

WRAGGLING

Faustina Chelloana Triatmojo_123210139

2024-10-30

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2     3.4.3      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## v purrr       1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dslabs)
```

```
## Warning: package 'dslabs' was built under R version 4.3.3
```

```
#Mendapatkan path data bawaan pustaka dslabs
path <- system.file("extdata", package="dslabs")

#Menentukan nama file yang akan dibaca
filename <- file.path(path, "fertility-two-countries-example.csv")

#Membaca data dari file CSV
wide_data <- read_csv(filename)
```

```
## Rows: 2 Columns: 57
## -- Column specification -----
## Delimiter: ","
## chr (1): country
## dbl (56): 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, ...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

cols: Parameter ini menentukan kolom-kolom mana yang akan “dilebur” dari bentuk wide ke bentuk long. Dalam contoh di atas, cols = 1960:2015 berarti bahwa kolom dari tahun 1960 sampai 2015 akan dipilih dan

“dilebur” menjadi satu kolom bernama year. Kolom-kolom ini menjadi satu variabel dalam data long (yaitu kolom year).

names_to: Parameter ini menentukan nama kolom baru untuk nama-nama yang sebelumnya menjadi header (judul kolom) di data wide. Pada contoh di atas, names_to = “year” berarti nama-nama kolom dari 1960 hingga 2015 akan menjadi isi dari kolom baru yang dinamai year.

values_to: Parameter ini menentukan nama kolom baru untuk nilai-nilai dari kolom-kolom yang dilebur. Dalam contoh ini, values_to = “fertility” berarti nilai-nilai yang ada di kolom 1960 hingga 2015 akan disimpan dalam kolom fertility di data long.

```
#new_tidy_data1 <- wide_data %>% gather(year, fertility, `1960`:`2015`)

new_tidy_data <- wide_data %>%
  pivot_longer(cols = `1960`:`2015`, names_to = "year", values_to = "fertility")

new_tidy_data
```

```
## # A tibble: 112 x 3
##   country year  fertility
##   <chr>   <chr>    <dbl>
## 1 Germany 1960      2.41
## 2 Germany 1961      2.44
## 3 Germany 1962      2.47
## 4 Germany 1963      2.49
## 5 Germany 1964      2.49
## 6 Germany 1965      2.48
## 7 Germany 1966      2.44
## 8 Germany 1967      2.37
## 9 Germany 1968      2.28
## 10 Germany 1969      2.17
## # i 102 more rows
```

names_from: Parameter ini menentukan kolom yang nilainya akan menjadi nama kolom baru dalam data wide. Dalam contoh di atas, names_from = “year” berarti nilai-nilai yang ada di kolom year di data long akan menjadi nama kolom-kolom baru di data wide (misalnya, 1960, 1961, 1962, dan seterusnya).

values_from: Parameter ini menentukan kolom yang nilainya akan mengisi nilai dalam kolom-kolom baru tersebut. Dalam contoh di atas, values_from = “fertility” berarti nilai-nilai di kolom fertility pada data long akan ditempatkan di bawah kolom-kolom yang baru dibuat berdasarkan names_from.

```
#new_wide_data1 <- new_tidy_data1 %>% spread(year, fertility)

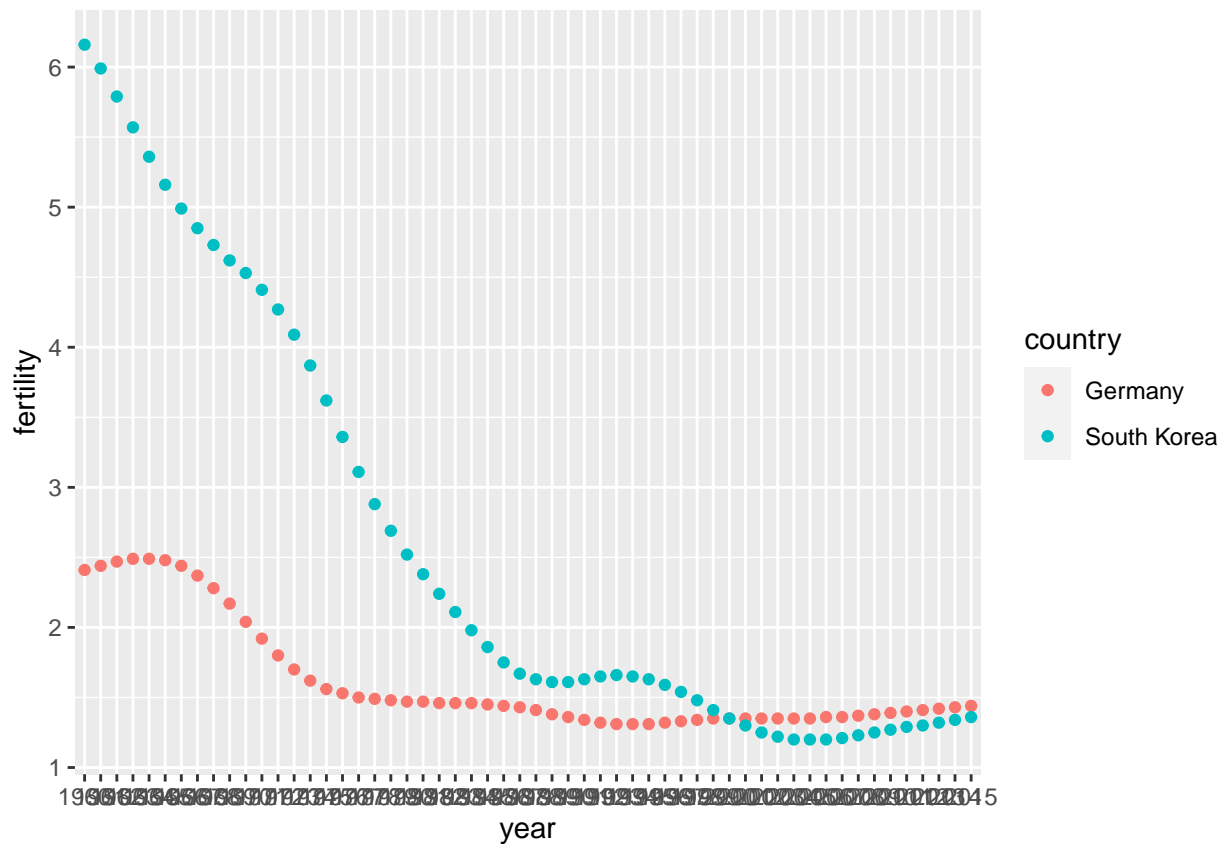
new_wide_data <- new_tidy_data %>%
  pivot_wider(names_from = "year", values_from = "fertility")

new_wide_data
```

```
## # A tibble: 2 x 57
##   country `1960` `1961` `1962` `1963` `1964` `1965` `1966` `1967` `1968` `1969`
##   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Germany 2.41  2.44  2.47  2.49  2.49  2.48  2.44  2.37  2.28  2.17
## 2 South K~ 6.16  5.99  5.79  5.57  5.36  5.16  4.99  4.85  4.73  4.62
## # i 46 more variables: `1970` <dbl>, `1971` <dbl>, `1972` <dbl>, `1973` <dbl>,
```

```
## # '1974' <dbl>, '1975' <dbl>, '1976' <dbl>, '1977' <dbl>, '1978' <dbl>,
## # '1979' <dbl>, '1980' <dbl>, '1981' <dbl>, '1982' <dbl>, '1983' <dbl>,
## # '1984' <dbl>, '1985' <dbl>, '1986' <dbl>, '1987' <dbl>, '1988' <dbl>,
## # '1989' <dbl>, '1990' <dbl>, '1991' <dbl>, '1992' <dbl>, '1993' <dbl>,
## # '1994' <dbl>, '1995' <dbl>, '1996' <dbl>, '1997' <dbl>, '1998' <dbl>,
## # '1999' <dbl>, '2000' <dbl>, '2001' <dbl>, '2002' <dbl>, '2003' <dbl>, ...
```

```
new_tidy_data %>% ggplot(aes(year, fertility, color = country)) + geom_point()
```



-Joining Table -

dataset results_us_election_2016 mungkin belum dimuat secara otomatis oleh dslabs

```
data(murders)
results_us_election_2016 <- dslabs::results_us_election_2016

murders
```

	state	abb	region	population	total
## 1	Alabama	AL	South	4779736	135
## 2	Alaska	AK	West	710231	19
## 3	Arizona	AZ	West	6392017	232
## 4	Arkansas	AR	South	2915918	93
## 5	California	CA	West	37253956	1257
## 6	Colorado	CO	West	5029196	65
## 7	Connecticut	CT	Northeast	3574097	97

## 8	Delaware	DE	South	897934	38
## 9	District of Columbia	DC	South	601723	99
## 10	Florida	FL	South	19687653	669
## 11	Georgia	GA	South	9920000	376
## 12	Hawaii	HI	West	1360301	7
## 13	Idaho	ID	West	1567582	12
## 14	Illinois	IL	North Central	12830632	364
## 15	Indiana	IN	North Central	6483802	142
## 16	Iowa	IA	North Central	3046355	21
## 17	Kansas	KS	North Central	2853118	63
## 18	Kentucky	KY	South	4339367	116
## 19	Louisiana	LA	South	4533372	351
## 20	Maine	ME	Northeast	1328361	11
## 21	Maryland	MD	South	5773552	293
## 22	Massachusetts	MA	Northeast	6547629	118
## 23	Michigan	MI	North Central	9883640	413
## 24	Minnesota	MN	North Central	5303925	53
## 25	Mississippi	MS	South	2967297	120
## 26	Missouri	MO	North Central	5988927	321
## 27	Montana	MT	West	989415	12
## 28	Nebraska	NE	North Central	1826341	32
## 29	Nevada	NV	West	2700551	84
## 30	New Hampshire	NH	Northeast	1316470	5
## 31	New Jersey	NJ	Northeast	8791894	246
## 32	New Mexico	NM	West	2059179	67
## 33	New York	NY	Northeast	19378102	517
## 34	North Carolina	NC	South	9535483	286
## 35	North Dakota	ND	North Central	672591	4
## 36	Ohio	OH	North Central	11536504	310
## 37	Oklahoma	OK	South	3751351	111
## 38	Oregon	OR	West	3831074	36
## 39	Pennsylvania	PA	Northeast	12702379	457
## 40	Rhode Island	RI	Northeast	1052567	16
## 41	South Carolina	SC	South	4625364	207
## 42	South Dakota	SD	North Central	814180	8
## 43	Tennessee	TN	South	6346105	219
## 44	Texas	TX	South	25145561	805
## 45	Utah	UT	West	2763885	22
## 46	Vermont	VT	Northeast	625741	2
## 47	Virginia	VA	South	8001024	250
## 48	Washington	WA	West	6724540	93
## 49	West Virginia	WV	South	1852994	27
## 50	Wisconsin	WI	North Central	5686986	97
## 51	Wyoming	WY	West	563626	5

results_us_election_2016

##	state	electoral_votes	clinton	trump	others
## 1	California	55	61.7	31.6	6.7
## 2	Texas	38	43.2	52.2	4.5
## 3	Florida	29	47.8	49.0	3.2
## 4	New York	29	59.0	36.5	4.5
## 5	Illinois	20	55.8	38.8	5.4
## 6	Pennsylvania	20	47.9	48.6	3.6

## 7	Ohio	18	43.5	51.7	4.8
## 8	Georgia	16	45.9	51.0	3.1
## 9	Michigan	16	47.3	47.5	5.2
## 10	North Carolina	15	46.2	49.8	4.0
## 11	New Jersey	14	55.5	41.4	3.2
## 12	Virginia	13	49.8	44.4	5.8
## 13	Washington	12	54.3	38.1	7.6
## 14	Arizona	11	45.1	48.7	6.2
## 15	Indiana	11	37.8	56.9	5.3
## 16	Massachusetts	11	60.0	32.8	7.2
## 17	Tennessee	11	34.7	60.7	4.6
## 18	Maryland	10	60.3	33.9	5.8
## 19	Minnesota	10	46.4	44.9	8.6
## 20	Missouri	10	38.1	56.8	5.1
## 21	Wisconsin	10	46.5	47.2	6.3
## 22	Alabama	9	34.4	62.1	3.6
## 23	Colorado	9	48.2	43.3	8.6
## 24	South Carolina	9	40.7	54.9	4.4
## 25	Kentucky	8	32.7	62.5	4.8
## 26	Louisiana	8	38.4	58.1	3.5
## 27	Connecticut	7	54.6	40.9	4.5
## 28	Oklahoma	7	28.9	65.3	5.7
## 29	Oregon	7	50.1	39.1	10.8
## 30	Arkansas	6	33.7	60.6	5.8
## 31	Iowa	6	41.7	51.1	7.1
## 32	Kansas	6	36.1	56.7	7.3
## 33	Mississippi	6	40.1	57.9	1.9
## 34	Nevada	6	47.9	45.5	6.6
## 35	Utah	6	27.5	45.5	27.0
## 36	Nebraska	5	34.3	59.9	5.8
## 37	New Mexico	5	48.3	40.0	11.7
## 38	West Virginia	5	26.5	68.6	4.9
## 39	Hawaii	4	62.2	30.0	7.7
## 40	Idaho	4	27.5	59.3	13.2
## 41	Maine	4	48.0	45.0	7.0
## 42	New Hampshire	4	46.8	46.5	6.7
## 43	Rhode Island	4	54.4	38.9	6.7
## 44	Alaska	3	36.6	51.3	12.2
## 45	Delaware	3	53.4	41.9	4.7
## 46	Montana	3	35.9	56.5	7.6
## 47	North Dakota	3	27.2	63.0	9.8
## 48	South Dakota	3	31.7	61.5	6.7
## 49	Vermont	3	56.7	30.3	13.1
## 50	Wyoming	3	21.9	68.2	10.0
## 51	District of Columbia	3	90.9	4.1	5.0

```
tab_1 <- slice(murders, 1:6) %>% select(state, population)
tab_1
```

##	state	population
## 1	Alabama	4779736
## 2	Alaska	710231
## 3	Arizona	6392017
## 4	Arkansas	2915918

```
## 5 California 37253956
## 6 Colorado 5029196
```

```
tab_2 <- results_us_election_2016 %>%
  filter(state %in% c("Alabama", "Alaska", "Arizona", "California", "Connecticut", "Delaware")) %>%
  select(state, electoral_votes) %>%
  rename(ev = electoral_votes)
tab_2
```

```
##      state ev
## 1 California 55
## 2 Arizona 11
## 3 Alabama 9
## 4 Connecticut 7
## 5 Alaska 3
## 6 Delaware 3
```

left_join

-karna berdasarkan state, terdapat NA karena Arkansas dan Colorado tidak ada pada tab_1

```
left_join(tab_1, tab_2, by = "state")
```

```
##      state population ev
## 1 Alabama 4779736 9
## 2 Alaska 710231 3
## 3 Arizona 6392017 11
## 4 Arkansas 2915918 NA
## 5 California 37253956 55
## 6 Colorado 5029196 NA
```

```
#alternatif
tab_1 %>% left_join(tab_2, by = "state")
```

```
##      state population ev
## 1 Alabama 4779736 9
## 2 Alaska 710231 3
## 3 Arizona 6392017 11
## 4 Arkansas 2915918 NA
## 5 California 37253956 55
## 6 Colorado 5029196 NA
```

right_join

```
tab_1 %>% right_join(tab_2, by = "state")
```

```
##      state population ev
## 1 Alabama 4779736 9
## 2 Alaska 710231 3
## 3 Arizona 6392017 11
## 4 California 37253956 55
## 5 Connecticut NA 7
## 6 Delaware NA 3
```

inner_join

```
inner_join(tab_1, tab_2, by = "state")
```

```
##      state population ev
## 1  Alabama      4779736  9
## 2   Alaska       710231  3
## 3  Arizona      6392017 11
## 4 California    37253956 55
```

full_join

```
full_join(tab_1, tab_2, by = "state")
```

```
##      state population ev
## 1  Alabama      4779736  9
## 2   Alaska       710231  3
## 3  Arizona      6392017 11
## 4  Arkansas      2915918 NA
## 5 California    37253956 55
## 6  Colorado      5029196 NA
## 7 Connecticut          NA  7
## 8  Delaware          NA   3
```

semi_join

-ga akan ada NA- Fungsi semi_join dapat digunakan untuk menampilkan bagian dari tabel pertama yang juga memiliki baris informasi yang sama di tabel kedua.

```
semi_join(tab_1, tab_2, by = "state")
```

```
##      state population
## 1  Alabama      4779736
## 2   Alaska       710231
## 3  Arizona      6392017
## 4 California    37253956
```

anti_join

Fungsi anti_join adalah kebalikan dari semi_join. Hasil dari fungsi ini akan menampilkan elemen dari tabel pertama yang tidak memiliki baris informasi yang sama di tabel kedua

```
anti_join(tab_1, tab_2, by = "state")
```

```
##      state population
## 1 Arkansas      2915918
## 2 Colorado      5029196
```