Data Frame

Pengenalan

Kita banyak menggunakan data frame di dalam kegiatan analisis data, karena matriks hanya mampu menampung tipe data yang seragam.

```
state.x77 # built-in df
```

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost	Area
Alabama	3615	3624	2.1	69.05	15.1	41.3	20	50708
Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
Arkansas	2110	3378	1.9	70.66	10.1	39.9	65	51945
California	21198	5114	1.1	71.71	10.3	62.6	20	156361
Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766
Connecticut	3100	5348	1.1	72.48	3.1	56.0	139	4862
Delaware	579	4809	0.9	70.06	6.2	54.6	103	1982
Florida	8277	4815	1.3	70.66	10.7	52.6	11	54090
Georgia	4931	4091	2.0	68.54	13.9	40.6	60	58073
Hawaii	868	4963	1.9	73.60	6.2	61.9	0	6425
Idaho	813	4119	0.6	71.87	5.3	59.5	126	82677
Illinois	11197	5107	0.9	70.14	10.3	52.6	127	55748
Indiana	5313	4458	0.7	70.88	7.1	52.9	122	36097
Iowa	2861	4628	0.5	72.56	2.3	59.0	140	55941
Kansas	2280	4669	0.6	72.58	4.5	59.9	114	81787
Kentucky	3387	3712	1.6	70.10	10.6	38.5	95	39650
Louisiana	3806	3545	2.8	68.76	13.2	42.2	12	44930
Maine	1058	3694	0.7	70.39	2.7	54.7	161	30920
Maryland	4122	5299	0.9	70.22	8.5	52.3	101	9891
Massachusetts	5814	4755	1.1	71.83	3.3	58.5	103	7826
Michigan	9111	4751	0.9	70.63	11.1	52.8	125	56817
Minnesota	3921	4675	0.6	72.96	2.3	57.6	160	79289
Mississippi	2341	3098	2.4	68.09	12.5	41.0	50	47296
Missouri	4767	4254	0.8	70.69	9.3	48.8	108	68995
Montana	746	4347	0.6	70.56	5.0	59.2	155	145587
Nebraska	1544	4508	0.6	72.60	2.9	59.3	139	76483
Nevada	590	5149	0.5	69.03	11.5	65.2	188	109889
New Hampshire	812	4281	0.7	71.23	3.3	57.6	174	9027
New Jersey	7333	5237	1.1	70.93	5.2	52.5	115	7521
New Mexico	1144	3601	2.2	70.32	9.7	55.2	120	121412
New York	18076	4903	1.4	70.55	10.9	52.7	82	47831

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost	Area
North Carolina	5441	3875	1.8	69.21	11.1	38.5	80	48798
North Dakota	637	5087	0.8	72.78	1.4	50.3	186	69273
Ohio	10735	4561	0.8	70.82	7.4	53.2	124	40975
Oklahoma	2715	3983	1.1	71.42	6.4	51.6	82	68782
Oregon	2284	4660	0.6	72.13	4.2	60.0	44	96184
Pennsylvania	11860	4449	1.0	70.43	6.1	50.2	126	44966
Rhode Island	931	4558	1.3	71.90	2.4	46.4	127	1049
South Carolina	2816	3635	2.3	67.96	11.6	37.8	65	30225
South Dakota	681	4167	0.5	72.08	1.7	53.3	172	75955
Tennessee	4173	3821	1.7	70.11	11.0	41.8	70	41328
Texas	12237	4188	2.2	70.90	12.2	47.4	35	262134
Utah	1203	4022	0.6	72.90	4.5	67.3	137	82096
Vermont	472	3907	0.6	71.64	5.5	57.1	168	9267
Virginia	4981	4701	1.4	70.08	9.5	47.8	85	39780
Washington	3559	4864	0.6	71.72	4.3	63.5	32	66570
West Virginia	1799	3617	1.4	69.48	6.7	41.6	100	24070
Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149	54464
Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203

```
USPersonalExpenditure
```

	1940	1945	1950	1955	1960
Food and Tobacco	22.200	44.500	59.60	73.2	86.80
Household Operation	10.500	15.500	29.00	36.5	46.20
Medical and Health	3.530	5.760	9.71	14.0	21.10
Personal Care	1.040	1.980	2.45	3.4	5.40
Private Education	0.341	0.974	1.80	2.6	3.64

```
# Mengetahui daftar built-in df
# data()
```

```
head(state.x77) # 6 baris pertama
```

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost	Area
Alabama	3615	3624	2.1	69.05	15.1	41.3	20	50708
Alaska	365	6315	1.5	69.31	11.3	66.7	152	566432
Arizona	2212	4530	1.8	70.55	7.8	58.1	15	113417
Arkansas	2110	3378	1.9	70.66	10.1	39.9	65	51945
California	21198	5114	1.1	71.71	10.3	62.6	20	156361
Colorado	2541	4884	0.7	72.06	6.8	63.9	166	103766

```
tail(state.x77) # 6 baris terakhir
```

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost	Area
Vermont	472	3907	0.6	71.64	5.5	57.1	168	9267
Virginia	4981	4701	1.4	70.08	9.5	47.8	85	39780
Washington	3559	4864	0.6	71.72	4.3	63.5	32	66570
West Virginia	1799	3617	1.4	69.48	6.7	41.6	100	24070
Wisconsin	4589	4468	0.7	72.48	3.0	54.5	149	54464
Wyoming	376	4566	0.6	70.29	6.9	62.9	173	97203

```
str(state.x77) # struktur dari df
```

```
 num [1:50, 1:8] 3615 365 2212 2110 21198 ...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
  ..$ : chr [1:8] "Population" "Income" "Illiteracy" "Life Exp" ...
```

```
summary(state.x77) # sari statistik dari df berdasarkan kolom
```

```
   Population        Income       Illiteracy       Life Exp
 Min.   :  365   Min.   :3098   Min.   :0.500   Min.   :67.96
 1st Qu.: 1080   1st Qu.:3993   1st Qu.:0.625   1st Qu.:70.12
 Median : 2838   Median :4519   Median :0.950   Median :70.67
 Mean   : 4246   Mean   :4436   Mean   :1.170   Mean   :70.88
 3rd Qu.: 4968   3rd Qu.:4814   3rd Qu.:1.575   3rd Qu.:71.89
```

```
 Max.   :21198   Max.   :6315   Max.   :2.800   Max.   :73.60
     Murder          HS Grad          Frost             Area
 Min.   : 1.400   Min.   :37.80   Min.   :  0.00   Min.   :  1049
 1st Qu.: 4.350   1st Qu.:48.05   1st Qu.: 66.25   1st Qu.: 36985
 Median : 6.850   Median :53.25   Median :114.50   Median : 54277
 Mean   : 7.378   Mean   :53.11   Mean   :104.46   Mean   : 70736
 3rd Qu.:10.675   3rd Qu.:59.15   3rd Qu.:139.75   3rd Qu.: 81162
 Max.   :15.100   Max.   :67.30   Max.   :188.00   Max.   :566432
```

```
# mendefinisikan df
nama <- c('Agus', 'Sugio', 'Bayu', 'Atmo', 'Roy')
umur <- c(42,35,37,28,27)
kawin <- c(F,F,T,F,T)
```

```
data.frame(nama, umur, kawin)
```

nama	umur	kawin
Agus	42	FALSE
Sugio	35	FALSE
Bayu	37	TRUE
Atmo	28	FALSE
Roy	27	TRUE

```
df <- data.frame(nama, umur, kawin)
df
```

nama	umur	kawin
Agus	42	FALSE
Sugio	35	FALSE
Bayu	37	TRUE
Atmo	28	FALSE
Roy	27	TRUE

```
str(df)
```

```
'data.frame':   5 obs. of  3 variables:
 $ nama : Factor w/ 5 levels "Agus","Atmo",..: 1 5 3 2 4
 $ umur : num  42 35 37 28 27
 $ kawin: logi  FALSE FALSE TRUE FALSE TRUE
```

```
summary(df)
```

```
    nama        umur          kawin
 Agus :1   Min.   :27.0   Mode :logical
 Atmo :1   1st Qu.:28.0   FALSE:3
 Bayu :1   Median :35.0   TRUE :2
 Roy  :1   Mean   :33.8
 Sugio:1   3rd Qu.:37.0
           Max.   :42.0
```

Seleksi dan pengindeksan Data Frame

```
df
```

nama	umur	kawin
Agus	42	FALSE
Sugio	35	FALSE
Bayu	37	TRUE
Atmo	28	FALSE
Roy	27	TRUE

```
df[3, ] # ambil baris ketiga
```

	nama	umur	kawin
3	Bayu	37	TRUE

```
df[,1]
```

1. Agus
2. Sugio
3. Bayu
4. Atmo
5. Roy

▶ **Levels**:

```
df[,'nama']
```

1. Agus
2. Sugio

3. Bayu
4. Atmo
5. Roy

▶ **Levels**:

```
df[1:4, c('nama', 'umur')]
```

nama	umur
Agus	42
Sugio	35
Bayu	37
Atmo	28

```
df$umur
```

1. 42
2. 35
3. 37
4. 28
5. 27

```
df[,'umur']
```

1. 42
2. 35
3. 37
4. 28
5. 27

```
# fungsi subset
df
```

nama	umur	kawin
Agus	42	FALSE
Sugio	35	FALSE
Bayu	37	TRUE
Atmo	28	FALSE
Roy	27	TRUE

```
subset(df,subset = kawin == T)
```

	nama	umur	kawin
3	Bayu	37	TRUE
5	Roy	27	TRUE

```
subset(df, subset = umur > 30)
```

nama	umur	kawin
Agus	42	FALSE
Sugio	35	FALSE
Bayu	37	TRUE

```
# Mengurutkan dataframe
urut.umur <- order(df['umur'])
urut.umur
```

1. 5
2. 4
3. 2
4. 3
5. 1

```
df[urut.umur,]
```

	nama	umur	kawin
5	Roy	27	TRUE
4	Atmo	28	FALSE
2	Sugio	35	FALSE
3	Bayu	37	TRUE
1	Agus	42	FALSE

```
umur.terbalik <- order(-df['umur'])
umur.terbalik
```

1. 1
2. 3
3. 2
4. 4
5. 5

```
df[umur.terbalik,]
```

	nama	umur	kawin
1	Agus	42	FALSE
3	Bayu	37	TRUE
2	Sugio	35	FALSE
4	Atmo	28	FALSE
5	Roy	27	TRUE

```
urut.umur <- order(df$umur)
df[urut.umur, ]
```

	nama	umur	kawin
5	Roy	27	TRUE
4	Atmo	28	FALSE
2	Sugio	35	FALSE
3	Bayu	37	TRUE
1	Agus	42	FALSE

Operasi - operasi data frame

Mendefinisikan data frame

```
c1 <- 1:10
c2 <- letters[1:10]
print(c1)
print(c2)
```

```
 [1]  1  2  3  4  5  6  7  8  9 10
 [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j"
```

```
df <- data.frame(c1,c2)
df
```

c1	c2
1	a
2	b
3	c
4	d
5	e
6	f
7	g
8	h
9	i
10	j

```
df <- data.frame(kolom1 = c1, kolom2 = c2) # nama kolom bisa kita ubah sesuka kita!
df
```

kolom1	kolom2
1	a
2	b
3	c
4	d
5	e
6	f
7	g
8	h
9	i
10	j

Mendapatkan info tentang data frame

```
nrow(df) # jumlah baris
```

10

```
ncol(df) # jumlah kolom
```

2

```
colnames(df)  # nama kolom
```

1. 'kolom1'
2. 'kolom2'

```
rownames(df) # nama baris
```

1. '1'
2. '2'
3. '3'
4. '4'
5. '5'
6. '6'
7. '7'
8. '8'
9. '9'
10. '10'

```
str(df) # struktur data frame
```

```
'data.frame':   10 obs. of  2 variables:
 $ kolom1: int  1 2 3 4 5 6 7 8 9 10
 $ kolom2: Factor w/ 10 levels "a","b","c","d",..: 1 2 3 4 5 6 7 8 9 10
```

```
summary(df) # sari statistik
```

```
     kolom1          kolom2
 Min.   : 1.00   a      :1
 1st Qu.: 3.25   b      :1
 Median : 5.50   c      :1
 Mean   : 5.50   d      :1
 3rd Qu.: 7.75   e      :1
 Max.   :10.00   f      :1
                 (Other):4
```

Referensi sel

```
df
```

kolom1	kolom2
1	a
2	b
3	c
4	d
5	e
6	f
7	g
8	h
9	i
10	j

```
df[5,1]
```

5

```
df[[5,1]]
```

5

```
df[1,'kolom1']
```

1

```
df[5, 'kolom1']
```

5

```
df[[5,'kolom1']]
```

5

```
df[8,'kolom1'] <- -999 # mengubah nilai
df
```

kolom1	kolom2
1	a
2	b
3	c
4	d
5	e
6	f
7	g
-999	h
9	i
10	j

```
df[[8,'kolom1']] <- 8
df
```

kolom1	kolom2
1	a
2	b
3	c
4	d
5	e
6	f
7	g
8	h
9	i
10	j

Referensi baris dan kolom

```
df[2,]
```

	kolom1	kolom2
2	2	b

```
df[1:3,]
```

kolom1	kolom2
1	a
2	b
3	c

```
head(mtcars)
```

	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

```
mtcars$mpg
```

1. 21
2. 21
3. 22.8
4. 21.4
5. 18.7
6. 18.1
7. 14.3
8. 24.4
9. 22.8
10. 19.2
11. 17.8
12. 16.4
13. 17.3
14. 15.2
15. 10.4
16. 10.4
17. 14.7
18. 32.4
19. 30.4
20. 33.9
21. 21.5
22. 15.5
23. 15.2
24. 13.3
25. 19.2
26. 27.3
27. 26
28. 30.4
29. 15.8
30. 19.7
31. 15
32. 21.4

```
mtcars[,1]
```

1. 21
2. 21
3. 22.8
4. 21.4
5. 18.7
6. 18.1
7. 14.3
8. 24.4
9. 22.8
10. 19.2
11. 17.8
12. 16.4
13. 17.3
14. 15.2
15. 10.4
16. 10.4
17. 14.7
18. 32.4
19. 30.4
20. 33.9
21. 21.5
22. 15.5
23. 15.2
24. 13.3
25. 19.2
26. 27.3
27. 26
28. 30.4
29. 15.8
30. 19.7
31. 15
32. 21.4

```
mtcars['mpg']
```

1. 21
2. 21
3. 22.8
4. 21.4
5. 18.7
6. 18.1
7. 14.3
8. 24.4
9. 22.8
10. 19.2
11. 17.8
12. 16.4
13. 17.3
14. 15.2
15. 10.4
16. 10.4

17. 14.7
18. 32.4
19. 30.4
20. 33.9
21. 21.5
22. 15.5
23. 15.2
24. 13.3
25. 19.2
26. 27.3
27. 26
28. 30.4
29. 15.8
30. 19.7
31. 15
32. 21.4

```
mtcars[['mpg']]
```

1. 21
2. 21
3. 22.8
4. 21.4
5. 18.7
6. 18.1
7. 14.3
8. 24.4
9. 22.8
10. 19.2
11. 17.8
12. 16.4
13. 17.3
14. 15.2
15. 10.4
16. 10.4
17. 14.7
18. 32.4
19. 30.4
20. 33.9
21. 21.5
22. 15.5
23. 15.2
24. 13.3
25. 19.2
26. 27.3
27. 26
28. 30.4
29. 15.8
30. 19.7
31. 15
32. 21.4

```r
mtcars[1] # indeks lokasi kolom:1, cara ini mereferensi kolom mpg sbg data frame
```

	mpg
Mazda RX4	21.0
Mazda RX4 Wag	21.0
Datsun 710	22.8
Hornet 4 Drive	21.4
Hornet Sportabout	18.7
Valiant	18.1
Duster 360	14.3
Merc 240D	24.4
Merc 230	22.8
Merc 280	19.2
Merc 280C	17.8
Merc 450SE	16.4
Merc 450SL	17.3
Merc 450SLC	15.2
Cadillac Fleetwood	10.4
Lincoln Continental	10.4
Chrysler Imperial	14.7
Fiat 128	32.4
Honda Civic	30.4
Toyota Corolla	33.9
Toyota Corona	21.5
Dodge Challenger	15.5
AMC Javelin	15.2
Camaro Z28	13.3
Pontiac Firebird	19.2
Fiat X1-9	27.3
Porsche 914-2	26.0
Lotus Europa	30.4
Ford Pantera L	15.8
Ferrari Dino	19.7
Maserati Bora	15.0
Volvo 142E	21.4

```
head(mtcars)
```

	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

```
mtcars[c('mpg', 'cyl')]
```

	mpg	cyl
Mazda RX4	21.0	6
Mazda RX4 Wag	21.0	6
Datsun 710	22.8	4
Hornet 4 Drive	21.4	6
Hornet Sportabout	18.7	8
Valiant	18.1	6
Duster 360	14.3	8
Merc 240D	24.4	4
Merc 230	22.8	4
Merc 280	19.2	6
Merc 280C	17.8	6
Merc 450SE	16.4	8
Merc 450SL	17.3	8
Merc 450SLC	15.2	8
Cadillac Fleetwood	10.4	8
Lincoln Continental	10.4	8
Chrysler Imperial	14.7	8
Fiat 128	32.4	4
Honda Civic	30.4	4
Toyota Corolla	33.9	4
Toyota Corona	21.5	4
Dodge Challenger	15.5	8
AMC Javelin	15.2	8
Camaro Z28	13.3	8
Pontiac Firebird	19.2	8
Fiat X1-9	27.3	4
Porsche 914-2	26.0	4
Lotus Europa	30.4	4
Ford Pantera L	15.8	8
Ferrari Dino	19.7	6
Maserati Bora	15.0	8
Volvo 142E	21.4	4

Menambahkan baris dan kolom

```
c1 <- c(10,20,30,40,50)
c2 <- letters[c(1:5)]
df <- data.frame(kol1=c1,kol2=c2)
df
```

kol1	kol2
10	a
20	b
30	c
40	d
50	e

```
df1 <- data.frame(kol1=128,kol2='Meteorologi')
df1
```

kol1	kol2
128	Meteorologi

```
# menambahkan df1 ke df
df <- rbind(df,df1)
df
```

kol1	kol2
10	a
20	b
30	c
40	d
50	e
128	Meteorologi

```
c3 <- c(11:16)
c3
```

1. 11
2. 12

3. 13
4. 14
5. 15
6. 16

```
# menambahkan kolom ke df
df <- cbind(df,kol3=c3)
df
```

kol1	kol2	kol3
10	a	11
20	b	12
30	c	13
40	d	14
50	e	15
128	Meteorologi	16

```
df$kol4 <- c(20,25,30,35,40,45)
df
```

kol1	kol2	kol3	kol4
10	a	11	20
20	b	12	25
30	c	13	30
40	d	14	35
50	e	15	40
128	Meteorologi	16	45

```
df$kol5 <- df$kol1 * 2
df
```

kol1	kol2	kol3	kol4	kol5
10	a	11	20	20
20	b	12	25	40
30	c	13	30	60
40	d	14	35	80
50	e	15	40	100
128	Meteorologi	16	45	256

Mengatur penamaan kolom

```
colnames(df) # mengetahui nama kolom
```

1. 'kol1'
2. 'kol2'
3. 'kol3'
4. 'kol4'
5. 'kol5'

```
colnames(df) <- c('A', 'B', 'C', 'D', 'E') # penamaan ulang nama kolom
df
```

A	B	C	D	E
10	a	11	20	20
20	b	12	25	40
30	c	13	30	60
40	d	14	35	80
50	e	15	40	100
128	Meteorologi	16	45	256

```
colnames(df)[1] <- 'X' # penamaan ulang kolom secara individual
df
```

X	B	C	D	E
10	a	11	20	20
20	b	12	25	40
30	c	13	30	60
40	d	14	35	80
50	e	15	40	100
128	Meteorologi	16	45	256

```
colnames(df)[c(2,3)] <- c('Y','Z')
df
```

X	Y	Z	D	E
10	a	11	20	20
20	b	12	25	40
30	c	13	30	60
40	d	14	35	80
50	e	15	40	100
128	Meteorologi	16	45	256

Menyeleksi banyak baris dan kolom

```
df[1:3,]
```

X	Y	Z	D	E
10	a	11	20	20
20	b	12	25	40
30	c	13	30	60

```
# menyeleksi seluruh baris, kecuali no 3
df[-3,]
```

	X	Y	Z	D	E
1	10	a	11	20	20
2	20	b	12	25	40
4	40	d	14	35	80
5	50	e	15	40	100
6	128	Meteorologi	16	45	256

```
# penyeleksian kondisional
head(mtcars)
```

	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

```
mtcars[mtcars$mpg > 20,]
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

```
mtcars[mtcars$mpg > 20 & mtcars$cyl == 6,]
```

	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
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1

```
mtcars[mtcars$mpg > 20 & mtcars$cyl == 6, c('mpg', 'cyl','hp')]
```

	mpg	cyl	hp
Mazda RX4	21.0	6	110
Mazda RX4 Wag	21.0	6	110
Hornet 4 Drive	21.4	6	110

```
subset(mtcars, mpg > 20 & cyl == 6) # pakai built-in subset function
```

	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
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1

```
df[,1:3]
```

X	Y	Z
10	a	11
20	b	12
30	c	13
40	d	14
50	e	15
128	Meteorologi	16

```
df[,c(3,5)]
```

Z	E
11	20
12	40
13	60
14	80
15	100
16	256

```
df[,c('Z', 'E')]
```

Z	E
11	20
12	40
13	60
14	80
15	100
16	256

Menangani data kosong

```
df
```

X	Y	Z	D	E
10	a	11	20	20
20	b	12	25	40
30	c	13	30	60
40	d	14	35	80
50	e	15	40	100
128	Meteorologi	16	45	256

```
is.na(df) # cek data kosong
```

X	Y	Z	D	E
FALSE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	FALSE

```
any(is.na(df))
```

FALSE

```
any(is.na(df$X))
```

FALSE

```
# Membuat na jadi 0
any(is.na(df)) <- 0.5 # error karena ga ada na
```

```
Error in any(is.na(df)) <- 0.5: could not find function "any<-"
Traceback:
```

```
any(is.na(df$D)) <- mean(df$D)  # error karena ga ada na
```

```
Error in any(is.na(df$D)) <- mean(df$D): could not find function "any<-"
Traceback:
```