Manipulasi data

dplyr

```
library(dplyr)
library(nycflights13)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
head(flights)
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	1	1	517	515	2	830	819	11	UA
2013	1	1	533	529	4	850	830	20	UA
2013	1	1	542	540	2	923	850	33	AA
2013	1	1	544	545	-1	1004	1022	-18	B6
2013	1	1	554	600	-6	812	837	-25	DL
2013	1	1	554	558	-4	740	728	12	UA

`filter()`

```
head(filter(flights,month == 5, day == 2, carrier ==
'AA'))
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	5	2	539	540	-1	850	840	10	AA
2013	5	2	549	600	-11	823	850	-27	AA
2013	5	2	558	605	-7	855	910	-15	AA
2013	5	2	603	610	-7	729	745	-16	AA
2013	5	2	611	615	-4	900	915	-15	AA
2013	5	2	627	630	-3	736	805	-29	AA

```
head(flights[flights$month == 5 & flights$day == 2 &
flights$carrier == 'AA',]) # ribet
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	5	2	539	540	-1	850	840	10	AA

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	5	2	549	600	-11	823	850	-27	AA
2013	5	2	558	605	-7	855	910	-15	AA
2013	5	2	603	610	-7	729	745	-16	AA
2013	5	2	611	615	-4	900	915	-15	AA
2013	5	2	627	630	-3	736	805	-29	AA

`slice()`

```
slice(flights, 1:10) # menyeleksi 10 baris pertama
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	1	1	517	515	2	830	819	11	UA
2013	1	1	533	529	4	850	830	20	UA
2013	1	1	542	540	2	923	850	33	AA
2013	1	1	544	545	-1	1004	1022	-18	B6
2013	1	1	554	600	-6	812	837	-25	DL
2013	1	1	554	558	-4	740	728	12	UA
2013	1	1	555	600	-5	913	854	19	B6
2013	1	1	557	600	-3	709	723	-14	EV
2013	1	1	557	600	-3	838	846	-8	B6
2013	1	1	558	600	-2	753	745	8	AA

`arrange()`

```
head(arrange(flights, year, month, day, arr_time))
# mengatur urutan sesuai kolomnya
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	1	1	1929	1920	9	3	7	-4	UA
2013	1	1	2121	2040	41	6	2323	43	B6
2013	1	1	2058	2100	-2	8	2359	9	UA
2013	1	1	2120	2130	-10	16	18	-2	B6
2013	1	1	2134	2045	49	20	2352	28	UA
2013	1	1	2312	2000	192	21	2110	191	EV

`select()`

```
head(select(flights, arr_time)) # seleksi kolom arr_time
```

ARR_TIME

ARR_TIME
830
850
923
1004
812
740

```
head(select(flights,carrier)) # seleksi kolom carrier
```

CARRIER
UA
UA
AA
B6
DL
UA

```
head(select(flights, arr_time, carrier, month)) # seleksi
3 kolom
```

ARR_TIME	CARRIER	MONTH
830	UA	1
850	UA	1
923	AA	1
1004	B6	1
812	DL	1
740	UA	1

`rename()`:

```
head(flights)
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	1	1	517	515	2	830	819	11	UA
2013	1	1	533	529	4	850	830	20	UA
2013	1	1	542	540	2	923	850	33	AA
2013	1	1	544	545	-1	1004	1022	-18	B6
2013	1	1	554	600	-6	812	837	-25	DL
2013	1	1	554	558	-4	740	728	12	UA

```
rename(flights,new_arr_time = arr_time) # mengubah nama
kolom
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	NEW_ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CAR
2013	1	1	517	515	2	830	819	11	UA

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	NEW_ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CAR
2013	1	1	533	529	4	850	830	20	UA
2013	1	1	542	540	2	923	850	33	AA
2013	1	1	544	545	-1	1004	1022	-18	B6
2013	1	1	554	600	-6	812	837	-25	DL
2013	1	1	554	558	-4	740	728	12	UA
2013	1	1	555	600	-5	913	854	19	B6
2013	1	1	557	600	-3	709	723	-14	EV
2013	1	1	557	600	-3	838	846	-8	B6
2013	1	1	558	600	-2	753	745	8	AA
2013	1	1	558	600	-2	849	851	-2	B6
2013	1	1	558	600	-2	853	856	-3	B6
2013	1	1	558	600	-2	924	917	7	UA
2013	1	1	558	600	-2	923	937	-14	UA
2013	1	1	559	600	-1	941	910	31	AA
2013	1	1	559	559	0	702	706	-4	B6
2013	1	1	559	600	-1	854	902	-8	UA
2013	1	1	600	600	0	851	858	-7	B6
2013	1	1	600	600	0	837	825	12	MQ
2013	1	1	601	600	1	844	850	-6	B6
2013	1	1	602	610	-8	812	820	-8	DL
2013	1	1	602	605	-3	821	805	16	MQ
2013	1	1	606	610	-4	858	910	-12	AA
2013	1	1	606	610	-4	837	845	-8	DL
2013	1	1	607	607	0	858	915	-17	UA
2013	1	1	608	600	8	807	735	32	MQ
2013	1	1	611	600	11	945	931	14	UA
2013	1	1	613	610	3	925	921	4	B6
2013	1	1	615	615	0	1039	1100	-21	B6
2013	1	1	615	615	0	833	842	-9	DL
...	...	...	...	...	...	...	...	...	...
2013	9	30	2123	2125	-2	2223	2247	-24	EV
2013	9	30	2127	2129	-2	2314	2323	-9	EV
2013	9	30	2128	2130	-2	2328	2359	-31	B6
2013	9	30	2129	2059	30	2230	2232	-2	EV
2013	9	30	2131	2140	-9	2225	2255	-30	MQ

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	NEW_ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CAR
2013	9	30	2140	2140	0	10	40	-30	AA
2013	9	30	2142	2129	13	2250	2239	11	EV
2013	9	30	2145	2145	0	115	140	-25	B6
2013	9	30	2147	2137	10	30	27	3	B6
2013	9	30	2149	2156	-7	2245	2308	-23	UA
2013	9	30	2150	2159	-9	2250	2306	-16	EV
2013	9	30	2159	1845	194	2344	2030	194	9E
2013	9	30	2203	2205	-2	2339	2331	8	EV
2013	9	30	2207	2140	27	2257	2250	7	MQ
2013	9	30	2211	2059	72	2339	2242	57	EV
2013	9	30	2231	2245	-14	2335	2356	-21	B6
2013	9	30	2233	2113	80	112	30	42	UA
2013	9	30	2235	2001	154	59	2249	130	B6
2013	9	30	2237	2245	-8	2345	2353	-8	B6
2013	9	30	2240	2245	-5	2334	2351	-17	B6
2013	9	30	2240	2250	-10	2347	7	-20	B6
2013	9	30	2241	2246	-5	2345	1	-16	B6
2013	9	30	2307	2255	12	2359	2358	1	B6
2013	9	30	2349	2359	-10	325	350	-25	B6
2013	9	30	NA	1842	NA	NA	2019	NA	EV
2013	9	30	NA	1455	NA	NA	1634	NA	9E
2013	9	30	NA	2200	NA	NA	2312	NA	9E
2013	9	30	NA	1210	NA	NA	1330	NA	MQ
2013	9	30	NA	1159	NA	NA	1344	NA	MQ
2013	9	30	NA	840	NA	NA	1020	NA	MQ

`distinct()`

Untuk menyeleksi nilai - nilai unik

```
distinct(select(flights, carrier)) # nilai - nilai unik
pada kolom carrier
```

CARRIER
UA
AA
B6
DL
EV
MQ
US

CARRIER
WN
VX
FL
AS
9E
F9
HA
YV
OO

```
distinct(select(flights,month))
```

MONTH
1
10
11
12
2
3
4
5
6
7
8
9

```
mutate()
```

Menambahkan kolom baru di data frame

```
mutate(flights, kol_baru = arr_delay - dep_delay)
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	1	1	517	515	2	830	819	11	UA
2013	1	1	533	529	4	850	830	20	UA
2013	1	1	542	540	2	923	850	33	AA
2013	1	1	544	545	-1	1004	1022	-18	B6
2013	1	1	554	600	-6	812	837	-25	DL
2013	1	1	554	558	-4	740	728	12	UA
2013	1	1	555	600	-5	913	854	19	B6
2013	1	1	557	600	-3	709	723	-14	EV
2013	1	1	557	600	-3	838	846	-8	B6
2013	1	1	558	600	-2	753	745	8	AA
2013	1	1	558	600	-2	849	851	-2	B6
2013	1	1	558	600	-2	853	856	-3	B6
2013	1	1	558	600	-2	924	917	7	UA

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	1	1	558	600	-2	923	937	-14	UA
2013	1	1	559	600	-1	941	910	31	AA
2013	1	1	559	559	0	702	706	-4	B6
2013	1	1	559	600	-1	854	902	-8	UA
2013	1	1	600	600	0	851	858	-7	B6
2013	1	1	600	600	0	837	825	12	MQ
2013	1	1	601	600	1	844	850	-6	B6
2013	1	1	602	610	-8	812	820	-8	DL
2013	1	1	602	605	-3	821	805	16	MQ
2013	1	1	606	610	-4	858	910	-12	AA
2013	1	1	606	610	-4	837	845	-8	DL
2013	1	1	607	607	0	858	915	-17	UA
2013	1	1	608	600	8	807	735	32	MQ
2013	1	1	611	600	11	945	931	14	UA
2013	1	1	613	610	3	925	921	4	B6
2013	1	1	615	615	0	1039	1100	-21	B6
2013	1	1	615	615	0	833	842	-9	DL
...	...	...	...	...	...	...	...	...	...
2013	9	30	2123	2125	-2	2223	2247	-24	EV
2013	9	30	2127	2129	-2	2314	2323	-9	EV
2013	9	30	2128	2130	-2	2328	2359	-31	B6
2013	9	30	2129	2059	30	2230	2232	-2	EV
2013	9	30	2131	2140	-9	2225	2255	-30	MQ
2013	9	30	2140	2140	0	10	40	-30	AA
2013	9	30	2142	2129	13	2250	2239	11	EV
2013	9	30	2145	2145	0	115	140	-25	B6
2013	9	30	2147	2137	10	30	27	3	B6
2013	9	30	2149	2156	-7	2245	2308	-23	UA
2013	9	30	2150	2159	-9	2250	2306	-16	EV
2013	9	30	2159	1845	194	2344	2030	194	9E
2013	9	30	2203	2205	-2	2339	2331	8	EV
2013	9	30	2207	2140	27	2257	2250	7	MQ
2013	9	30	2211	2059	72	2339	2242	57	EV
2013	9	30	2231	2245	-14	2335	2356	-21	B6
2013	9	30	2233	2113	80	112	30	42	UA

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	9	30	2235	2001	154	59	2249	130	B6
2013	9	30	2237	2245	-8	2345	2353	-8	B6
2013	9	30	2240	2245	-5	2334	2351	-17	B6
2013	9	30	2240	2250	-10	2347	7	-20	B6
2013	9	30	2241	2246	-5	2345	1	-16	B6
2013	9	30	2307	2255	12	2359	2358	1	B6
2013	9	30	2349	2359	-10	325	350	-25	B6
2013	9	30	NA	1842	NA	NA	2019	NA	EV
2013	9	30	NA	1455	NA	NA	1634	NA	9E
2013	9	30	NA	2200	NA	NA	2312	NA	9E
2013	9	30	NA	1210	NA	NA	1330	NA	MQ
2013	9	30	NA	1159	NA	NA	1344	NA	MQ
2013	9	30	NA	840	NA	NA	1020	NA	MQ

`transmute()`

Sama seperti `mutate()`, namu hanya mengeluarkan *output* kolom baru yang dihasilkan.

```
transmute(flights, kol_baru = arr_delay - dep_delay)
```

KOL_BARU
9
16
31
-17
-19
16
24
-11
-5
10
0
-1
9
-12
32
-4
-7
-7
12
-7
0
19
-8
-4
-17
24
3
1

KOL_BARU
-21
-9
...
-22
-7
-29
-32
-21
-30
-2
-25
-7
-16
-7
0
10
-20
-15
-7
-38
-24
0
-12
-10
-11
-11
-15
NA
NA
NA
NA
NA
NA

`summarise()`

```
summarise(flights, rata2wktTerbang = mean(air_time, na.rm
= T))
# sama seperti fungsi agregate di R
```

RATA2WKTTERBANG
150.6865

```
summarise(flights, JmlhwktTerbang = sum(air_time, na.rm =
T))
```

JMLHWKTTERBANG
49326610

`sample_n()` dan `sample_frac()`

```
sample_n(flights, 3) # mensampel 3 baris acak
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	8	22	552	600	-8	647	700	-13	US
2013	4	5	1433	1345	48	1813	1700	73	AA
2013	10	29	836	835	1	1100	1050	10	EV

```
sample_frac(flights, 0.01) # mensampel 1 % dari seluruh
baris
```

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	8	3	745	752	-7	856	913	-17	B6
2013	10	19	653	700	-7	953	1003	-10	B6
2013	7	17	2107	2030	37	2224	2211	13	9E
2013	4	4	1059	930	89	1421	1255	86	UA
2013	5	13	1825	1829	-4	2042	2031	11	US
2013	9	15	620	625	-5	835	850	-15	MQ
2013	5	25	1609	1557	12	1930	1908	22	DL
2013	9	1	1550	1600	-10	1819	1849	-30	B6
2013	3	1	703	650	13	846	858	-12	EV
2013	5	30	906	900	6	1146	1210	-24	UA
2013	11	18	955	1000	-5	1326	1333	-7	DL
2013	4	11	1415	1415	0	1611	1610	1	MQ
2013	9	2	603	611	-8	714	722	-8	EV
2013	5	6	1052	1055	-3	1156	1228	-32	UA
2013	10	28	1823	1725	58	2050	2019	31	UA
2013	8	9	2216	2040	96	2344	2154	110	B6
2013	10	6	1747	1732	15	2035	1959	36	FL
2013	2	4	1624	1630	-6	1836	1838	-2	DL
2013	6	17	1956	1959	-3	2123	2140	-17	DL
2013	10	29	1321	1257	24	1425	1414	11	EV
2013	12	24	941	945	-4	1153	1202	-9	EV
2013	4	29	2146	2130	16	35	16	19	B6
2013	3	4	1647	1630	17	1953	1954	-1	B6
2013	7	17	1920	1930	-10	2044	2051	-7	EV
2013	1	15	NA	1359	NA	NA	1656	NA	UA
2013	2	5	1453	1500	-7	1711	1655	16	MQ
2013	1	11	1557	1600	-3	1720	1712	8	US

YEAR	MONTH	DAY	DEP_TIME	SCHED_DEP_TIME	DEP_DELAY	ARR_TIME	SCHED_ARR_TIME	ARR_DELAY	CARRIER
2013	11	16	2108	2059	9	2237	2238	-1	EV
2013	6	11	2013	2015	-2	2240	2308	-28	B6
2013	5	24	758	759	-1	956	1020	-24	EV
...	...	...	...	...	...	...	...	...	...
2013	7	14	825	830	-5	1008	1028	-20	EV
2013	3	28	1535	1540	-5	1736	1820	-44	DL
2013	7	20	2008	2005	3	2317	2300	17	UA
2013	3	29	1952	1955	-3	2230	2240	-10	EV
2013	9	22	1255	1300	-5	1415	1415	0	US
2013	10	6	837	840	-3	1001	955	6	UA
2013	9	24	814	820	-6	1056	1110	-14	B6
2013	2	17	1947	1910	37	2256	2228	28	B6
2013	3	15	1514	1456	18	1813	1759	14	UA
2013	9	11	2334	2125	129	37	2247	110	EV
2013	8	18	839	843	-4	1126	1135	-9	UA
2013	5	30	1651	1529	82	1857	1740	77	EV
2013	10	17	1832	1820	12	2140	2125	15	UA
2013	8	12	1859	1827	32	2047	2033	14	US
2013	2	5	1342	1330	12	1658	1616	42	B6
2013	4	10	1703	1655	8	2129	2010	79	AA
2013	6	25	1336	1100	156	1606	1349	137	DL
2013	9	3	957	1000	-3	1218	1237	-19	DL
2013	7	25	746	745	1	929	1000	-31	9E
2013	3	29	1815	1815	0	2123	2147	-24	B6
2013	5	4	1403	1359	4	1509	1515	-6	EV
2013	4	7	1556	1600	-4	1829	1847	-18	DL
2013	11	7	1557	1610	-13	1810	1819	-9	EV
2013	7	11	801	800	1	1044	1104	-20	DL
2013	6	18	1313	1253	20	1553	1520	33	UA
2013	3	5	745	600	105	1006	825	101	MQ
2013	7	2	1258	1144	74	1433	1316	77	EV
2013	8	21	2144	2152	-8	2240	2258	-18	EV
2013	10	16	752	746	6	1027	1021	6	UA
2013	4	30	2150	2155	-5	2247	2312	-25	UA

Operator *pipe*

```
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

Nesting

```
filter(df, 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

```
sample_n(filter(df, mpg > 20), size = 5) # dua operasi
secara bersamaan
```

	MPG	CYL	DISP	HP	DRAT	WT	QSEC	VS	AM	GEAR	CARB
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1

```
hasil <- arrange(sample_n(filter(df, mpg > 20), size = 5),
desc(mpg))
# memfilter df untuk mpg > 20
# mengambil 5 baris sampel acak
# mengurutkannya berdasarkan kolom mpg secara terbalik
(descending order)
hasil
```

	MPG	CYL	DISP	HP	DRAT	WT	QSEC	VS	AM	GEAR	CARB
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4

```
# susah dibaca!!!
```

Penugasan berganda

```
a <- filter(df, mpg > 20)
b <- sample_n(a, size = 5)
hasil <- arrange(b, desc(mpg))
hasil
```

	MPG	CYL	DISP	HP	DRAT	WT	QSEC	VS	AM	GEAR	CARB
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
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
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1

Operator *pipe*

```
hasil <- df %>% filter(mpg > 20) %>% sample_n(size = 5)
%>% arrange(desc(mpg))
# lebih mudah diBACA!
hasil
```

	MPG	CYL	DISP	HP	DRAT	WT	QSEC	VS	AM	GEAR	CARB
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	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
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4

tidyr

- dplyr → manipulasi data.
- tidyr → pembersihan data.

```r
library(tidyr)
library(data.table) # punya kelebihan soal waktu eksekusi
dibandingkan data frame
```

```
Attaching package: 'data.table'

The following objects are masked from 'package:dplyr':

    between, first, last
```

`gather()`

Membagi kolom menjadi pasangan *key-value*

```r
v <- c('A', 'B', 'C','A', 'B', 'C','A', 'B', 'C')
thn <- c(2020,2021,2022,2020,2021,2022,2020,2021,2022)

q1 <- runif(n = 9, min = 0, max = 100)
q2 <- runif(n = 9, min = 0, max = 100)
q3 <- runif(n = 9, min = 0, max = 100)
q4 <- runif(n = 9, min = 0, max = 100)

df <- data.frame(Perusahaan = v, Tahun = thn, q1, q2, q3,
q4)
df
```

PERUSAHAAN	TAHUN	Q1	Q2	Q3	Q4
A	2020	71.401874	34.04511	75.01010	35.93238
B	2021	6.254729	37.22828	58.15716	5.82554
C	2022	58.959634	87.88993	43.93756	14.01085
A	2020	60.140881	69.06750	33.05298	69.82548
B	2021	12.791612	76.31263	66.74539	73.95468
C	2022	77.511686	15.82185	32.11450	84.96146
A	2020	31.628151	45.47135	77.67680	77.78378
B	2021	87.868777	72.97178	55.51417	39.13752
C	2022	99.692759	15.36461	73.52894	42.38900

```r
gather(data = df, key = 'Kuarter', value = 'Keuntungan',
q1:q4)
# 4 kolom dibuat jadi 2 kolom
```

PERUSAHAAN	TAHUN	KUARTER	KEUNTUNGAN
A	2020	q1	71.401874
B	2021	q1	6.254729
C	2022	q1	58.959634
A	2020	q1	60.140881
B	2021	q1	12.791612
C	2022	q1	77.511686
A	2020	q1	31.628151
B	2021	q1	87.868777
C	2022	q1	99.692759
A	2020	q2	34.045112
B	2021	q2	37.228278
C	2022	q2	87.889935
A	2020	q2	69.067497
B	2021	q2	76.312626
C	2022	q2	15.821849

PERUSAHAAN	TAHUN	KUARTER	KEUNTUNGAN
A	2020	q2	45.471355
B	2021	q2	72.971783
C	2022	q2	15.364613
A	2020	q3	75.010101
B	2021	q3	58.157156
C	2022	q3	43.937558
A	2020	q3	33.052977
B	2021	q3	66.745387
C	2022	q3	32.114503
A	2020	q3	77.676802
B	2021	q3	55.514168
C	2022	q3	73.528938
A	2020	q4	35.932385
B	2021	q4	5.825540
C	2022	q4	14.010853
A	2020	q4	69.825478
B	2021	q4	73.954683
C	2022	q4	84.961463
A	2020	q4	77.783779
B	2021	q4	39.137522
C	2022	q4	42.389005

```
# menggunakan fungsi gather() dengan menggunakan operator
pipe

df %>% gather(key = 'Kuarter', value = 'Keuntungan',
q1:q4)
```

PERUSAHAAN	TAHUN	KUARTER	KEUNTUNGAN
A	2020	q1	71.401874
B	2021	q1	6.254729
C	2022	q1	58.959634
A	2020	q1	60.140881
B	2021	q1	12.791612
C	2022	q1	77.511686
A	2020	q1	31.628151
B	2021	q1	87.868777
C	2022	q1	99.692759
A	2020	q2	34.045112
B	2021	q2	37.228278
C	2022	q2	87.889935
A	2020	q2	69.067497
B	2021	q2	76.312626
C	2022	q2	15.821849
A	2020	q2	45.471355
B	2021	q2	72.971783
C	2022	q2	15.364613
A	2020	q3	75.010101
B	2021	q3	58.157156
C	2022	q3	43.937558
A	2020	q3	33.052977
B	2021	q3	66.745387
C	2022	q3	32.114503
A	2020	q3	77.676802
B	2021	q3	55.514168
C	2022	q3	73.528938
A	2020	q4	35.932385
B	2021	q4	5.825540
C	2022	q4	14.010853

PERUSAHAAN	TAHUN	KUARTER	KEUNTUNGAN
A	2020	q4	69.825478
B	2021	q4	73.954683
C	2022	q4	84.961463
A	2020	q4	77.783779
B	2021	q4	39.137522
C	2022	q4	42.389005

spread()

```r
saham <- data.frame(
    waktu = as.Date('2009-01-01') + 0:9,
    X = rnorm(10,0,1),
    Y = rnorm(10,0,2),
    Z = rnorm(10,0,4)
)
```

```r
saham
```

WAKTU	X	Y	Z
2009-01-01	-0.9984820	-0.9431436	-3.33749766
2009-01-02	1.2059759	1.5118200	1.06892403
2009-01-03	-2.2298649	1.0133796	4.37434315
2009-01-04	-0.1377994	-1.5819430	0.52897651
2009-01-05	-0.4644349	-2.3425200	-0.02850392
2009-01-06	0.6431310	-3.0102303	-4.28047760
2009-01-07	-0.9540437	2.1494300	0.59964915
2009-01-08	-0.4403926	0.4593915	-4.46067291
2009-01-09	-1.2862645	1.2761994	1.99288550
2009-01-10	0.3102627	-1.6672000	10.22607936

```r
saham.gather <- gather(saham, key = saham, value = harga,
X,Y,Z)
saham.gather
```

WAKTU	SAHAM	HARGA
2009-01-01	X	-0.99848196
2009-01-02	X	1.20597592
2009-01-03	X	-2.22986494
2009-01-04	X	-0.13779945
2009-01-05	X	-0.46443485
2009-01-06	X	0.64313100
2009-01-07	X	-0.95404366
2009-01-08	X	-0.44039255
2009-01-09	X	-1.28626448
2009-01-10	X	0.31026273
2009-01-01	Y	-0.94314360
2009-01-02	Y	1.51182002
2009-01-03	Y	1.01337955
2009-01-04	Y	-1.58194299
2009-01-05	Y	-2.34251996
2009-01-06	Y	-3.01023026
2009-01-07	Y	2.14943001
2009-01-08	Y	0.45939146
2009-01-09	Y	1.27619940
2009-01-10	Y	-1.66720001
2009-01-01	Z	-3.33749766
2009-01-02	Z	1.06892403
2009-01-03	Z	4.37434315

WAKTU	SAHAM	HARGA
2009-01-04	Z	0.52897651
2009-01-05	Z	-0.02850392
2009-01-06	Z	-4.28047760
2009-01-07	Z	0.59964915
2009-01-08	Z	-4.46067291
2009-01-09	Z	1.99288550
2009-01-10	Z	10.22607936

```
spread(data = saham.gather, key = 'saham', value =
'harga')
# menyebar data hasil gather (kebalikan)
```

WAKTU	X	Y	Z
2009-01-01	-0.9984820	-0.9431436	-3.33749766
2009-01-02	1.2059759	1.5118200	1.06892403
2009-01-03	-2.2298649	1.0133796	4.37434315
2009-01-04	-0.1377994	-1.5819430	0.52897651
2009-01-05	-0.4644349	-2.3425200	-0.02850392
2009-01-06	0.6431310	-3.0102303	-4.28047760
2009-01-07	-0.9540437	2.1494300	0.59964915
2009-01-08	-0.4403926	0.4593915	-4.46067291
2009-01-09	-1.2862645	1.2761994	1.99288550
2009-01-10	0.3102627	-1.6672000	10.22607936

```
# menggunakan spread() dengan operator pipe
saham.gather %>% spread(key = 'saham', value = 'harga')
```

WAKTU	X	Y	Z
2009-01-01	-0.9984820	-0.9431436	-3.33749766
2009-01-02	1.2059759	1.5118200	1.06892403
2009-01-03	-2.2298649	1.0133796	4.37434315
2009-01-04	-0.1377994	-1.5819430	0.52897651
2009-01-05	-0.4644349	-2.3425200	-0.02850392
2009-01-06	0.6431310	-3.0102303	-4.28047760
2009-01-07	-0.9540437	2.1494300	0.59964915
2009-01-08	-0.4403926	0.4593915	-4.46067291
2009-01-09	-1.2862645	1.2761994	1.99288550
2009-01-10	0.3102627	-1.6672000	10.22607936

`separate()`

Memisahkan satu kolom ke banyak kolom.

```
df <- data.frame(kol.baru = c('a.x', 'b.y', 'c.z'))
df
```

KOL.BARU
a.x
b.y
c.z

```
separate(df, kol.baru, c("ABC", "XYZ"))
```

ABC	XYZ
a	x
b	y
c	z

```r
df <- data.frame(kol.baru = c('a-x', 'b-y', 'c-z'))
df
```

KOL.BARU
a-x
b-y
c-z

```r
separate(df, kol.baru, c("ABC", "XYZ"))
```

ABC	XYZ
a	x
b	y
c	z

```r
# Sintaks lengkapnya:
separate(data = df, col = kol.baru, c('Pertama', 'Kedua'),
sep="-")
```

PERTAMA	KEDUA
a	x
b	y
c	z

unite()

Merupakan kebalikan dari `separate()`. Digunakan untuk menggabungkan kolom.

```r
df1 <- separate(data = df, col = kol.baru, c('Pertama',
'Kedua'), sep="-")
df1
```

PERTAMA	KEDUA
a	x
b	y
c	z

```r
# menggabungkan jd 1 kolom
unite(df1, kol.gab.baru, Pertama, Kedua, sep = '-')
```

KOL.GAB.BARU
a-x
b-y
c-z