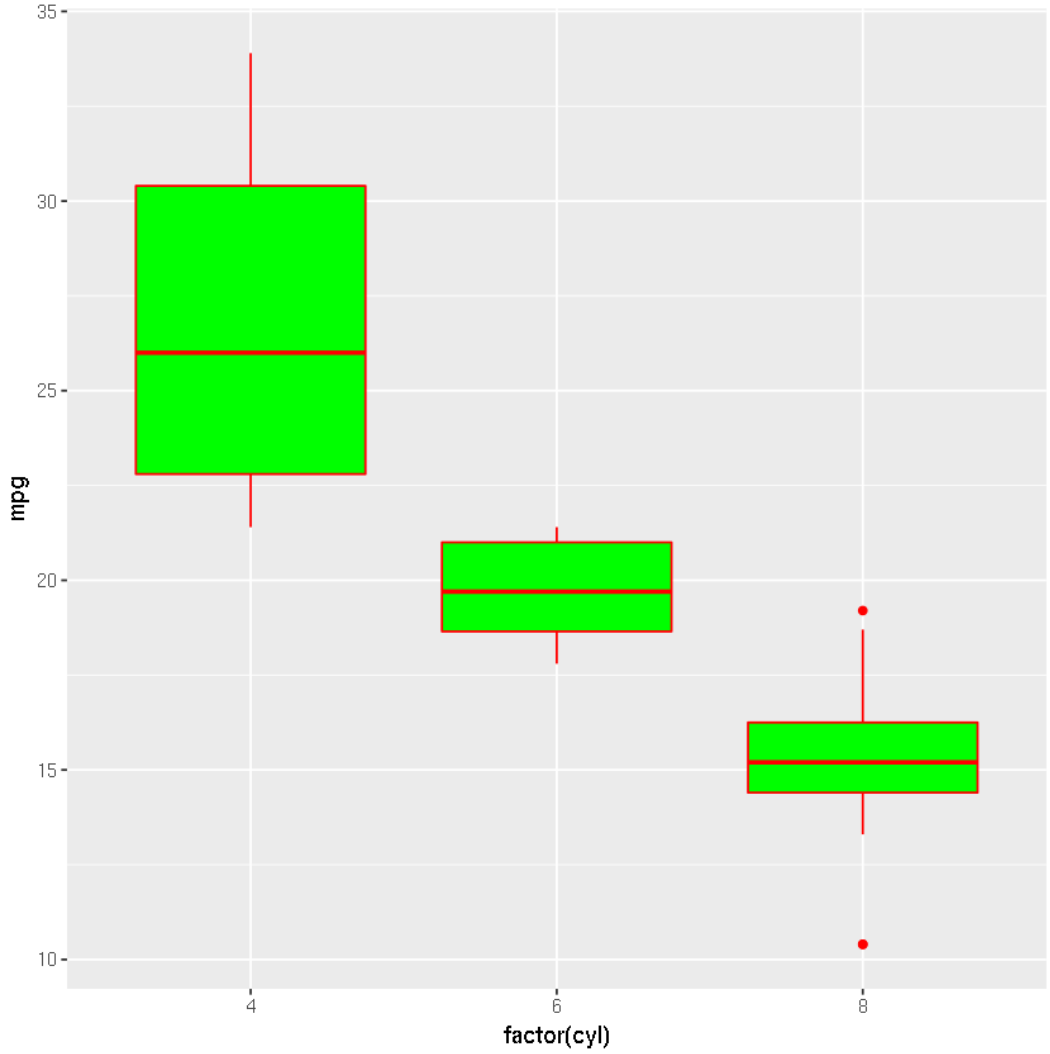

```
pl <- ggplot(df, aes(x = cyl, y = mpg)) # tanpa factor
pl + geom_boxplot()
```

```
Warning message:
“Continuous x aesthetic -- did you forget aes(group=...)?”
```

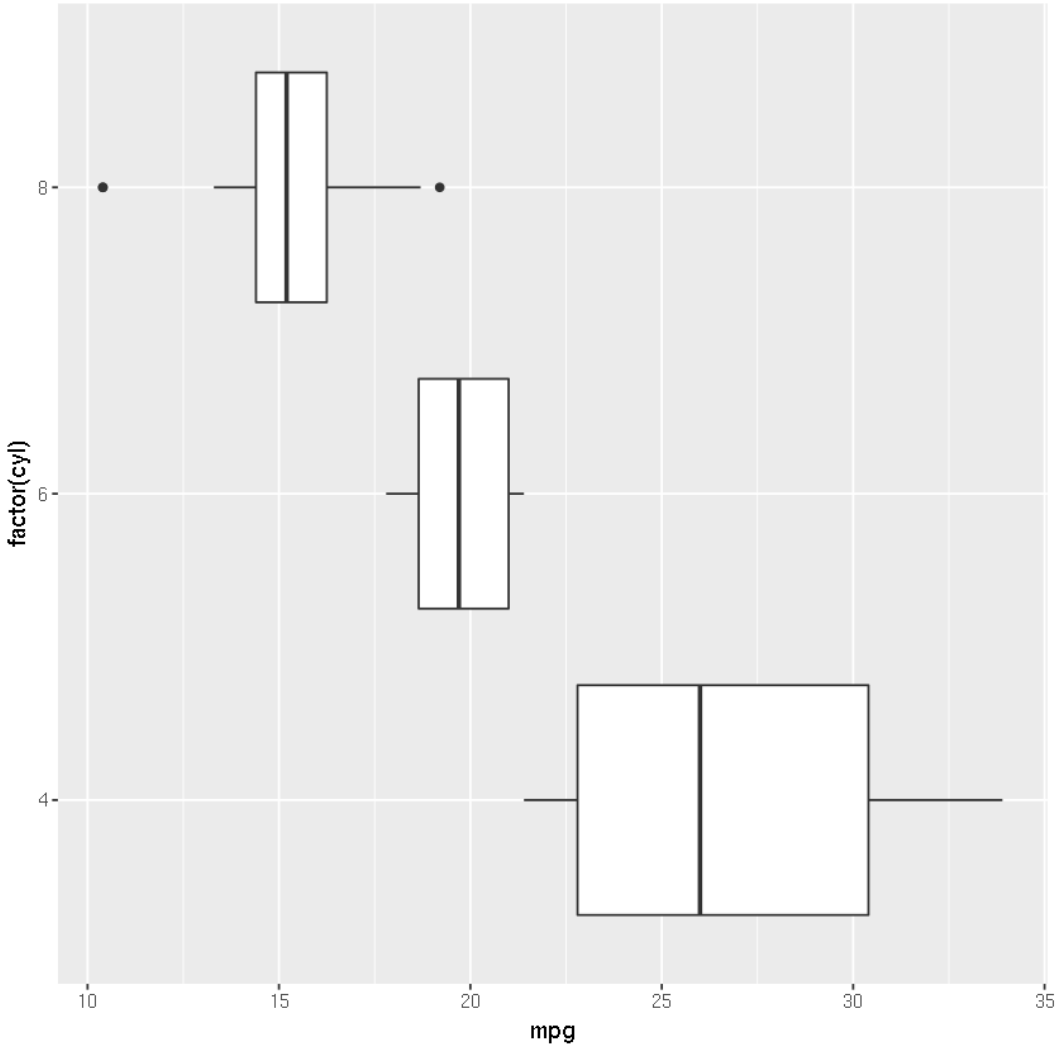



```
# help("geom_boxplot")
```

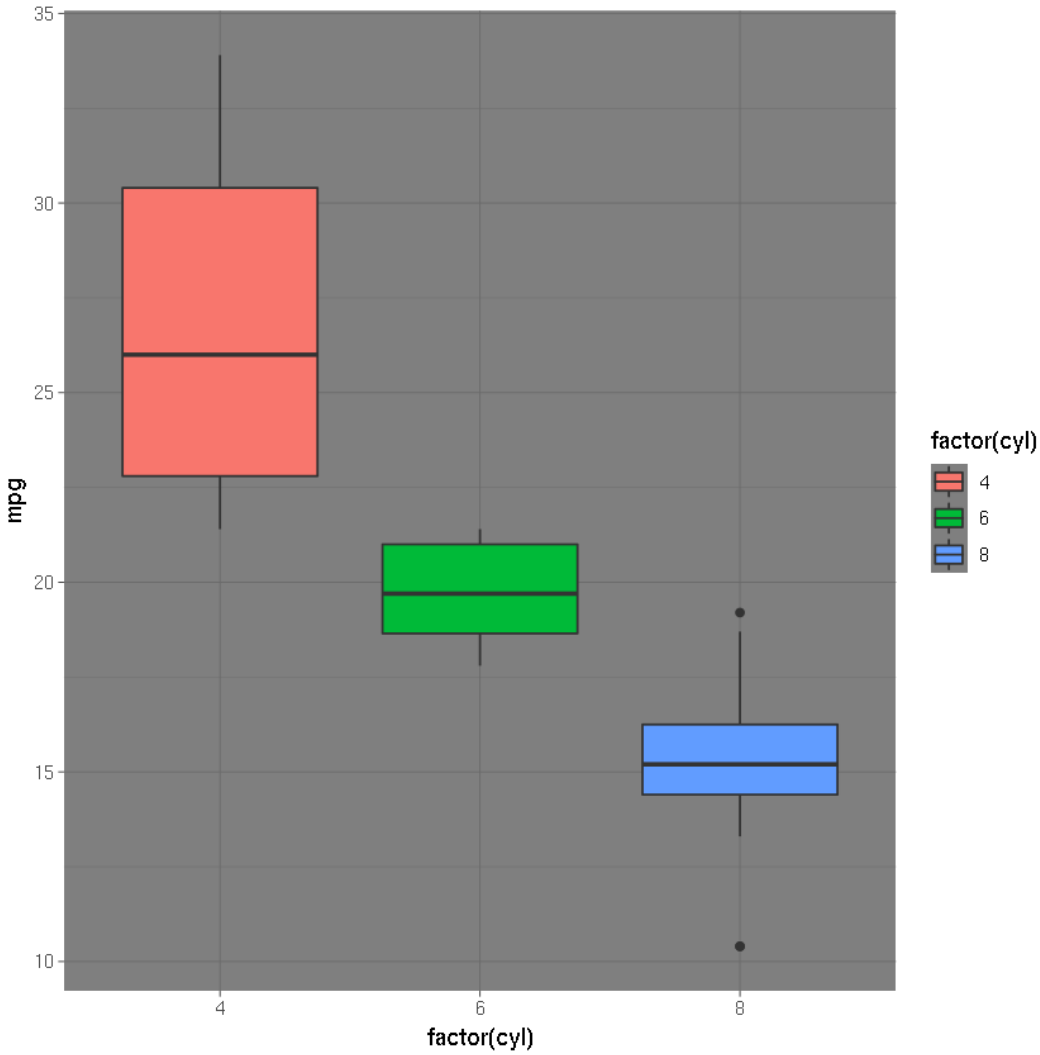
```
pl <- ggplot(df, aes(x = factor(cyl), y = mpg))
pl + geom_boxplot(color='red', fill = 'green')
```




```
# memutar koordinat
pl + geom_boxplot() + coord_flip()
```




```
pl + geom_boxplot(aes(fill=factor(cyl))) + theme_dark()
```



Visualisasi dua variabel

```
head(movies)
```

title	year	length	budget	rating	votes	r1	r2	r3	r4	...	r9	r10	mpaa	Action	Ar
$	1971	121	NA	6.4	348	4.5	4.5	4.5	4.5	...	4.5	4.5		0	0
$1000 a Touchdown	1939	71	NA	6.0	20	0.0	14.5	4.5	24.5	...	4.5	14.5		0	0
$21 a Day Once a Month	1941	7	NA	8.2	5	0.0	0.0	0.0	0.0	...	24.5	24.5		0	1
$40,000	1996	70	NA	8.2	6	14.5	0.0	0.0	0.0	...	34.5	45.5		0	0
$50,000 Climax Show, The	1975	71	NA	3.4	17	24.5	4.5	0.0	14.5	...	0.0	24.5		0	0
$pent	2000	91	NA	4.3	45	4.5	4.5	4.5	14.5	...	14.5	14.5		0	0

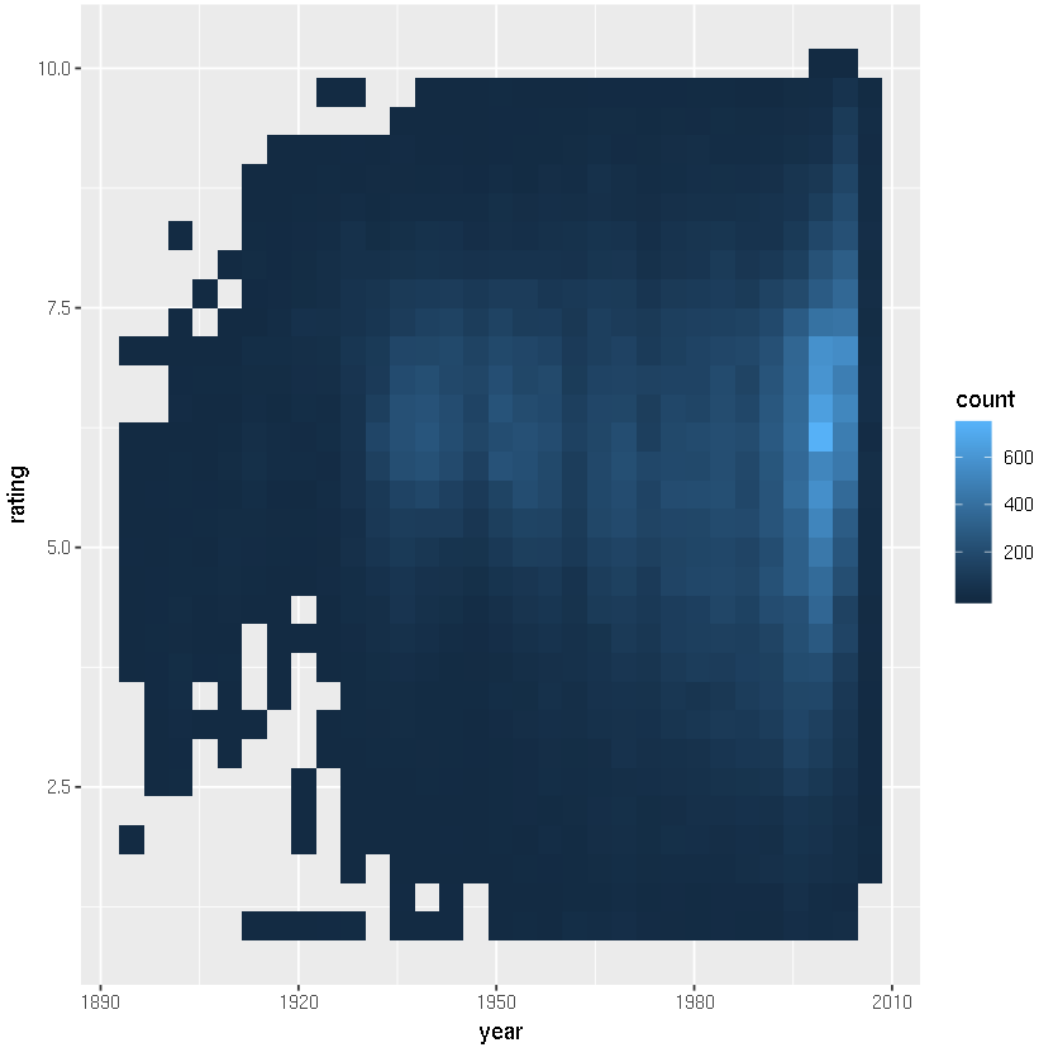
```
colnames(movies)
```

1. 'title'
2. 'year'
3. 'length'
4. 'budget'
5. 'rating'
6. 'votes'
7. 'r1'
8. 'r2'
9. 'r3'
10. 'r4'

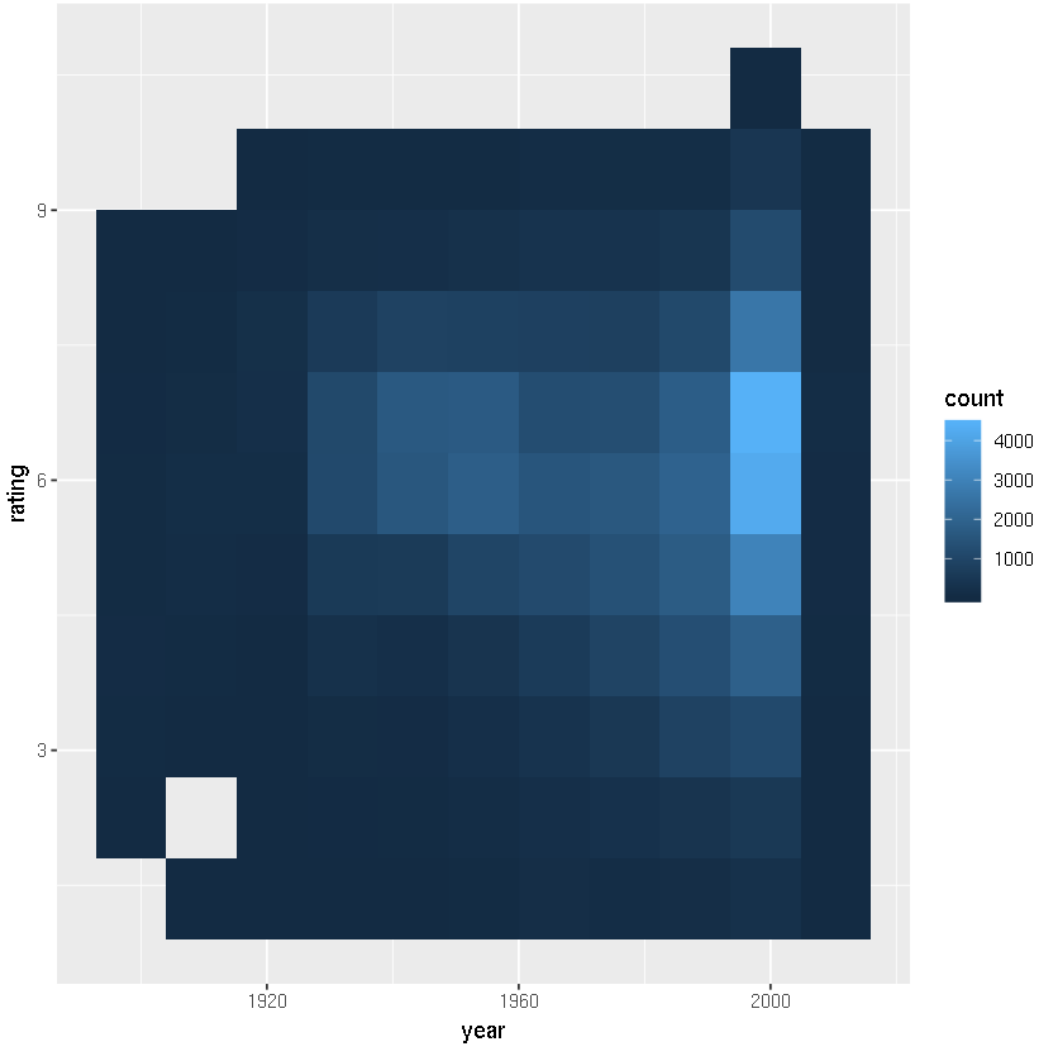
11. 'r5'
12. 'r6'
13. 'r7'
14. 'r8'
15. 'r9'
16. 'r10'
17. 'mpaa'
18. 'Action'
19. 'Animation'
20. 'Comedy'
21. 'Drama'
22. 'Documentary'
23. 'Romance'
24. 'Short'

```
pl <- ggplot(movies, aes(x=year, y=rating))
```

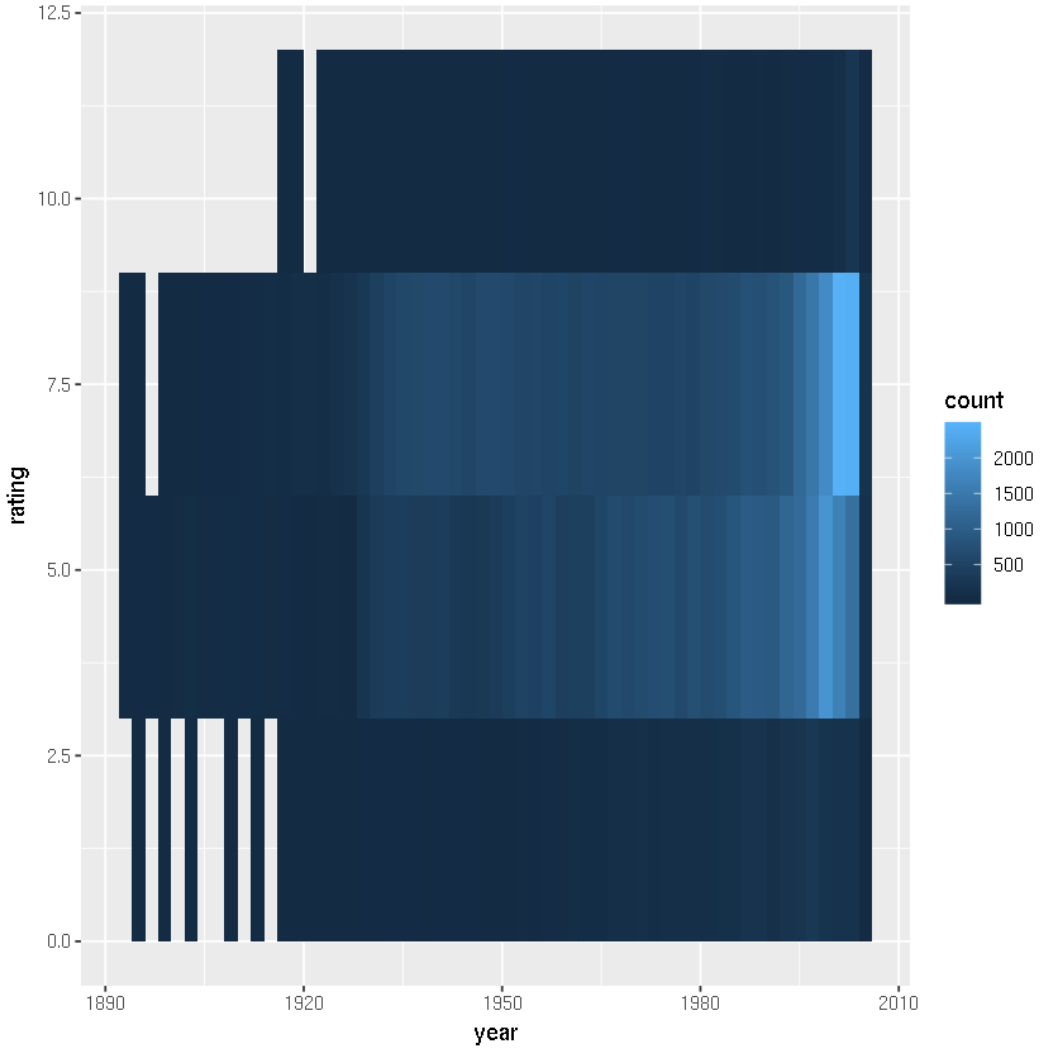
```
pl + geom_bin2d()
# mirip dengan heatmap
# jumlah kejadian dihitung berdasarkan warna
```




```
pl + geom_bin2d(bins=10)
```

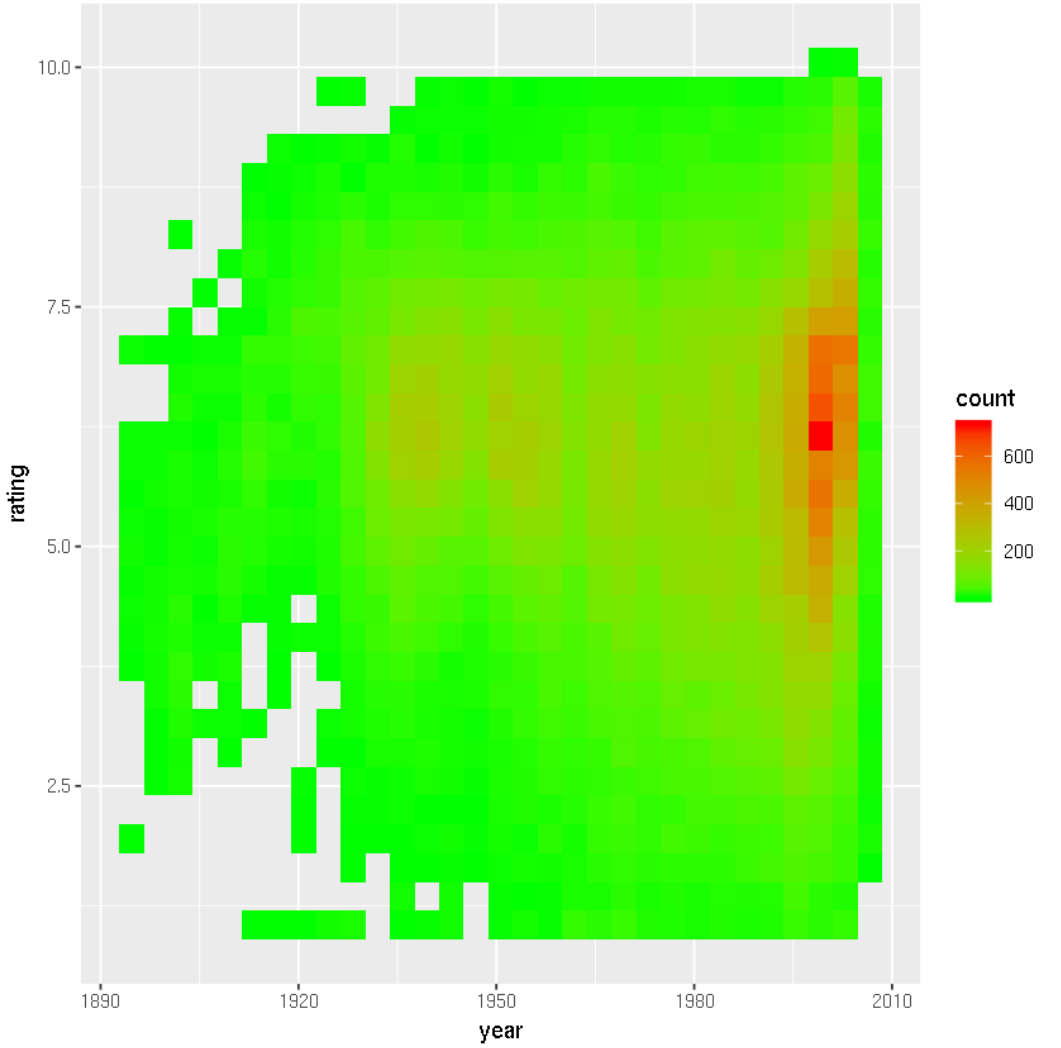



```
pl + geom_bin2d(binwidth=c(2,3), bins=10)
```

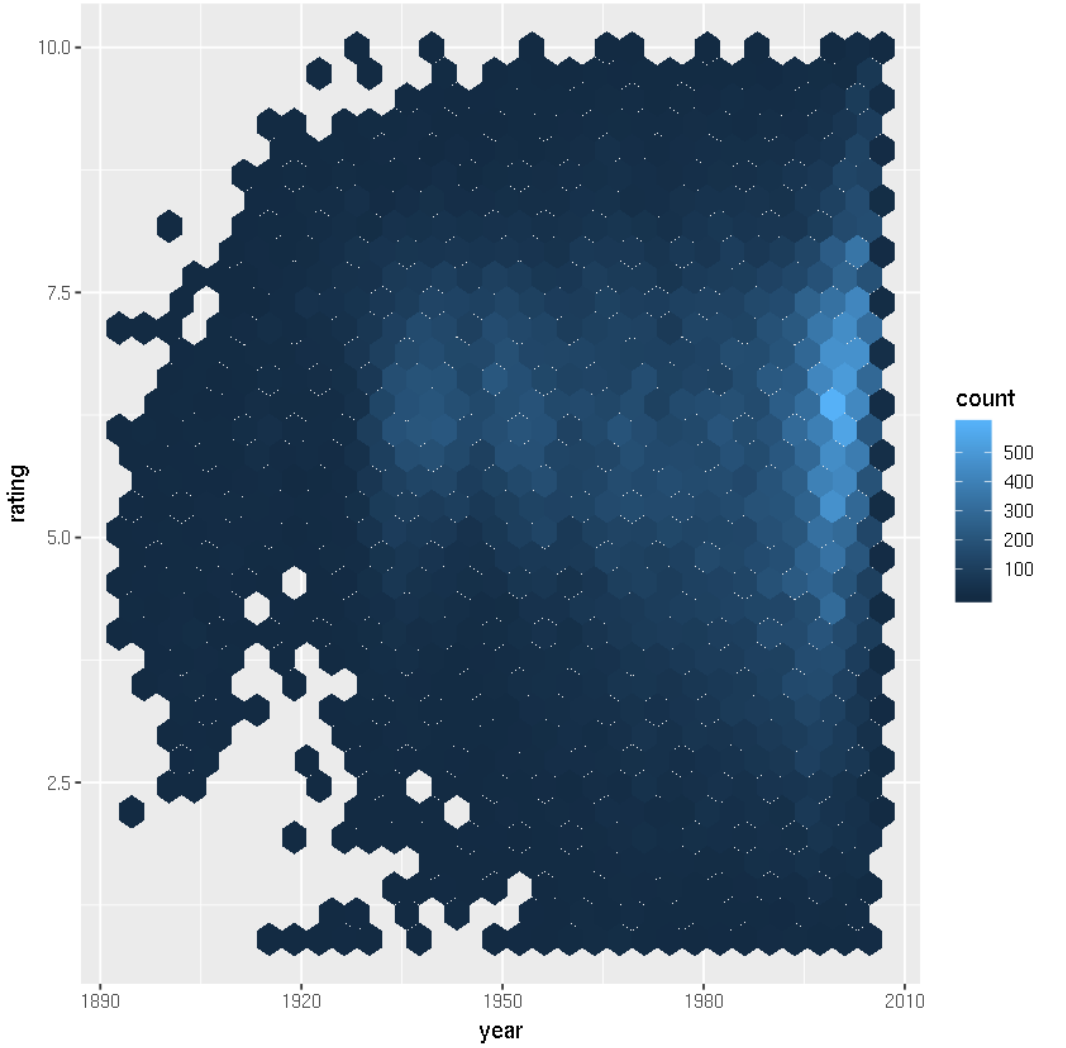



```
# help("geom_bin2d")
```

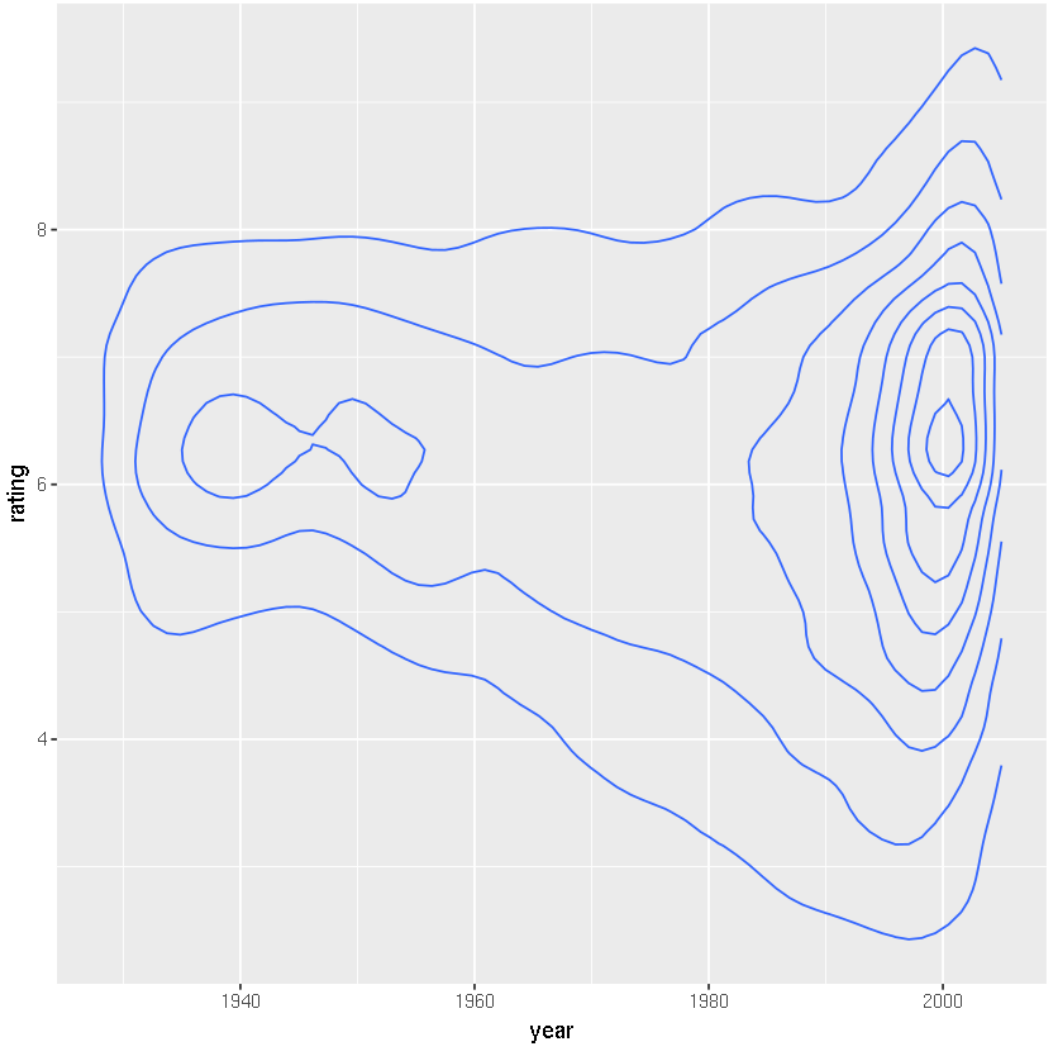
```
# mengubah warna
pl2 <- pl + geom_bin2d()
pl2 + scale_fill_gradient(high = 'red', low='green')
```




```
# mengubah shape jadi hexagon
library(hexbin)
pl + geom_hex()
```




```
pl + geom_density2d()
```



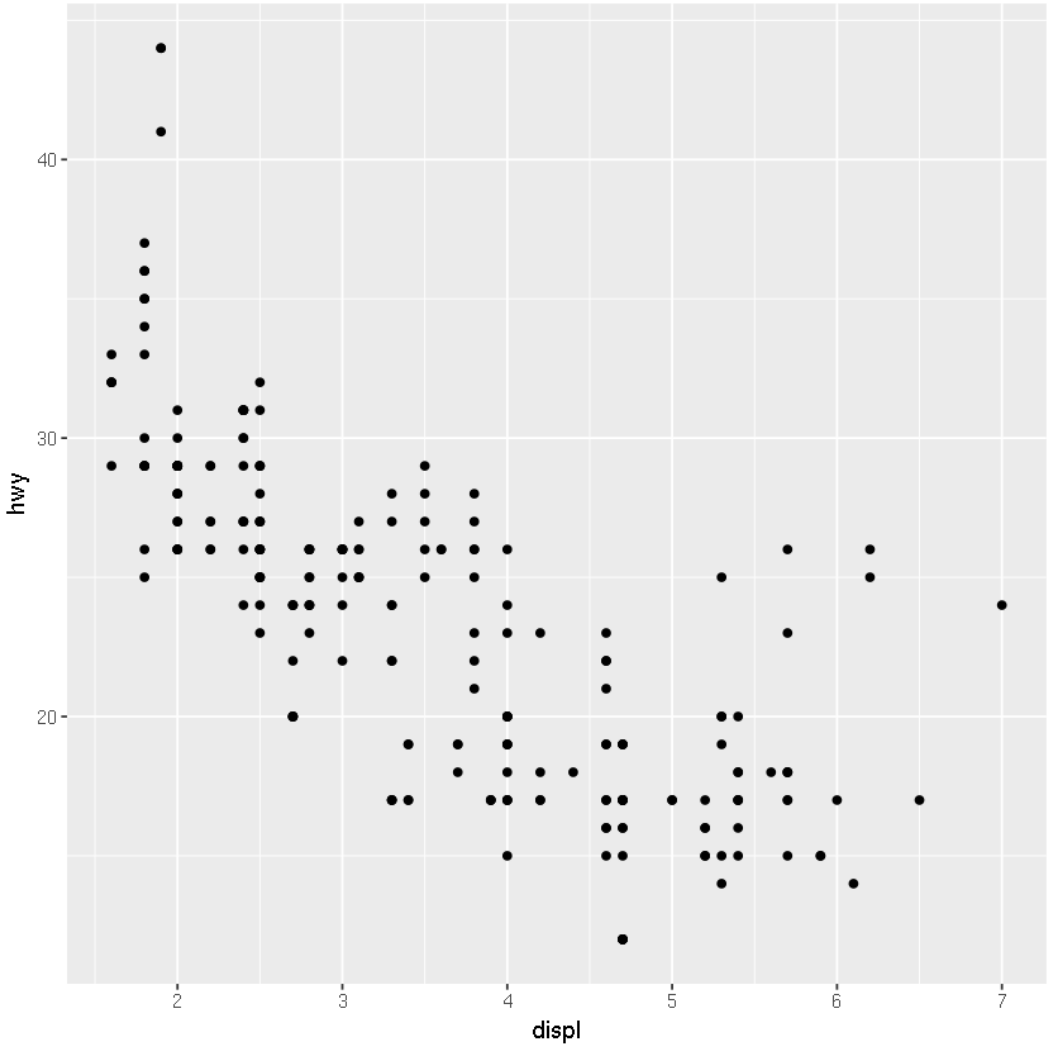
Koordinat dan *faceting*

```
head(mpg)
```

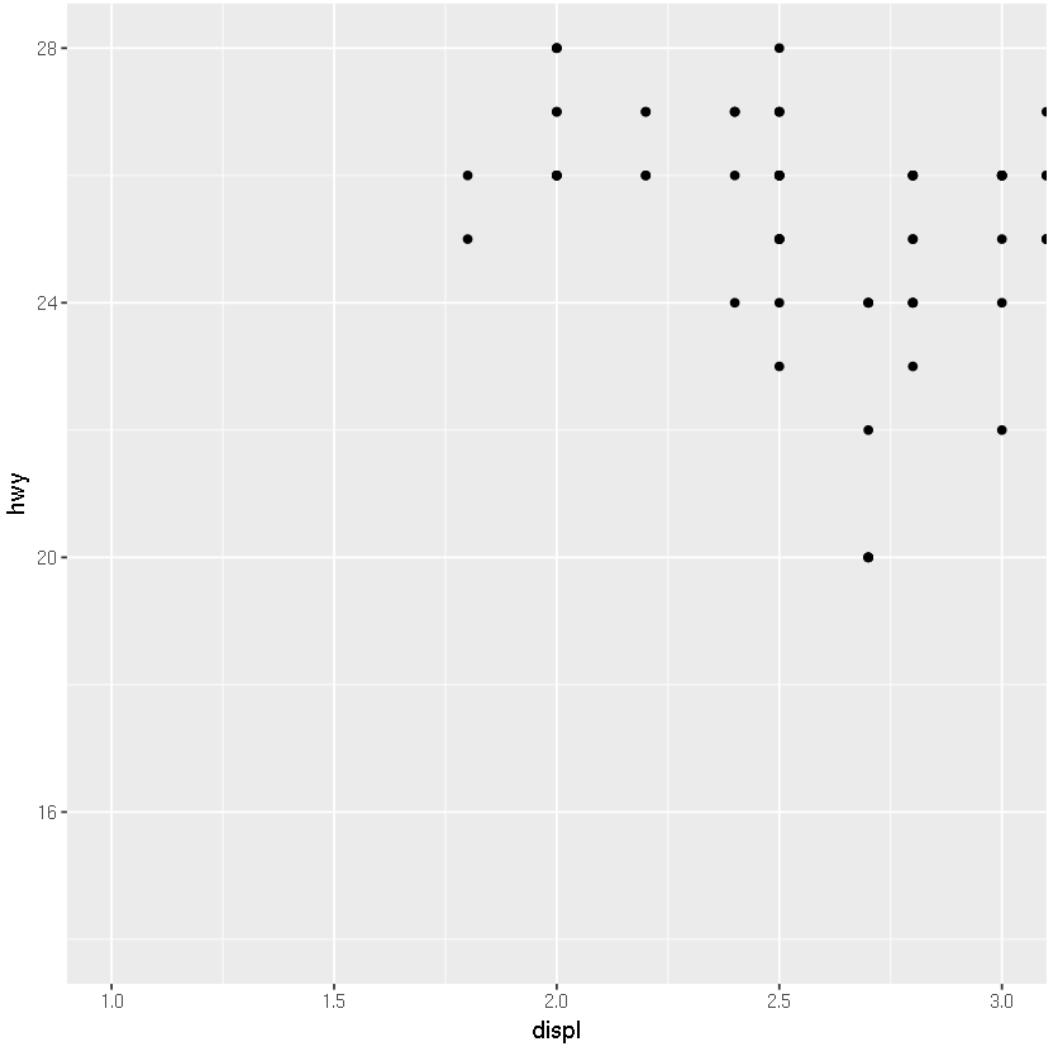
manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact

```
# scatterplot simpel
pl <- ggplot(mpg, aes(x=displ, y=hwy)) +
            geom_point()
```

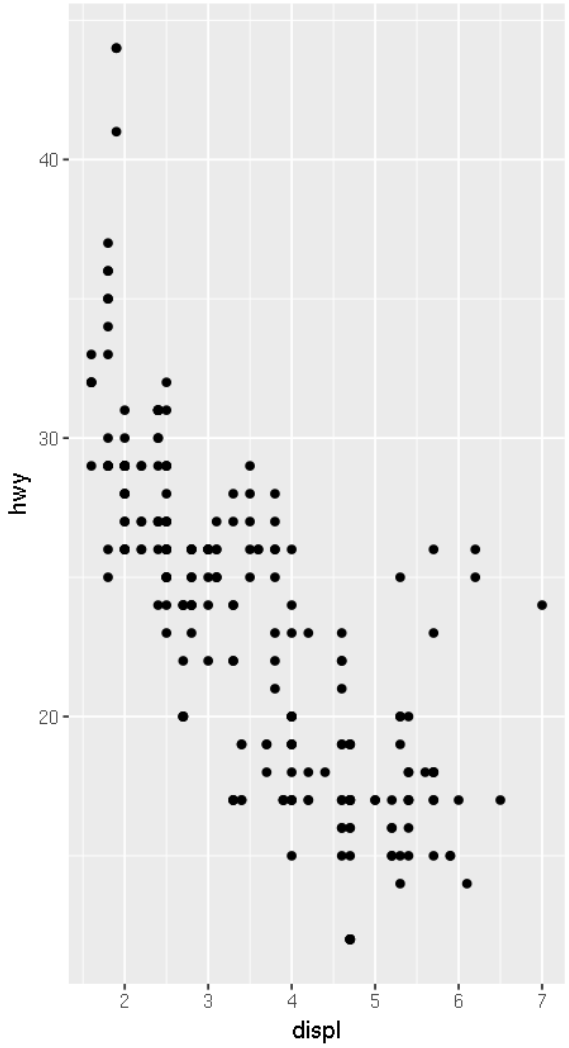
```
pl
```



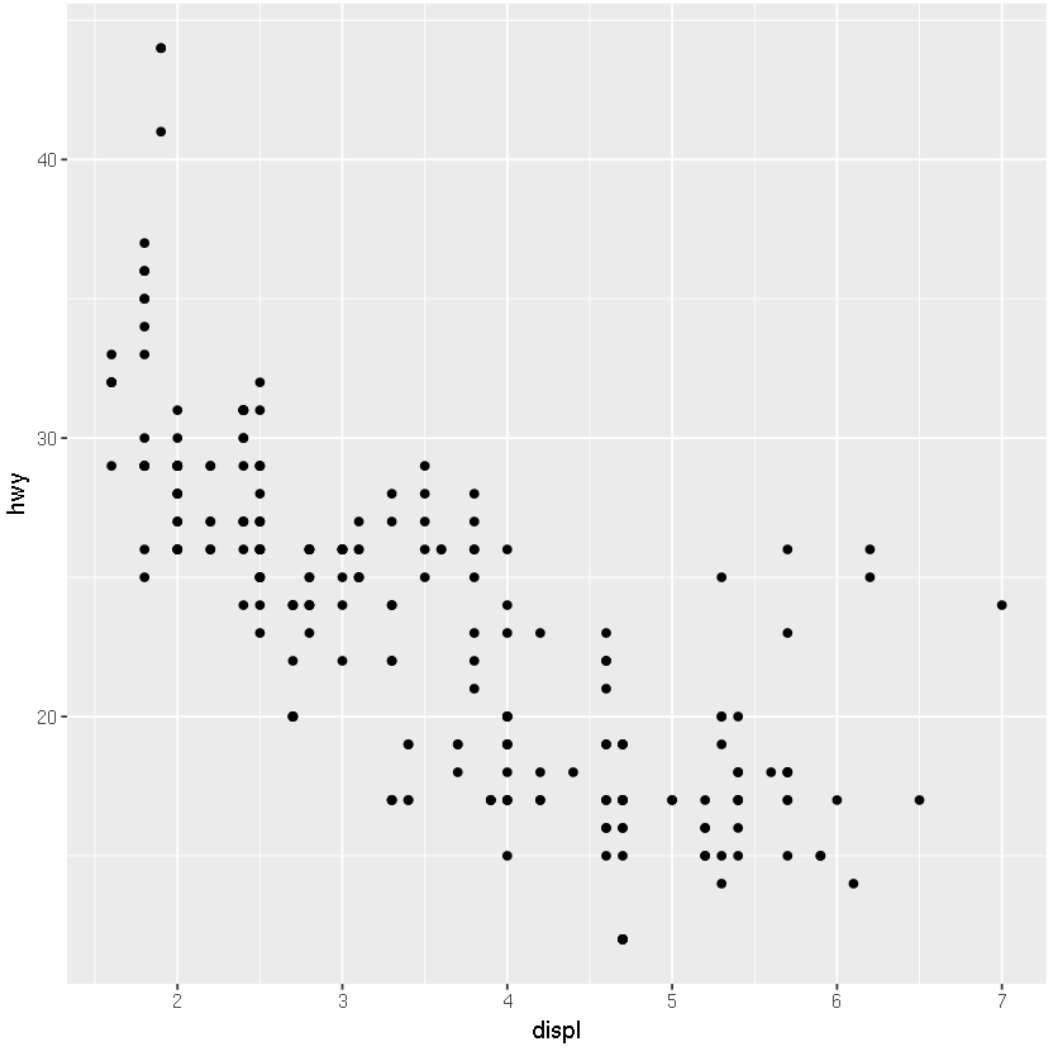

```
# Mengatur limit sumbu-x dan y
pl + coord_cartesian(xlim = c(1,3), ylim = c(14,28))
```



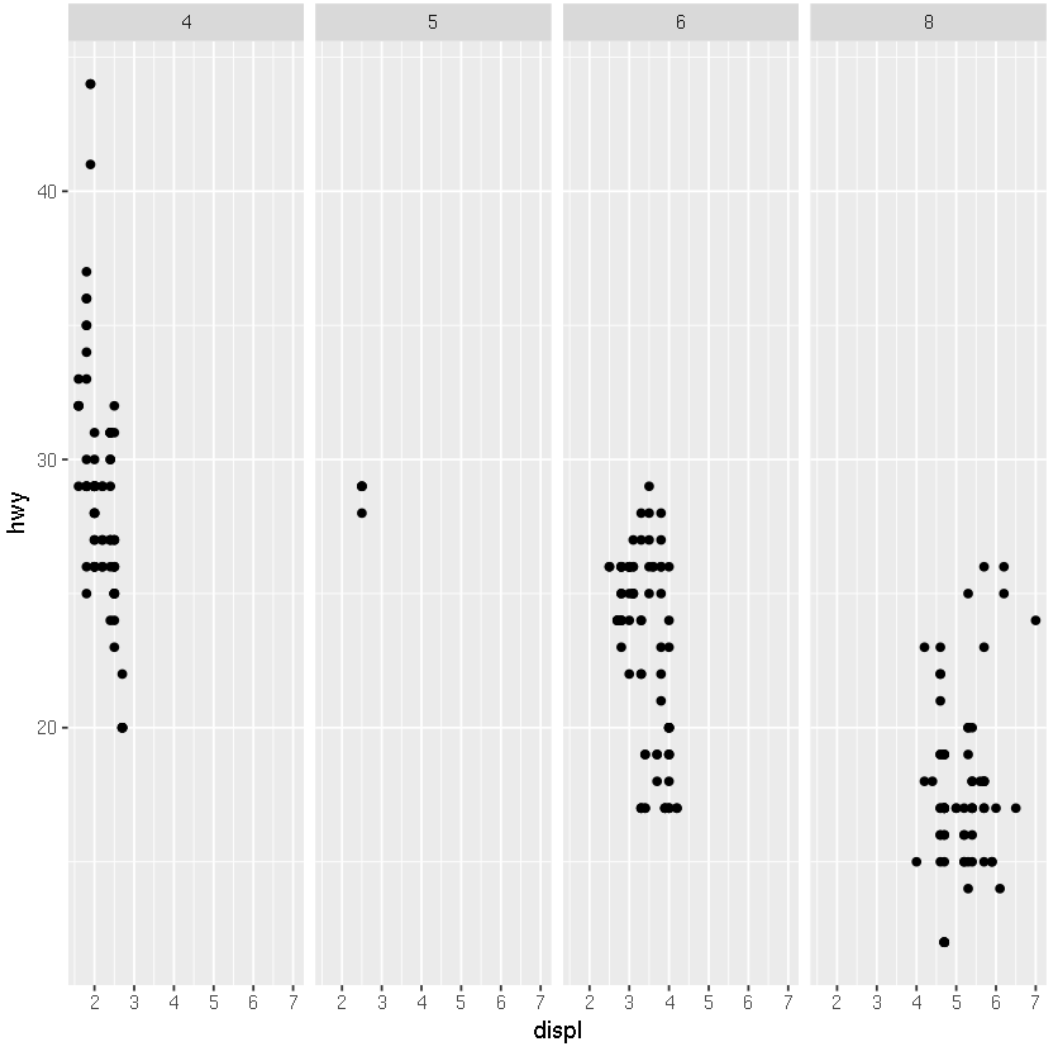

```
# Mengatur rasio sumbu
pl + coord_fixed(ratio = 1/3) # y/x
```




```
# Facets
pl
```

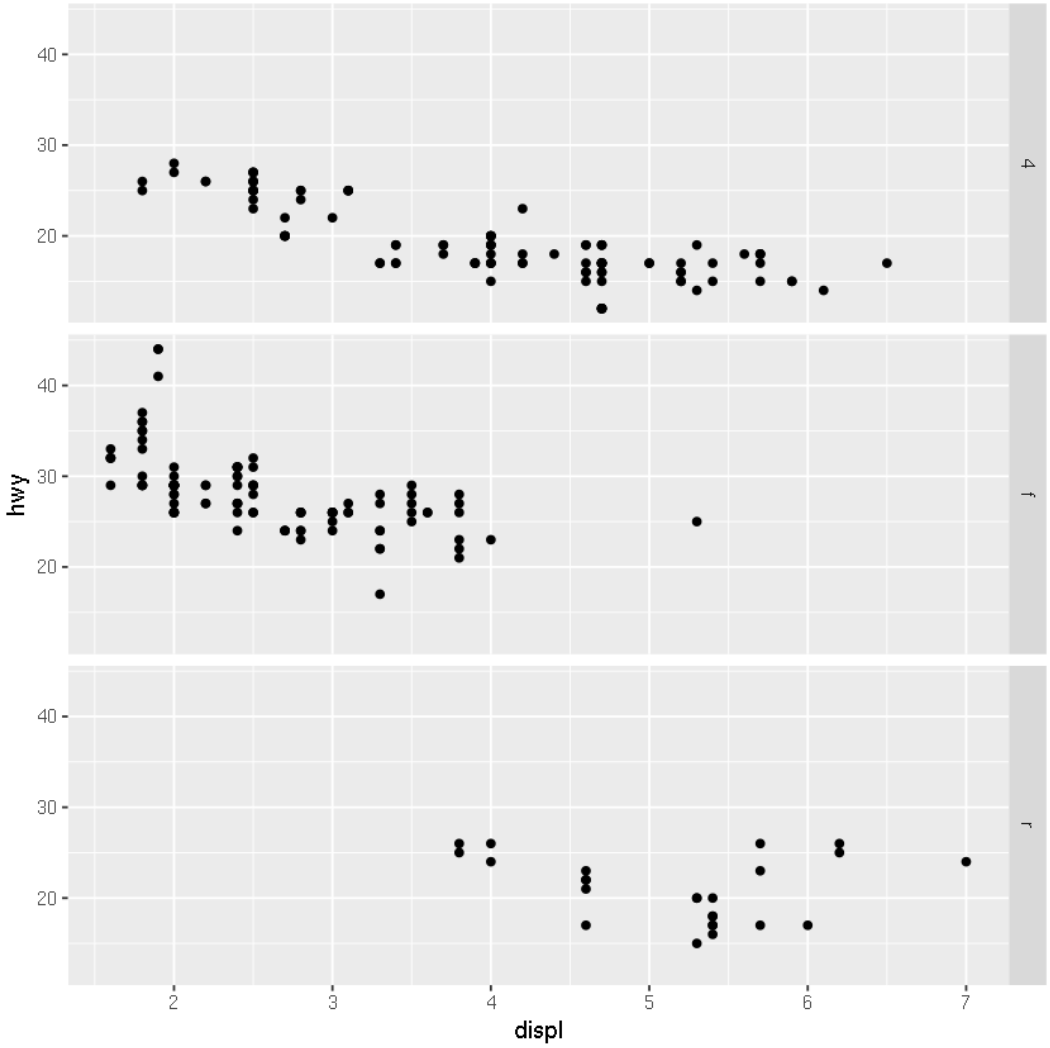



```
pl + facet_grid(.~cyl) # dipisahkan menurut silinder pada sumbu-x
```

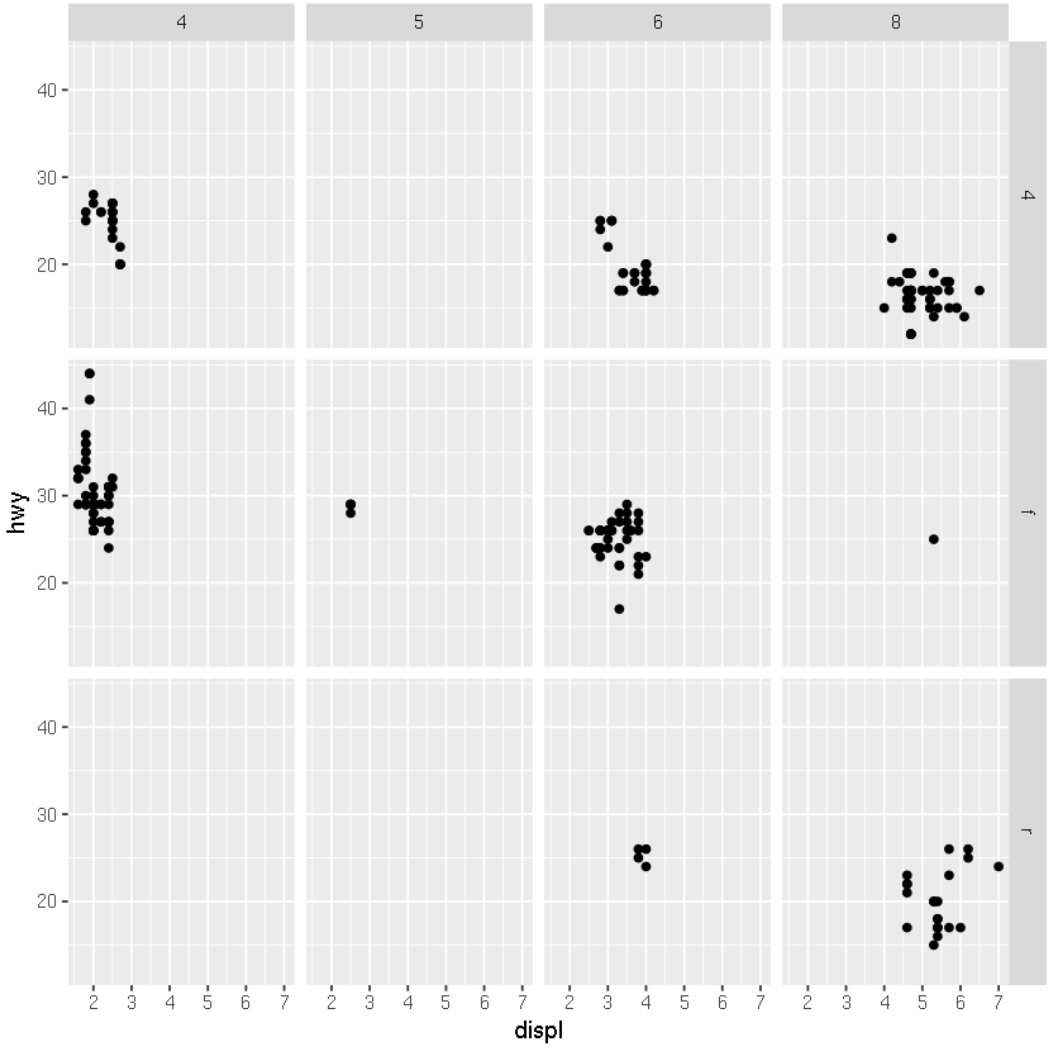


sintaks: `facet_grid(sb-x~sb-y)`

```
pl + facet_grid(drv~.) # membagi facet sumbu-y dengan menggunakan drv
```

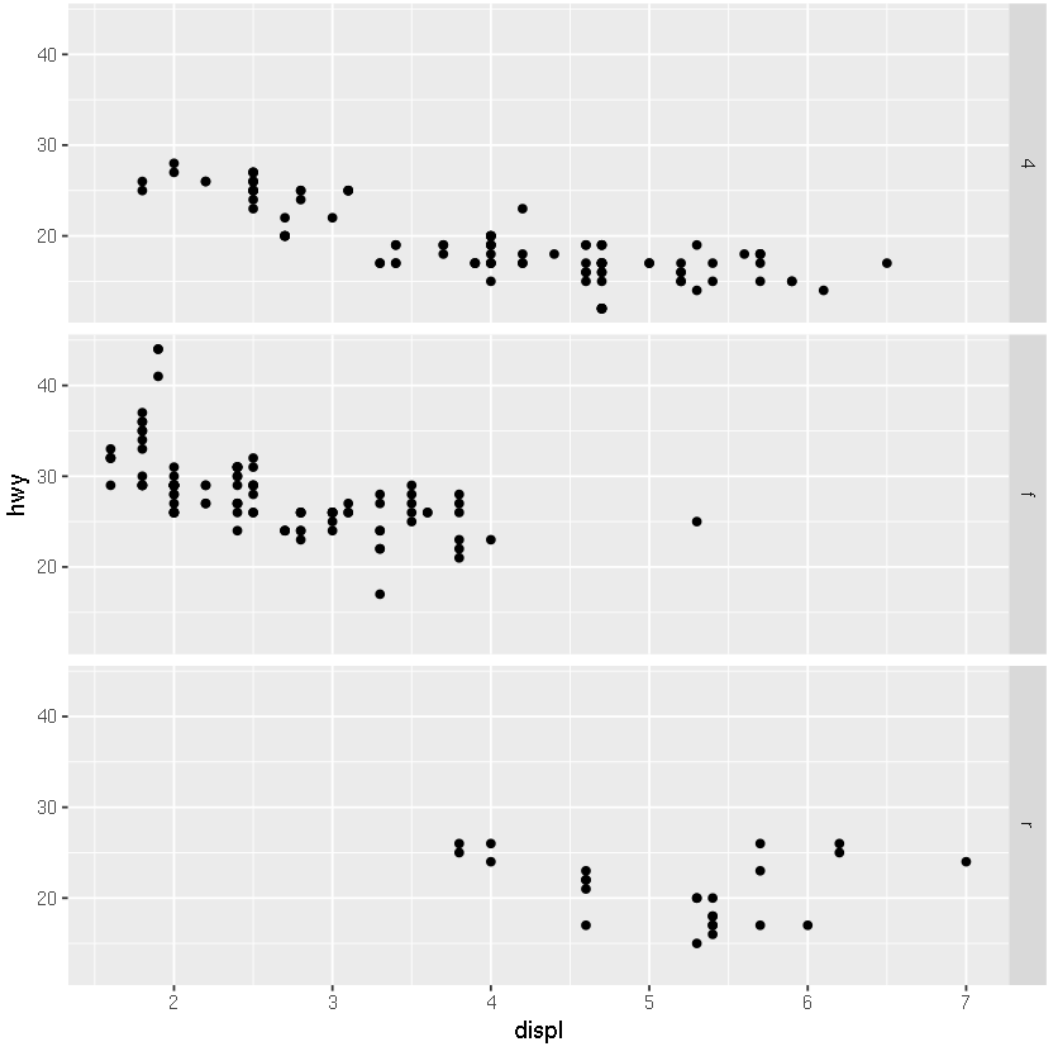



```
pl + facet_grid(drv~cyl)
```

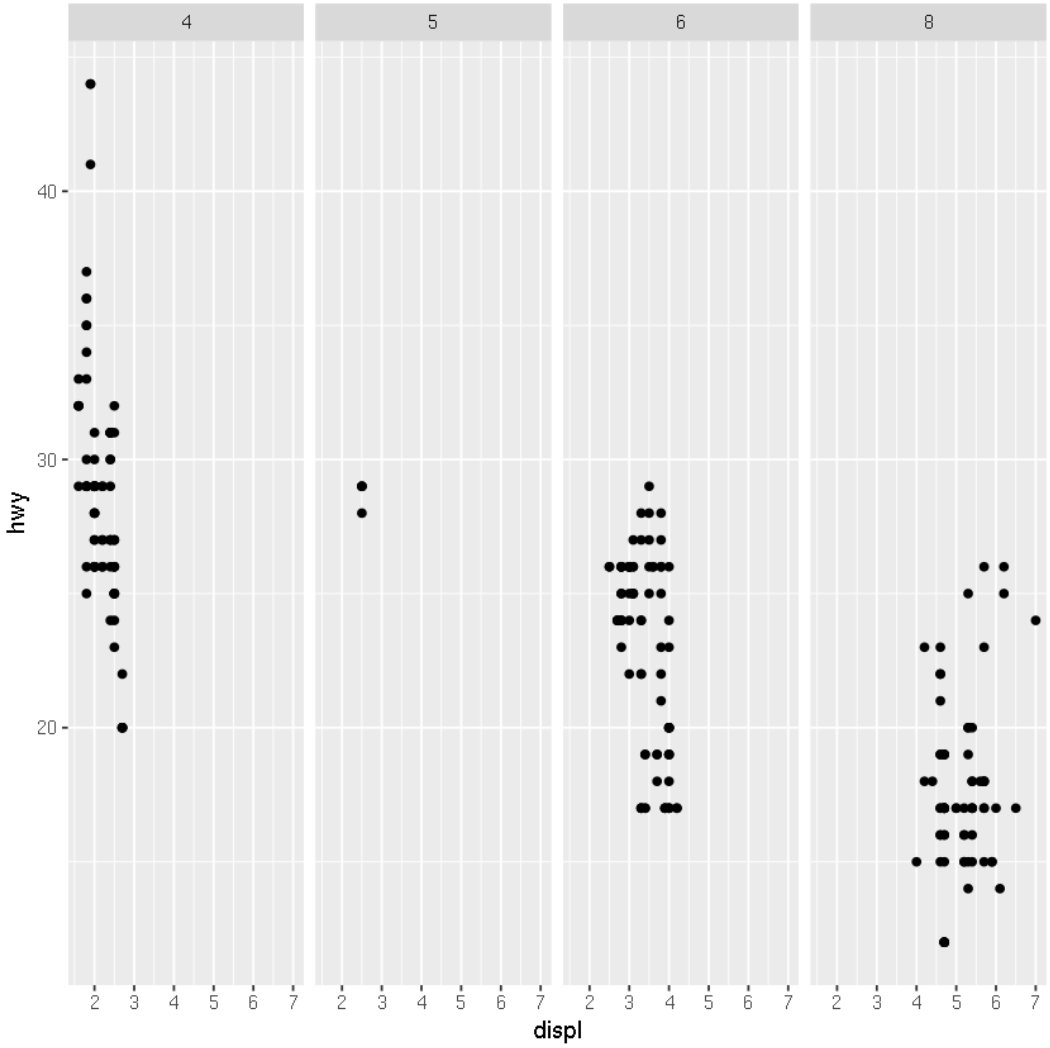



```
# Untuk mengetahui secara lebih lanjut, jalankan perintah:
help("facet_grid")
```

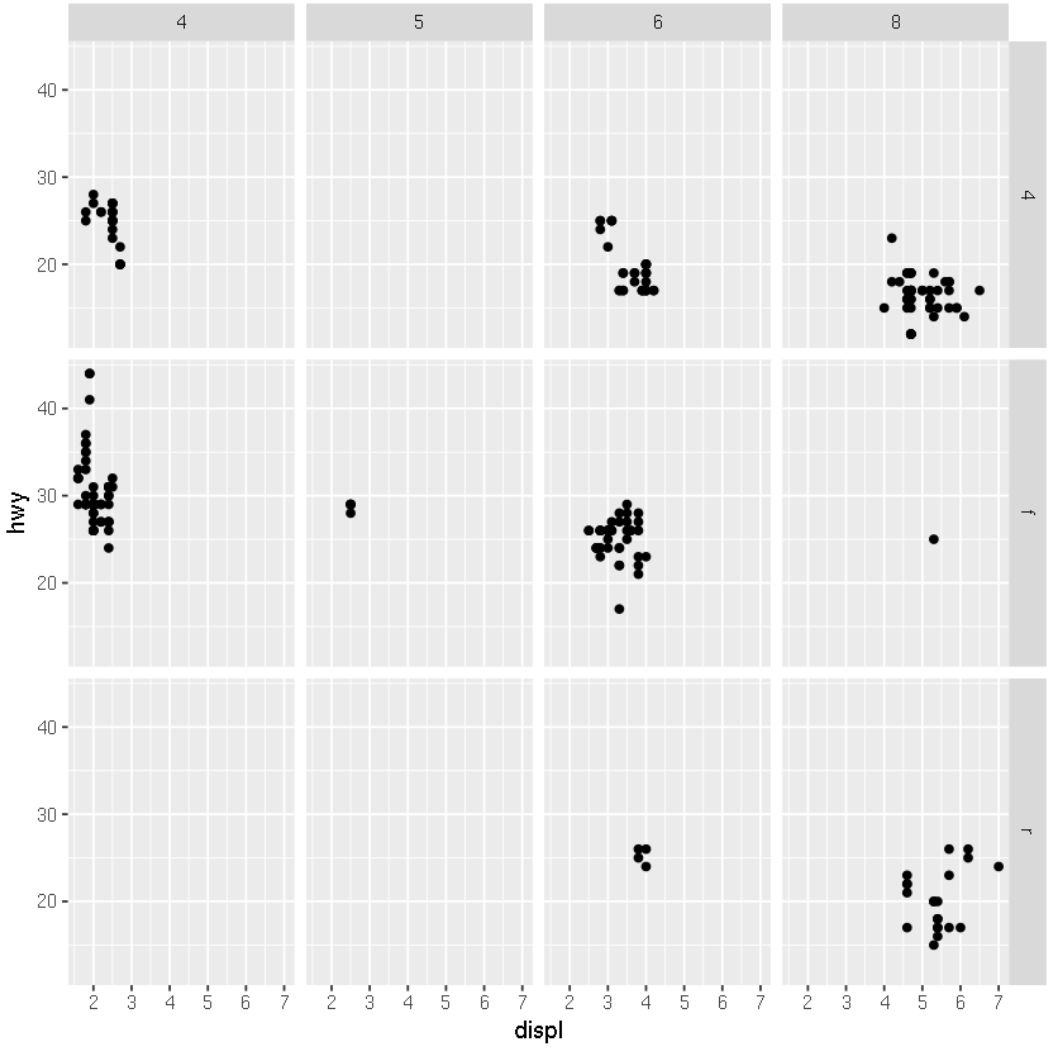
```
# Use vars() to supply variables from the dataset: (berbasis baris, kolom)
pl + facet_grid(rows = vars(drv))
```




```
pl + facet_grid(cols = vars(cyl))
```




```
pl + facet_grid(vars(drv), vars(cyl))
```

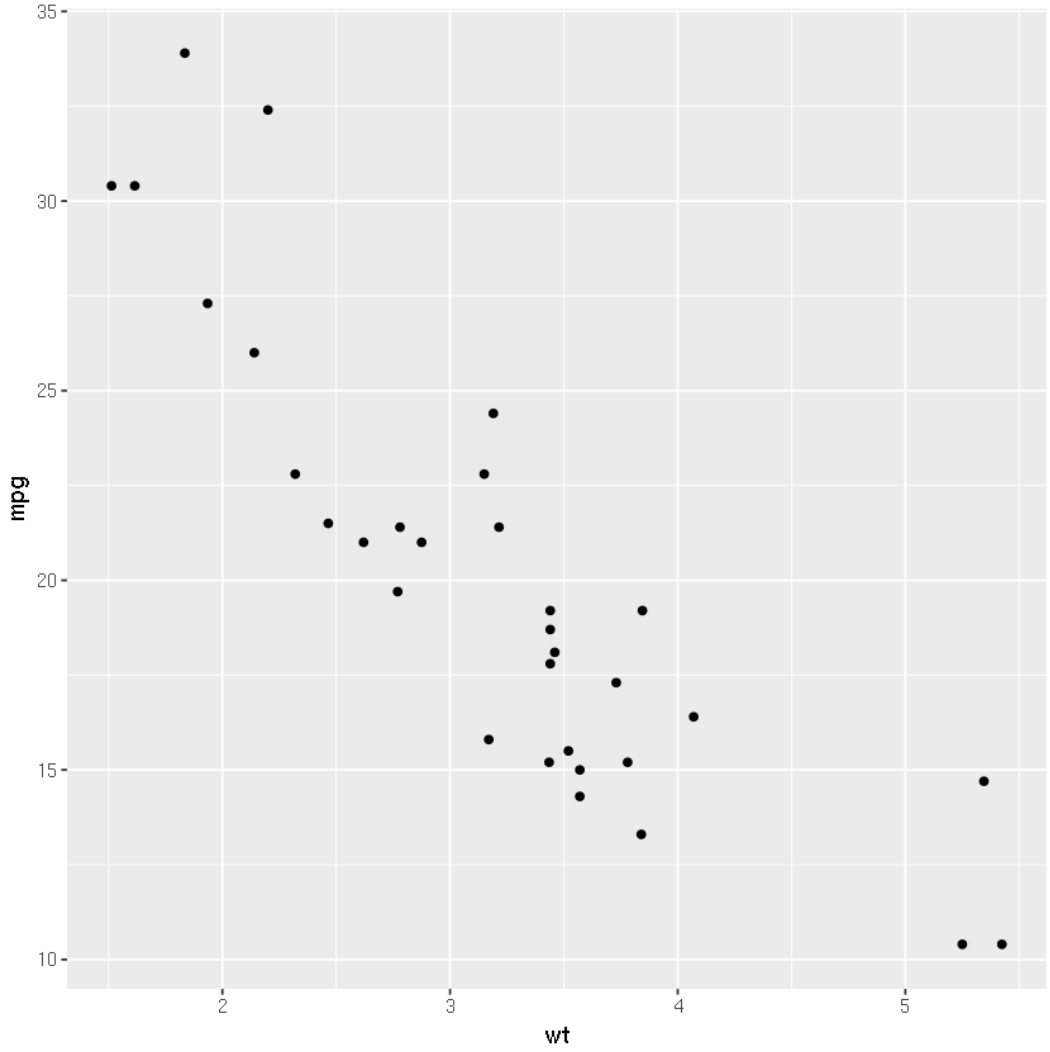


Tema

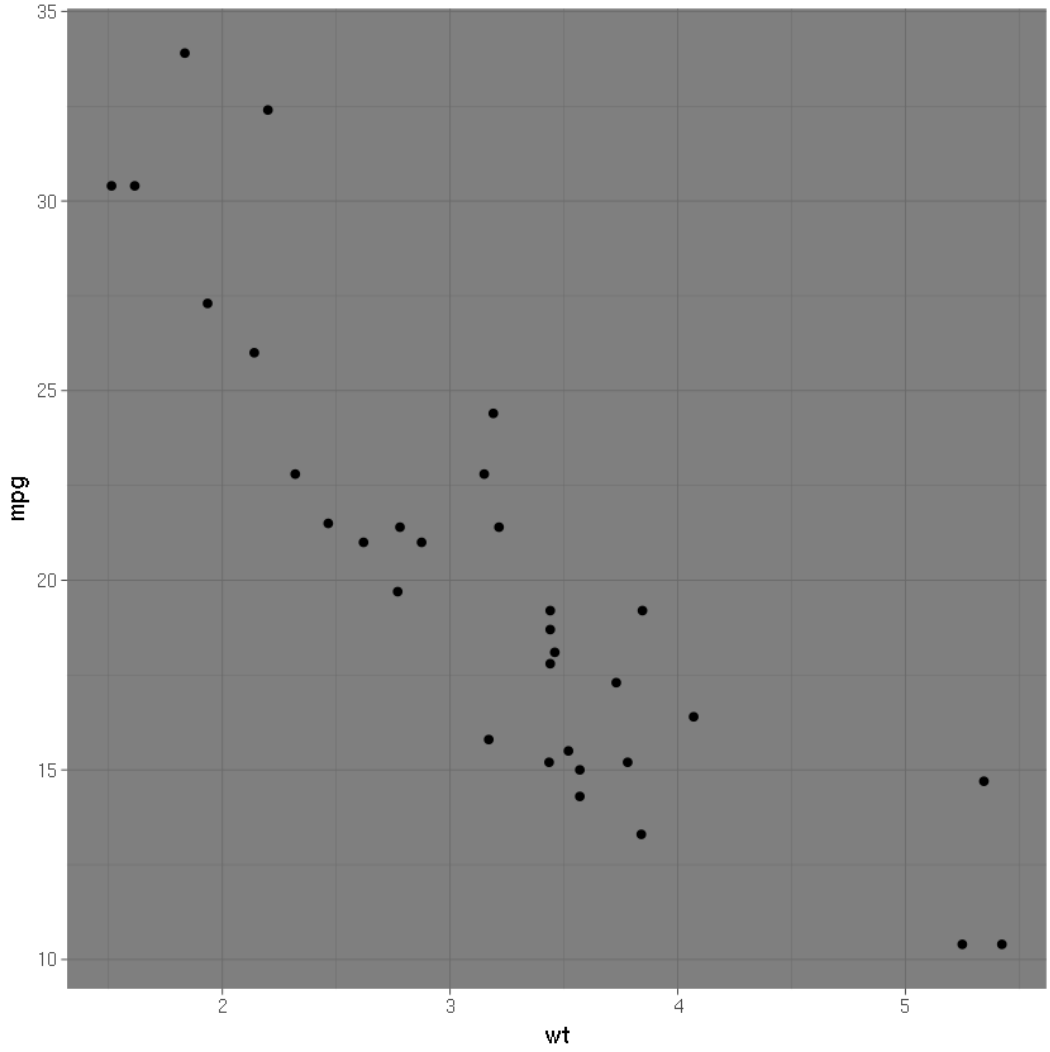
```
df <- mtcars
head(df)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

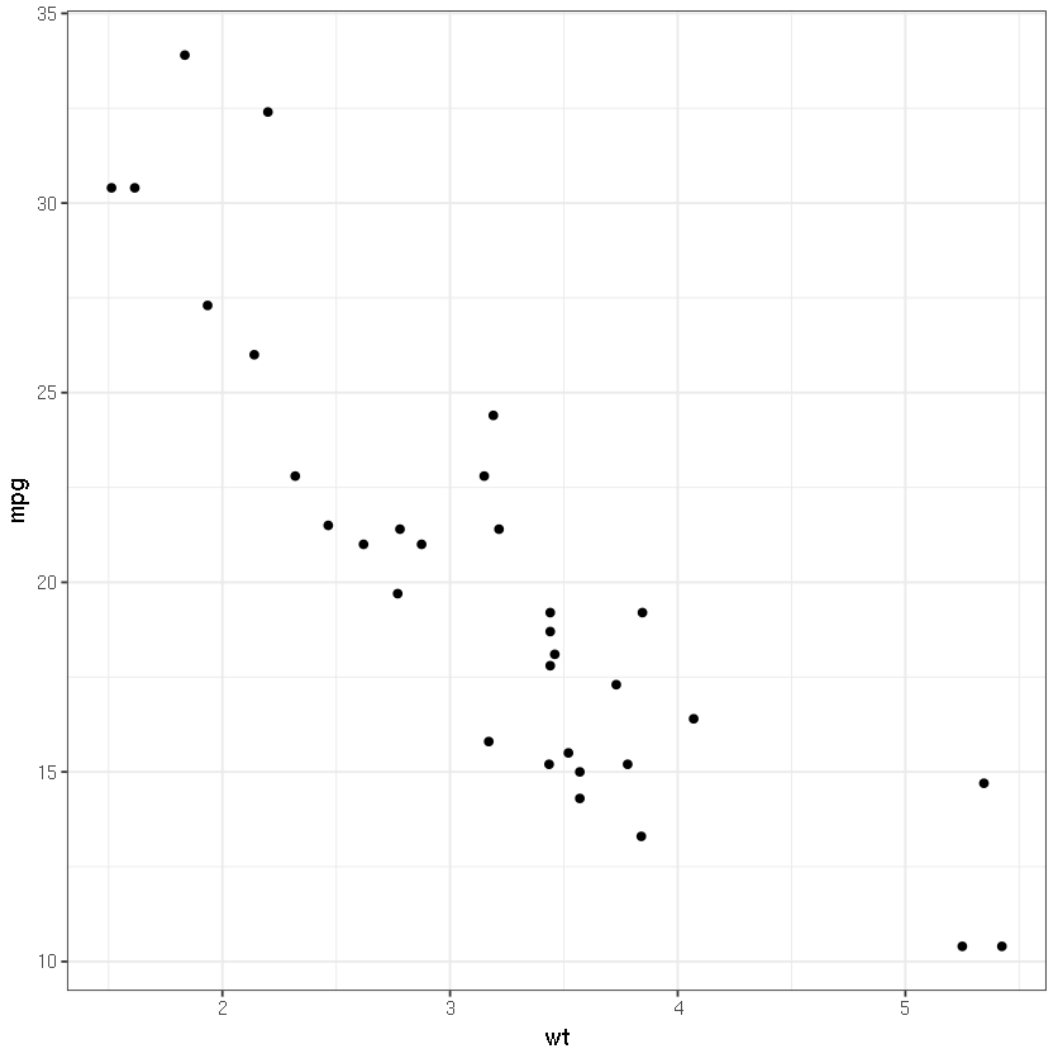
```
pl <- ggplot(df, aes(x=wt, y = mpg)) + geom_point()
pl
```




```
theme_set(theme_dark()) # mengatur tema untuk seluruh plot di dalam script
pl
```



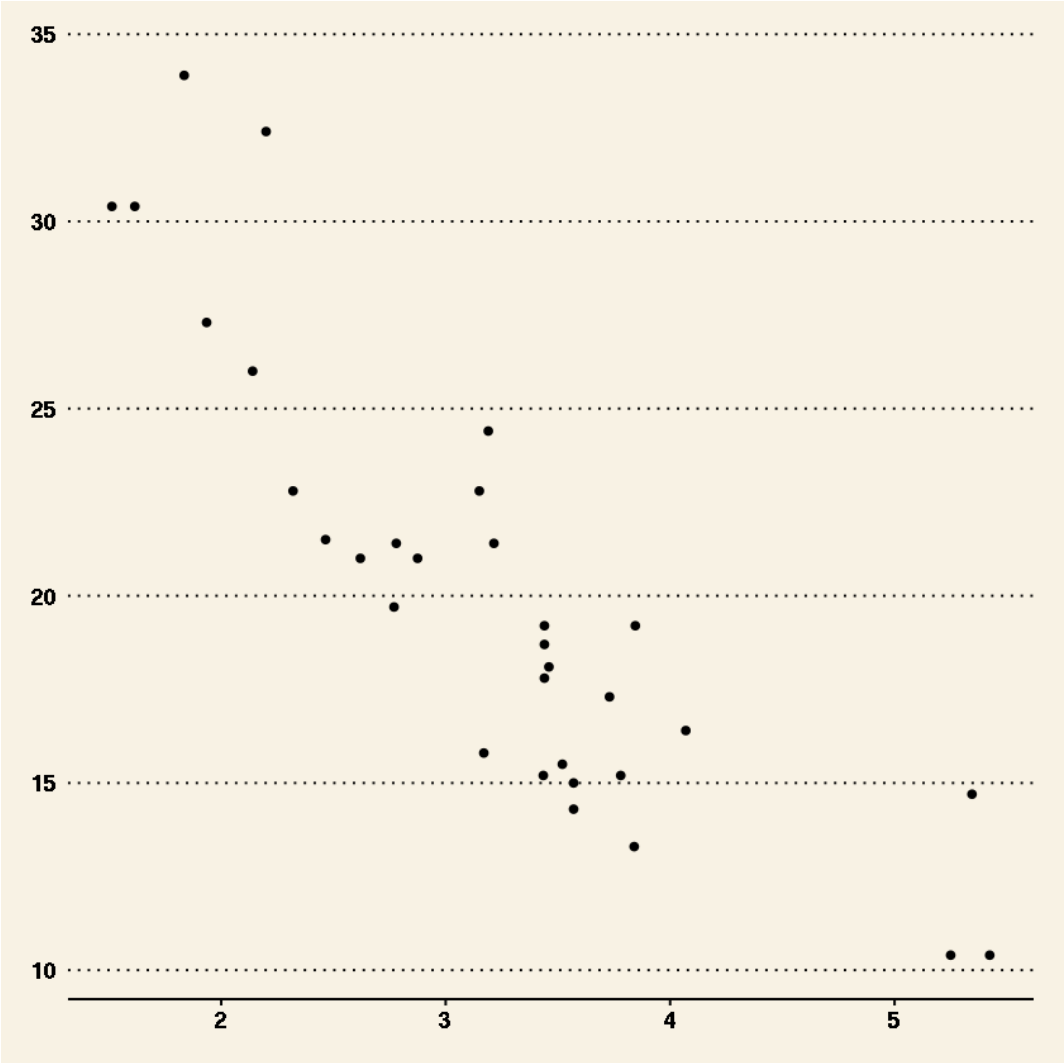

```
pl + theme_bw()
```



Untuk tema tambahan kita dapat menjalankan perintah sebagai berikut:

```
library(ggthemes)
```

```
pl + theme_wsj() # tema dari Wall Street Journal
```


35
30
25
20
15
10
2
3
4
5