



R ile Veri Manipölasyonu

Muhammed Fatih TÖZEN

06 01 2022

İÇİNDEKİLER

1	dplyr Paketi	3
2	Counties veri seti	3
3	select	5
4	arrange	8
5	filter	9
6	mutate	11
7	rename	15
8	count	16
9	group_by ve summarize	17
10	case when	23
11	reshaping	24
12	join	28

1 dplyr Paketi

dplyr, RStudio'dan Hadley Wickham tarafından geliştirilmiş ve en yaygın veri işleme zorluklarını çözenize yardımcı olan bir veri işleme dilbilgisidir. **dplyr** paketi, **devtools** paketi ve `install_github()` fonksiyonu kullanılarak **CRAN**'dan veya **GitHub**'dan kurulabilir. GitHub deposu genellikle paketteki en son güncellemeleri ve geliştirme sürümünü içerir.

CRAN sayfasından yüklemek için;

```
> install.packages("dplyr")
```

GitHub sayfasından yüklemek için;

```
> install_github("hadley/dplyr")
```

dplyr paketinde sıklıkla kullanılan fonksiyonlar şunlardır:

- **select** : veri çerçevesinden istenilen sütunları seçer.
- **filter** : mantıksal koşullara dayalı olarak bir veri çerçevesinden satırları filtreler.
- **arrange** : satırları sıralar.
- **rename** : sütun isimlerini yeniden isimlendirir.
- **mutate** : yeni değişkenler/sütunlar ekler veya mevcut değişkenleri dönüştürür.
- **summarise/ summarize** : veri çerçevesindeki farklı değişkenlerin özet istatistiklerini oluşturur
- **%>%** (pipe) operatörü birden çok eylemi ardışık düzende birbirine bağlamak için kullanılır.

2 Counties veri seti

Veri manipülasyonu için 2015 yılı ABD nüfus sayımına ilişkin **counties** veri seti kullanılacaktır. Bu veri setinde eyalet ve şehir detayında nüfus, gelir, ırk, coğrafi yapı, işgücü gibi değişkenler yer almaktadır.

```
library(dplyr)
counties <- readRDS("datasets/counties.rds")

# veri setinin yapısı hakkında bilgi sağlar
glimpse(counties)
```

```

## Rows: 3,138
## Columns: 40
## $ census_id      <chr> "1001", "1003", "1005", "1007", "1009", "1011", "10~
## $ state          <chr> "Alabama", "Alabama", "Alabama", "Alabama", "Alabam~
## $ county         <chr> "Autauga", "Baldwin", "Barbour", "Bibb", "Blount", ~
## $ region         <chr> "South", "South", "South", "South", "South", "South~
## $ metro          <chr> "Metro", "Metro", "Nonmetro", "Metro", "Metro", "No~
## $ population     <dbl> 55221, 195121, 26932, 22604, 57710, 10678, 20354, 1~
## $ men            <dbl> 26745, 95314, 14497, 12073, 28512, 5660, 9502, 5627~
## $ women          <dbl> 28476, 99807, 12435, 10531, 29198, 5018, 10852, 603~
## $ hispanic       <dbl> 2.6, 4.5, 4.6, 2.2, 8.6, 4.4, 1.2, 3.5, 0.4, 1.5, 7~
## $ white          <dbl> 75.8, 83.1, 46.2, 74.5, 87.9, 22.2, 53.3, 73.0, 57.~
## $ black          <dbl> 18.5, 9.5, 46.7, 21.4, 1.5, 70.7, 43.8, 20.3, 40.3,~
## $ native         <dbl> 0.4, 0.6, 0.2, 0.4, 0.3, 1.2, 0.1, 0.2, 0.2, 0.6, 0~
## $ asian          <dbl> 1.0, 0.7, 0.4, 0.1, 0.1, 0.2, 0.4, 0.9, 0.8, 0.3, 0~
## $ pacific        <dbl> 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0~
## $ citizens       <dbl> 40725, 147695, 20714, 17495, 42345, 8057, 15581, 88~
## $ income         <dbl> 51281, 50254, 32964, 38678, 45813, 31938, 32229, 41~
## $ income_err     <dbl> 2391, 1263, 2973, 3995, 3141, 5884, 1793, 925, 2949~
## $ income_per_cap <dbl> 24974, 27317, 16824, 18431, 20532, 17580, 18390, 21~
## $ income_per_cap_err <dbl> 1080, 711, 798, 1618, 708, 2055, 714, 489, 1366, 15~
## $ poverty        <dbl> 12.9, 13.4, 26.7, 16.8, 16.7, 24.6, 25.4, 20.5, 21.~
## $ child_poverty  <dbl> 18.6, 19.2, 45.3, 27.9, 27.2, 38.4, 39.2, 31.6, 37.~
## $ professional   <dbl> 33.2, 33.1, 26.8, 21.5, 28.5, 18.8, 27.5, 27.3, 23.~
## $ service        <dbl> 17.0, 17.7, 16.1, 17.9, 14.1, 15.0, 16.6, 17.7, 14.~
## $ office         <dbl> 24.2, 27.1, 23.1, 17.8, 23.9, 19.7, 21.9, 24.2, 26.~
## $ construction   <dbl> 8.6, 10.8, 10.8, 19.0, 13.5, 20.1, 10.3, 10.5, 11.5~
## $ production     <dbl> 17.1, 11.2, 23.1, 23.7, 19.9, 26.4, 23.7, 20.4, 24.~
## $ drive          <dbl> 87.5, 84.7, 83.8, 83.2, 84.9, 74.9, 84.5, 85.3, 85.~
## $ carpool        <dbl> 8.8, 8.8, 10.9, 13.5, 11.2, 14.9, 12.4, 9.4, 11.9, ~
## $ transit        <dbl> 0.1, 0.1, 0.4, 0.5, 0.4, 0.7, 0.0, 0.2, 0.2, 0.2, 0~
## $ walk           <dbl> 0.5, 1.0, 1.8, 0.6, 0.9, 5.0, 0.8, 1.2, 0.3, 0.6, 1~
## $ other_transp    <dbl> 1.3, 1.4, 1.5, 1.5, 0.4, 1.7, 0.6, 1.2, 0.4, 0.7, 1~
## $ work_at_home    <dbl> 1.8, 3.9, 1.6, 0.7, 2.3, 2.8, 1.7, 2.7, 2.1, 2.5, 1~
## $ mean_commute    <dbl> 26.5, 26.4, 24.1, 28.8, 34.9, 27.5, 24.6, 24.1, 25.~
## $ employed       <dbl> 23986, 85953, 8597, 8294, 22189, 3865, 7813, 47401,~
## $ private_work    <dbl> 73.6, 81.5, 71.8, 76.8, 82.0, 79.5, 77.4, 74.1, 85.~
## $ public_work     <dbl> 20.9, 12.3, 20.8, 16.1, 13.5, 15.1, 16.2, 20.8, 12.~
## $ self_employed   <dbl> 5.5, 5.8, 7.3, 6.7, 4.2, 5.4, 6.2, 5.0, 2.8, 7.9, 4~
## $ family_work     <dbl> 0.0, 0.4, 0.1, 0.4, 0.4, 0.0, 0.2, 0.1, 0.0, 0.5, 0~
## $ unemployment    <dbl> 7.6, 7.5, 17.6, 8.3, 7.7, 18.0, 10.9, 12.3, 8.9, 7.~
## $ land_area       <dbl> 594.44, 1589.78, 884.88, 622.58, 644.78, 622.81, 77~

```

3 select

```
# belirli sütunları seçmek
counties %>%
select(state, county, population, unemployment)
```

```
## # A tibble: 3,138 x 4
##   state   county   population unemployment
##   <chr>   <chr>         <dbl>         <dbl>
## 1 Alabama Autauga      55221          7.6
## 2 Alabama Baldwin    195121         7.5
## 3 Alabama Barbour    26932         17.6
## 4 Alabama Bibb       22604          8.3
## 5 Alabama Blount     57710          7.7
## 6 Alabama Bullock    10678          18
## 7 Alabama Butler     20354         10.9
## 8 Alabama Calhoun    116648         12.3
## 9 Alabama Chambers   34079          8.9
## 10 Alabama Cherokee  26008          7.9
## # ... with 3,128 more rows
```

```
# belli aralıkta bütün sütunların seçilmesi
counties %>%
select(state, county, drive:work_at_home)
```

```
## # A tibble: 3,138 x 8
##   state   county   drive carpool transit  walk other_transp work_at_home
##   <chr>   <chr>     <dbl>   <dbl>   <dbl> <dbl>         <dbl>         <dbl>
## 1 Alabama Autauga    87.5     8.8     0.1   0.5         1.3         1.8
## 2 Alabama Baldwin   84.7     8.8     0.1   1         1.4         3.9
## 3 Alabama Barbour   83.8    10.9     0.4   1.8         1.5         1.6
## 4 Alabama Bibb      83.2    13.5     0.5   0.6         1.5         0.7
## 5 Alabama Blount    84.9    11.2     0.4   0.9         0.4         2.3
## 6 Alabama Bullock   74.9    14.9     0.7   5          1.7         2.8
## 7 Alabama Butler    84.5    12.4     0     0.8         0.6         1.7
## 8 Alabama Calhoun   85.3     9.4     0.2   1.2         1.2         2.7
## 9 Alabama Chambers  85.1    11.9     0.2   0.3         0.4         2.1
## 10 Alabama Cherokee 83.9    12.1     0.2   0.6         0.7         2.5
## # ... with 3,128 more rows
```

```
# belirli bir ifadeyi içeren sütunları seçmek
counties %>%
select(state, county, contains("employed"))
```

```
## # A tibble: 3,138 x 4
##   state  county  employed self_employed
##   <chr>  <chr>      <dbl>      <dbl>
## 1 Alabama Autauga    23986        5.5
## 2 Alabama Baldwin    85953        5.8
## 3 Alabama Barbour     8597        7.3
## 4 Alabama Bibb       8294        6.7
## 5 Alabama Blount    22189        4.2
## 6 Alabama Bullock    3865        5.4
## 7 Alabama Butler     7813        6.2
## 8 Alabama Calhoun   47401         5
## 9 Alabama Chambers  13689        2.8
## 10 Alabama Cherokee  10155        7.9
## # ... with 3,128 more rows
```

```
# belirli bir ifade ile başlayan sütunları seçmek
counties %>%
select(state, county, starts_with("income"))
```

```
## # A tibble: 3,138 x 6
##   state  county  income income_err income_per_cap income_per_cap_err
##   <chr>  <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Alabama Autauga    51281      2391      24974      1080
## 2 Alabama Baldwin   50254      1263      27317       711
## 3 Alabama Barbour   32964      2973      16824       798
## 4 Alabama Bibb     38678      3995      18431      1618
## 5 Alabama Blount   45813      3141      20532       708
## 6 Alabama Bullock   31938      5884      17580      2055
## 7 Alabama Butler   32229      1793      18390       714
## 8 Alabama Calhoun   41703       925      21374       489
## 9 Alabama Chambers  34177      2949      21071      1366
## 10 Alabama Cherokee  36296      1710      21811      1556
## # ... with 3,128 more rows
```

```
# belirli bir ifade ile biten sütunları seçmek
counties %>%
select(state, county, ends_with("work"))
```

```
## # A tibble: 3,138 x 5
##   state   county private_work public_work family_work
##   <chr>   <chr>         <dbl>         <dbl>         <dbl>
## 1 Alabama Autauga          73.6           20.9           0
## 2 Alabama Baldwin          81.5           12.3           0.4
## 3 Alabama Barbour          71.8           20.8           0.1
## 4 Alabama Bibb             76.8           16.1           0.4
## 5 Alabama Blount           82            13.5           0.4
## 6 Alabama Bullock          79.5           15.1           0
## 7 Alabama Butler           77.4           16.2           0.2
## 8 Alabama Calhoun          74.1           20.8           0.1
## 9 Alabama Chambers          85.1           12.1           0
## 10 Alabama Cherokee        73.1           18.5           0.5
## # ... with 3,128 more rows
```

```
# belirli sütunları hariç tutarak seçmek
counties %>%
select(census_id:population,-c(men:land_area))
```

```
## # A tibble: 3,138 x 6
##   census_id state   county   region metro   population
##   <chr>      <chr>   <chr>   <chr> <chr>         <dbl>
## 1 1001      Alabama Autauga   South  Metro      55221
## 2 1003      Alabama Baldwin  South  Metro     195121
## 3 1005      Alabama Barbour  South  Nonmetro   26932
## 4 1007      Alabama Bibb     South  Metro      22604
## 5 1009      Alabama Blount   South  Metro      57710
## 6 1011      Alabama Bullock  South  Nonmetro   10678
## 7 1013      Alabama Butler   South  Nonmetro   20354
## 8 1015      Alabama Calhoun  South  Metro     116648
## 9 1017      Alabama Chambers South  Nonmetro   34079
## 10 1019      Alabama Cherokee South  Nonmetro   26008
## # ... with 3,128 more rows
```

```
# belirli veri tipindeki sütunları seçmek
counties %>%
select(where(is.character))
```

```
## # A tibble: 3,138 x 5
##   census_id state   county   region metro
##   <chr>      <chr>   <chr>   <chr> <chr>
## 1 1001      Alabama Autauga   South  Metro
## 2 1003      Alabama Baldwin  South  Metro
```

```
## 3 1005      Alabama Barbour  South  Nonmetro
## 4 1007      Alabama Bibb    South  Metro
## 5 1009      Alabama Blount  South  Metro
## 6 1011      Alabama Bullock South  Nonmetro
## 7 1013      Alabama Butler  South  Nonmetro
## 8 1015      Alabama Calhoun South  Metro
## 9 1017      Alabama Chambers South  Nonmetro
## 10 1019     Alabama Cherokee South  Nonmetro
## # ... with 3,128 more rows
```

4 arrange

```
counties_selected <- counties %>%
select(state, county, population, unemployment)

# artan sıralama (ascending)
counties_selected %>%
arrange(population)
```

```
## # A tibble: 3,138 x 4
##   state      county      population unemployment
##   <chr>      <chr>          <dbl>          <dbl>
## 1 Hawaii    Kalawao             85             0
## 2 Texas     King              267            5.1
## 3 Nebraska  McPherson          433            0.9
## 4 Montana   Petroleum          443            6.6
## 5 Nebraska  Arthur             448             4
## 6 Nebraska  Loup              548            0.7
## 7 Nebraska  Blaine            551            0.7
## 8 New Mexico Harding        565             6
## 9 Texas     Kenedy            565             0
## 10 Colorado San Juan           606           13.8
## # ... with 3,128 more rows
```

```
# azalan sıralama (descending)
counties_selected %>%
arrange(desc(population))
```

```
## # A tibble: 3,138 x 4
##   state      county      population unemployment
##   <chr>      <chr>          <dbl>          <dbl>
```



```
## 1 California Los Angeles 10038388 10
## 2 Illinois Cook 5236393 10.7
## 3 Texas Harris 4356362 7.5
## 4 Arizona Maricopa 4018143 7.7
## 5 California San Diego 3223096 8.7
## 6 California Orange 3116069 7.6
## 7 Florida Miami-Dade 2639042 10
## 8 New York Kings 2595259 10
## 9 Texas Dallas 2485003 7.6
## 10 New York Queens 2301139 8.6
## # ... with 3,128 more rows
```

```
counties_selected %>%
  arrange(state, desc(population))
```

```
## # A tibble: 3,138 x 4
##   state county population unemployment
##   <chr> <chr>      <dbl>      <dbl>
## 1 Alabama Jefferson 659026 9.1
## 2 Alabama Mobile 414251 9.8
## 3 Alabama Madison 346438 8.5
## 4 Alabama Montgomery 228138 8.8
## 5 Alabama Shelby 203530 5.5
## 6 Alabama Tuscaloosa 200458 7.6
## 7 Alabama Baldwin 195121 7.5
## 8 Alabama Lee 150982 7.3
## 9 Alabama Morgan 119786 9.9
## 10 Alabama Calhoun 116648 12.3
## # ... with 3,128 more rows
```

5 filter

```
# sadece New York'u filtrele
counties_selected %>%
  arrange(desc(population)) %>%
  filter(state == "New York")
```

```
## # A tibble: 62 x 4
##   state county population unemployment
##   <chr> <chr>      <dbl>      <dbl>
## 1 New York Kings 2595259 10
```

```
## 2 New York Queens      2301139      8.6
## 3 New York New York    1629507      7.5
## 4 New York Suffolk     1501373      6.4
## 5 New York Bronx       1428357      14
## 6 New York Nassau      1354612      6.4
## 7 New York Westchester 967315      7.6
## 8 New York Erie        921584      7
## 9 New York Monroe      749356      7.7
## 10 New York Richmond   472481      6.9
## # ... with 52 more rows
```

```
# işsizlik oranı 6'dan küçük olanları filtrele
counties_selected %>%
  arrange(desc(population)) %>%
  filter(unemployment < 6)
```

```
## # A tibble: 949 x 4
##   state   county      population unemployment
##   <chr>   <chr>         <dbl>         <dbl>
## 1 Virginia Fairfax     1128722         4.9
## 2 Utah    Salt Lake    1078958         5.8
## 3 Hawaii  Honolulu     984178         5.6
## 4 Texas   Collin       862215         4.9
## 5 Texas   Denton       731851         5.7
## 6 Texas   Fort Bend    658331         5.1
## 7 Kansas  Johnson      566814         4.5
## 8 Maryland Anne Arundel  555280         5.9
## 9 Colorado Jefferson   552344         5.9
## 10 Utah    Utah         551957         5.5
## # ... with 939 more rows
```

```
# birden fazla koşul
counties_selected %>%
  arrange(desc(population)) %>%
  filter(state == "New York", unemployment < 6)
```

```
## # A tibble: 5 x 4
##   state   county      population unemployment
##   <chr>   <chr>         <dbl>         <dbl>
## 1 New York Tompkins    103855         5.9
## 2 New York Chemung     88267         5.4
## 3 New York Madison     72427         5.1
## 4 New York Livingston   64801         5.4
## 5 New York Seneca      35144         5.5
```

6 mutate

```
# işsiz nüfus sayısına ilişkin değişken üretme
```

```
counties_selected %>%
```

```
mutate(unemployed_population = population * unemployment / 100)
```

```
## # A tibble: 3,138 x 5
```

```
##   state   county   population unemployment unemployed_population
```

```
##   <chr>   <chr>         <dbl>         <dbl>         <dbl>
```

```
## 1 Alabama Autauga      55221          7.6          4197.
```

```
## 2 Alabama Baldwin    195121         7.5         14634.
```

```
## 3 Alabama Barbour     26932         17.6         4740.
```

```
## 4 Alabama Bibb        22604          8.3         1876.
```

```
## 5 Alabama Blount      57710          7.7         4444.
```

```
## 6 Alabama Bullock     10678          18          1922.
```

```
## 7 Alabama Butler      20354         10.9         2219.
```

```
## 8 Alabama Calhoun     116648         12.3         14348.
```

```
## 9 Alabama Chambers    34079          8.9          3033.
```

```
## 10 Alabama Cherokee   26008          7.9          2055.
```

```
## # ... with 3,128 more rows
```

```
# yeni sütun ekle
```

```
counties_selected %>%
```

```
mutate(unemployed_population = population * unemployment / 100) %>%
```

```
arrange(desc(unemployed_population))
```

```
## # A tibble: 3,138 x 5
```

```
##   state   county   population unemployment unemployed_population
```

```
##   <chr>   <chr>         <dbl>         <dbl>         <dbl>
```

```
## 1 California Los Angeles  10038388      10      1003839.
```

```
## 2 Illinois   Cook          5236393     10.7     560294.
```

```
## 3 Texas      Harris       4356362      7.5     326727.
```

```
## 4 Arizona    Maricopa     4018143      7.7     309397.
```

```
## 5 California Riverside  2298032     12.9     296446.
```

```
## 6 California San Diego   3223096      8.7     280409.
```

```
## 7 Michigan   Wayne       1778969     14.9     265066.
```

```
## 8 California San Bernardino 2094769     12.6     263941.
```

```
## 9 Florida    Miami-Dade   2639042      10     263904.
```

```
## 10 New York   Kings        2595259      10     259526.
```

```
## # ... with 3,128 more rows
```

```
# var olan sütunu güncelle
counties %>%
  select(state, county, population, men,women) %>%
  mutate(population = men + women)
```

```
## # A tibble: 3,138 x 5
##   state   county   population   men women
##   <chr>   <chr>         <dbl> <dbl> <dbl>
## 1 Alabama Autauga      55221 26745 28476
## 2 Alabama Baldwin     195121 95314 99807
## 3 Alabama Barbour      26932 14497 12435
## 4 Alabama Bibb         22604 12073 10531
## 5 Alabama Blount       57710 28512 29198
## 6 Alabama Bullock      10678  5660  5018
## 7 Alabama Butler       20354  9502 10852
## 8 Alabama Calhoun     116648 56274 60374
## 9 Alabama Chambers     34079 16258 17821
## 10 Alabama Cherokee    26008 12975 13033
## # ... with 3,128 more rows
```

```
# birden fazla yeni değişken üretme
counties %>%
  select(state, county, population, men,women) %>%
  mutate(men_ratio = men/population*100,
         women_ratio = women/population*100)
```

```
## # A tibble: 3,138 x 7
##   state   county   population   men women men_ratio women_ratio
##   <chr>   <chr>         <dbl> <dbl> <dbl>      <dbl>      <dbl>
## 1 Alabama Autauga      55221 26745 28476      48.4      51.6
## 2 Alabama Baldwin     195121 95314 99807      48.8      51.2
## 3 Alabama Barbour      26932 14497 12435      53.8      46.2
## 4 Alabama Bibb         22604 12073 10531      53.4      46.6
## 5 Alabama Blount       57710 28512 29198      49.4      50.6
## 6 Alabama Bullock      10678  5660  5018      53.0      47.0
## 7 Alabama Butler       20354  9502 10852      46.7      53.3
## 8 Alabama Calhoun     116648 56274 60374      48.2      51.8
## 9 Alabama Chambers     34079 16258 17821      47.7      52.3
## 10 Alabama Cherokee    26008 12975 13033      49.9      50.1
## # ... with 3,128 more rows
```

transmute sadece yeni eklenen değişkenleri gösterir

```
counties %>%
  select(state, county, population, men,women) %>%
  transmute(men_ratio = men/population*100,
            women_ratio = women/population*100)
```

```
## # A tibble: 3,138 x 2
##   men_ratio women_ratio
##   <dbl>      <dbl>
## 1     48.4       51.6
## 2     48.8       51.2
## 3     53.8       46.2
## 4     53.4       46.6
## 5     49.4       50.6
## 6     53.0       47.0
## 7     46.7       53.3
## 8     48.2       51.8
## 9     47.7       52.3
## 10    49.9       50.1
## # ... with 3,128 more rows
```

mutate_at ile koşula göre birden fazla değişkene aynı fonksiyon uygulanabilir.

```
scale2 <- function(x, na.rm = FALSE) (x - mean(x, na.rm = na.rm)) / sd(x, na.rm)
```

```
counties_selected %>%
  mutate_at(c("population","unemployment"),scale2)
```

```
## # A tibble: 3,138 x 4
##   state    county    population unemployment
##   <chr>   <chr>      <dbl>      <dbl>
## 1 Alabama Autauga    -0.141     -0.0563
## 2 Alabama Baldwin     0.292     -0.0846
## 3 Alabama Barbour    -0.228      2.78
## 4 Alabama Bibb       -0.242      0.142
## 5 Alabama Blount     -0.133     -0.0279
## 6 Alabama Bullock    -0.278      2.89
## 7 Alabama Butler     -0.249      0.880
## 8 Alabama Calhoun     0.0495     1.28
## 9 Alabama Chambers   -0.206      0.313
## 10 Alabama Cherokee   -0.231     0.0288
## # ... with 3,128 more rows
```

```
counties_selected %>% # birden fazla argüman kullanımı
  mutate_at(c("population", "unemployment"), scale2, na.rm = TRUE)
```

```
## # A tibble: 3,138 x 4
##   state   county   population unemployment
##   <chr>   <chr>         <dbl>         <dbl>
## 1 Alabama Autauga     -0.141         -0.0563
## 2 Alabama Baldwin      0.292         -0.0846
## 3 Alabama Barbour     -0.228          2.78
## 4 Alabama Bibb        -0.242          0.142
## 5 Alabama Blount     -0.133        -0.0279
## 6 Alabama Bullock    -0.278          2.89
## 7 Alabama Butler     -0.249          0.880
## 8 Alabama Calhoun     0.0495          1.28
## 9 Alabama Chambers   -0.206          0.313
## 10 Alabama Cherokee  -0.231          0.0288
## # ... with 3,128 more rows
```

```
# mutate_if ile koşula göre birden fazla değişkende değişiklik yapılabilir.
str(counties_selected)
```

```
## spec_tbl_df [3,138 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ state      : chr [1:3138] "Alabama" "Alabama" "Alabama" "Alabama" ...
## $ county     : chr [1:3138] "Autauga" "Baldwin" "Barbour" "Bibb" ...
## $ population : num [1:3138] 55221 195121 26932 22604 57710 ...
## $ unemployment: num [1:3138] 7.6 7.5 17.6 8.3 7.7 18 10.9 12.3 8.9 7.9 ...
```

```
counties_selected <- counties_selected %>%
  mutate_if(is.character, as.factor)

str(counties_selected)
```

```
## spec_tbl_df [3,138 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ state      : Factor w/ 50 levels "Alabama","Alaska",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ county     : Factor w/ 1847 levels "Abbeville","Acadia",...: 82 89 100 149 164 225 ...
## $ population : num [1:3138] 55221 195121 26932 22604 57710 ...
## $ unemployment: num [1:3138] 7.6 7.5 17.6 8.3 7.7 18 10.9 12.3 8.9 7.9 ...
```

```
counties_selected %>%
  mutate_if(is.numeric, scale2, na.rm = TRUE)
```

```
## # A tibble: 3,138 x 4
##   state   county   population unemployment
##   <fct>   <fct>         <dbl>         <dbl>
## 1 Alabama Autauga      -0.141         -0.0563
## 2 Alabama Baldwin       0.292         -0.0846
## 3 Alabama Barbour      -0.228          2.78
## 4 Alabama Bibb        -0.242          0.142
## 5 Alabama Blount      -0.133        -0.0279
## 6 Alabama Bullock     -0.278          2.89
## 7 Alabama Butler      -0.249          0.880
## 8 Alabama Calhoun       0.0495          1.28
## 9 Alabama Chambers    -0.206          0.313
## 10 Alabama Cherokee    -0.231          0.0288
## # ... with 3,128 more rows
```

7 rename

```
# yeniden isimlendirmede eşitliği sol tarafı yeni isim olmalı
counties_selected %>%
  rename(unemployment_rate = unemployment)
```

```
## # A tibble: 3,138 x 4
##   state   county   population unemployment_rate
##   <fct>   <fct>         <dbl>         <dbl>
## 1 Alabama Autauga      55221          7.6
## 2 Alabama Baldwin    195121          7.5
## 3 Alabama Barbour     26932         17.6
## 4 Alabama Bibb        22604          8.3
## 5 Alabama Blount     57710          7.7
## 6 Alabama Bullock    10678          18
## 7 Alabama Butler     20354         10.9
## 8 Alabama Calhoun    116648         12.3
## 9 Alabama Chambers   34079          8.9
## 10 Alabama Cherokee   26008          7.9
## # ... with 3,128 more rows
```

```
# select ile beraber de yeniden isimlendirme yapılabilir
counties_selected %>%
  select(state, county, population, unemployment_rate = unemployment)
```

```
## # A tibble: 3,138 x 4
```

```
##      state   county   population unemployment_rate
##      <fct>   <fct>         <dbl>             <dbl>
##  1 Alabama Autauga      55221              7.6
##  2 Alabama Baldwin    195121             7.5
##  3 Alabama Barbour     26932            17.6
##  4 Alabama Bibb       22604             8.3
##  5 Alabama Blount     57710             7.7
##  6 Alabama Bullock    10678             18
##  7 Alabama Butler     20354            10.9
##  8 Alabama Calhoun    116648            12.3
##  9 Alabama Chambers   34079             8.9
## 10 Alabama Cherokee   26008             7.9
## # ... with 3,128 more rows
```

8 count

```
# count ile veri setinde sayma işlemleri yapılır
counties %>%
count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1  3138
```

```
# state dağılımını elde etmek
counties %>%
count(state)
```

```
## # A tibble: 50 x 2
##       state         n
##   <chr>       <int>
## 1 Alabama      67
## 2 Alaska       28
## 3 Arizona      15
## 4 Arkansas     75
## 5 California   58
## 6 Colorado     64
## 7 Connecticut   8
## 8 Delaware      3
## 9 Florida     67
## 10 Georgia    159
## # ... with 40 more rows
```



```
# sort = TRUE ile büyükten küçüğe sıralama yapılabilir
counties %>%
count(state, sort = TRUE)
```

```
## # A tibble: 50 x 2
##   state      n
##   <chr>    <int>
## 1 Texas      253
## 2 Georgia    159
## 3 Virginia   133
## 4 Kentucky   120
## 5 Missouri   115
## 6 Kansas     105
## 7 Illinois   102
## 8 North Carolina 100
## 9 Iowa        99
## 10 Tennessee   95
## # ... with 40 more rows
```

```
# wt argümanı ile değişken toplamaları hesaplanabilir
counties %>%
count(state, wt = population, sort = TRUE)
```

```
## # A tibble: 50 x 2
##   state      n
##   <chr>    <dbl>
## 1 California 38421464
## 2 Texas      26538497
## 3 New York   19673174
## 4 Florida    19645772
## 5 Illinois   12873761
## 6 Pennsylvania 12779559
## 7 Ohio       11575977
## 8 Georgia    10006693
## 9 Michigan    9900571
## 10 North Carolina 9845333
## # ... with 40 more rows
```

9 group_by ve summarize

```
counties %>%
  summarize(total_population = sum(population))
```

```
## # A tibble: 1 x 1
##   total_population
##           <dbl>
## 1           315845353
```

```
counties %>%
  summarize(total_population = sum(population),
            average_unemployment = mean(unemployment))
```

```
## # A tibble: 1 x 2
##   total_population average_unemployment
##           <dbl>           <dbl>
## 1           315845353             7.80
```

istenilen düzeye göre hesaplamalar group_by ile yapılır

```
counties %>%
  group_by(state) %>%
  summarize(total_pop = sum(population),
            average_unemployment = sum(unemployment))
```

```
## # A tibble: 50 x 3
##   state      total_pop average_unemployment
##   <chr>          <dbl>           <dbl>
## 1 Alabama      4830620             758.
## 2 Alaska       725461             257.
## 3 Arizona     6641928             180.
## 4 Arkansas     2958208             674.
## 5 California  38421464             626.
## 6 Colorado     5278906             477.
## 7 Connecticut  3593222              65.3
## 8 Delaware      926454             23.8
## 9 Florida     19645772             696.
## 10 Georgia     10006693            1586.
## # ... with 40 more rows
```

```
counties %>%
  group_by(state) %>%
  summarize(total_pop = sum(population),
            average_unemployment = mean(unemployment)) %>%
  arrange(desc(average_unemployment))
```

```
## # A tibble: 50 x 3
##   state      total_pop average_unemployment
##   <chr>      <dbl>          <dbl>
## 1 Mississippi 2988081          12.0
## 2 Arizona      6641928          12.0
## 3 South Carolina 4777576          11.3
## 4 Alabama      4830620          11.3
## 5 California   38421464          10.8
## 6 Nevada       2798636          10.5
## 7 North Carolina 9845333          10.5
## 8 Florida      19645772          10.4
## 9 Georgia      10006693           9.97
## 10 Michigan     9900571           9.96
## # ... with 40 more rows
```

```
# birden fazla değişken düzeyinde grupta
counties %>%
  group_by(state, metro) %>%
  summarize(total_pop = sum(population))
```

```
## # A tibble: 97 x 3
## # Groups:   state [50]
##   state      metro      total_pop
##   <chr>      <chr>      <dbl>
## 1 Alabama    Metro      3671377
## 2 Alabama    Nonmetro   1159243
## 3 Alaska     Metro      494990
## 4 Alaska     Nonmetro   230471
## 5 Arizona    Metro      6295145
## 6 Arizona    Nonmetro   346783
## 7 Arkansas   Metro      1806867
## 8 Arkansas   Nonmetro   1151341
## 9 California Metro      37587429
## 10 California Nonmetro   834035
## # ... with 87 more rows
```

```
# elde edilen veri üzerinden devam edilecekse ungroup kullanılmalı.
# ungroup kullanılmazsa sonradan yapılan işlemler group_by değişkenleri düzeyinde
# devam eder

counties %>%
  group_by(state, metro) %>%
  summarize(total_pop = sum(population)) %>%
  ungroup()
```

```
## # A tibble: 97 x 3
##   state      metro  total_pop
##   <chr>      <chr>      <dbl>
## 1 Alabama    Metro      3671377
## 2 Alabama    Nonmetro   1159243
## 3 Alaska     Metro      494990
## 4 Alaska     Nonmetro   230471
## 5 Arizona    Metro      6295145
## 6 Arizona    Nonmetro   346783
## 7 Arkansas   Metro      1806867
## 8 Arkansas   Nonmetro   1151341
## 9 California Metro      37587429
## 10 California Nonmetro   834035
## # ... with 87 more rows
```

```
# top_n en yüksek ya da en düşük sonuçları listeleme
counties_selected %>%
  group_by(state) %>%
  top_n(1, population) # her eyaletteki en yüksek nüfuslu yer
```

```
## # A tibble: 50 x 4
## # Groups:   state [50]
##   state      county      population unemployment
##   <fct>      <fct>      <dbl>          <dbl>
## 1 Alabama    Jefferson      659026           9.1
## 2 Alaska     Anchorage Municipality 299107           6.7
## 3 Arizona    Maricopa      4018143          7.7
## 4 Arkansas   Pulaski       390463           7.5
## 5 California Los Angeles   10038388         10
## 6 Colorado    El Paso       655024           8.4
## 7 Connecticut Fairfield      939983           9
## 8 Delaware    New Castle    549643           7.4
## 9 Florida     Miami-Dade    2639042          10
## 10 Georgia     Fulton       983903           9.9
## # ... with 40 more rows
```

```
counties_selected %>%
  group_by(state) %>%
  top_n(-1, population) # her eyaletteki en düşük nüfuslu yer
```

```
## # A tibble: 50 x 4
## # Groups:   state [50]
##   state      county      population unemployment
```

```
##      <fct>      <fct>      <dbl>      <dbl>
##  1 Alabama    Greene        8697        20.4
##  2 Alaska     Yakutat City and Borough    643         7.9
##  3 Arizona     Greenlee       9023         10
##  4 Arkansas    Calhoun        5245         7.2
##  5 California  Alpine         1131        10.7
##  6 Colorado    San Juan         606        13.8
##  7 Connecticut Windham      117470         9.3
##  8 Delaware    Kent        169509         8.4
##  9 Florida     Liberty       8295        10.2
## 10 Georgia     Taliaferro     1721        12.1
## # ... with 40 more rows
```

```
counties_selected %>%
group_by(state) %>%
top_n(2, population) # her eyaletteki en yüksek nüfuslu 2 yer
```

```
## # A tibble: 100 x 4
## # Groups:   state [50]
##   state      county      population unemployment
##   <fct>      <fct>      <dbl>      <dbl>
##  1 Alabama    Jefferson    659026         9.1
##  2 Alabama    Mobile      414251         9.8
##  3 Alaska     Anchorage Municipality    299107         6.7
##  4 Alaska     Fairbanks North Star Borough    99705         7.9
##  5 Arizona     Maricopa    4018143         7.7
##  6 Arizona     Pima        998537         10
##  7 Arkansas    Benton      238198         4.2
##  8 Arkansas    Pulaski     390463         7.5
##  9 California Los Angeles  10038388        10
## 10 California San Diego    3223096         8.7
## # ... with 90 more rows
```

```
# summarise_all bütün değişkenler için özetleme yapar
counties_selected %>% summarise_all(nlevels)
```

```
## # A tibble: 1 x 4
##   state county population unemployment
##   <int> <int>      <int>      <int>
## 1     50   1847          0          0
```

```
counties_selected %>%
  select(-county) %>%
  group_by(state) %>%
  summarise_all(mean)
```

```
## # A tibble: 50 x 3
##   state      population unemployment
##   <fct>      <dbl>         <dbl>
## 1 Alabama      72099.          11.3
## 2 Alaska       25909.           9.19
## 3 Arizona     442795.         12.0
## 4 Arkansas     39443.           8.98
## 5 California  662439.         10.8
## 6 Colorado     82483.           7.46
## 7 Connecticut 449153.           8.16
## 8 Delaware    308818           7.93
## 9 Florida     293220.         10.4
## 10 Georgia     62935.           9.97
## # ... with 40 more rows
```

```
# summarise_at belli değişkenler için özetleme yapar
counties_selected %>%
  select(-county) %>%
  group_by(state) %>%
  summarise_at("population",mean)
```

```
## # A tibble: 50 x 2
##   state      population
##   <fct>      <dbl>
## 1 Alabama      72099.
## 2 Alaska       25909.
## 3 Arizona     442795.
## 4 Arkansas     39443.
## 5 California  662439.
## 6 Colorado     82483.
## 7 Connecticut 449153.
## 8 Delaware    308818
## 9 Florida     293220.
## 10 Georgia     62935.
## # ... with 40 more rows
```

```
# summarise_if ile koşula göre özetleme yapar
counties_selected %>%
  summarize_if(is.numeric, mean, na.rm = TRUE)
```

```
## # A tibble: 1 x 2
##   population unemployment
##   <dbl>         <dbl>
## 1    100652.         7.80
```

10 case when

```
# case_when ile yeniden kodlama yapılabilir.
# gelir değişkenini sınıflandıralım
summary(counties$income)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  19328  38827   45113   46832  52249  123453
```

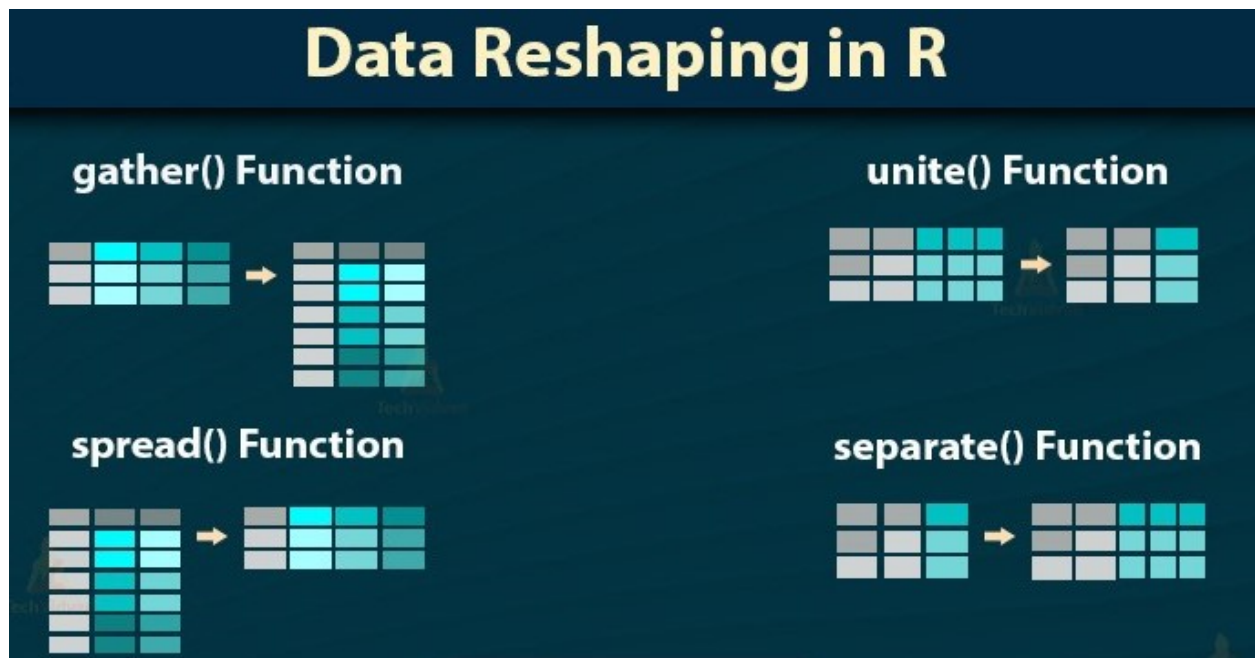
```
counties_income <- counties %>%
  select(state, county, income) %>%
  mutate(income_new= case_when(
    between(income, 19328, 38827) ~ "low",
    between(income, 38828, 52249) ~ "medium",
    income > 52249 ~ "high")
  )
counties_income
```

```
## # A tibble: 3,138 x 4
##   state  county  income income_new
##   <chr>  <chr>    <dbl> <chr>
## 1 Alabama Autauga    51281 medium
## 2 Alabama Baldwin   50254 medium
## 3 Alabama Barbour   32964 low
## 4 Alabama Bibb      38678 low
## 5 Alabama Blount    45813 medium
## 6 Alabama Bullock   31938 low
## 7 Alabama Butler    32229 low
## 8 Alabama Calhoun   41703 medium
## 9 Alabama Chambers  34177 low
## 10 Alabama Cherokee 36296 low
## # ... with 3,128 more rows
```

```
table(counties_income$income_new)
```

```
##
##   high    low medium
##    785    785   1568
```

11 reshaping



```
library(tidyr)

# gather tabloyu yatay formattan dikey formata dönüştürür (transpose).
```

```
counties_gender <- counties %>%
  select(state,county,men,women) %>%
  gather(key="gender",value = "population",-c(state,county))
counties_gender
```

```
## # A tibble: 6,276 x 4
##   state  county  gender population
##   <chr>  <chr>   <chr>         <dbl>
## 1 Alabama Autauga  men           26745
## 2 Alabama Baldwin men           95314
## 3 Alabama Barbour men           14497
```



```
## 4 Alabama Bibb      men      12073
## 5 Alabama Blount   men      28512
## 6 Alabama Bullock  men       5660
## 7 Alabama Butler   men      9502
## 8 Alabama Calhoun  men     56274
## 9 Alabama Chambers men     16258
## 10 Alabama Cherokee men    12975
## # ... with 6,266 more rows
```

```
counties_race <- counties %>%
  select(state,county,hispanic:pacific) %>%
  gather(key="race",value = "ratio",-c(state,county))
counties_race
```

```
## # A tibble: 18,828 x 4
##   state  county  race    ratio
##   <chr>  <chr>   <chr>  <dbl>
## 1 Alabama Autauga  hispanic  2.6
## 2 Alabama Baldwin hispanic  4.5
## 3 Alabama Barbour hispanic  4.6
## 4 Alabama Bibb    hispanic  2.2
## 5 Alabama Blount  hispanic  8.6
## 6 Alabama Bullock hispanic  4.4
## 7 Alabama Butler  hispanic  1.2
## 8 Alabama Calhoun hispanic  3.5
## 9 Alabama Chambers hispanic  0.4
## 10 Alabama Cherokee hispanic  1.5
## # ... with 18,818 more rows
```

spread tabloyu dikey formattan yatay formata dönüştürür (transpose).

```
counties_gender_hor <- counties_gender %>%
  spread(key=gender,value = population)
counties_gender_hor
```

```
## # A tibble: 3,138 x 4
##   state  county  men women
##   <chr>  <chr>   <dbl> <dbl>
## 1 Alabama Autauga  26745 28476
## 2 Alabama Baldwin 95314 99807
## 3 Alabama Barbour 14497 12435
## 4 Alabama Bibb    12073 10531
## 5 Alabama Blount  28512 29198
```

```
## 6 Alabama Bullock 5660 5018
## 7 Alabama Butler 9502 10852
## 8 Alabama Calhoun 56274 60374
## 9 Alabama Chambers 16258 17821
## 10 Alabama Cherokee 12975 13033
## # ... with 3,128 more rows
```

```
# unite
```

```
counties_unite <- counties %>%
  select(state:population) %>%
  unite("region_metro",c(region,metro))
counties_unite
```

```
## # A tibble: 3,138 x 4
##   state county region_metro population
##   <chr> <chr> <chr> <dbl>
## 1 Alabama Autauga South_Metro 55221
## 2 Alabama Baldwin South_Metro 195121
## 3 Alabama Barbour South_Nonmetro 26932
## 4 Alabama Bibb South_Metro 22604
## 5 Alabama Blount South_Metro 57710
## 6 Alabama Bullock South_Nonmetro 10678
## 7 Alabama Butler South_Nonmetro 20354
## 8 Alabama Calhoun South_Metro 116648
## 9 Alabama Chambers South_Nonmetro 34079
## 10 Alabama Cherokee South_Nonmetro 26008
## # ... with 3,128 more rows
```

```
counties_unite <- counties %>%
  select(state:population) %>%
  unite("region_metro",c(region,metro),sep = "-",remove = FALSE)
counties_unite
```

```
## # A tibble: 3,138 x 6
##   state county region_metro region metro population
##   <chr> <chr> <chr> <chr> <chr> <dbl>
## 1 Alabama Autauga South-Metro South Metro 55221
## 2 Alabama Baldwin South-Metro South Metro 195121
## 3 Alabama Barbour South-Nonmetro South Nonmetro 26932
## 4 Alabama Bibb South-Metro South Metro 22604
## 5 Alabama Blount South-Metro South Metro 57710
## 6 Alabama Bullock South-Nonmetro South Nonmetro 10678
```

```
## 7 Alabama Butler South-Nonmetro South Nonmetro 20354
## 8 Alabama Calhoun South-Metro South Metro 116648
## 9 Alabama Chambers South-Nonmetro South Nonmetro 34079
## 10 Alabama Cherokee South-Nonmetro South Nonmetro 26008
## # ... with 3,128 more rows
```

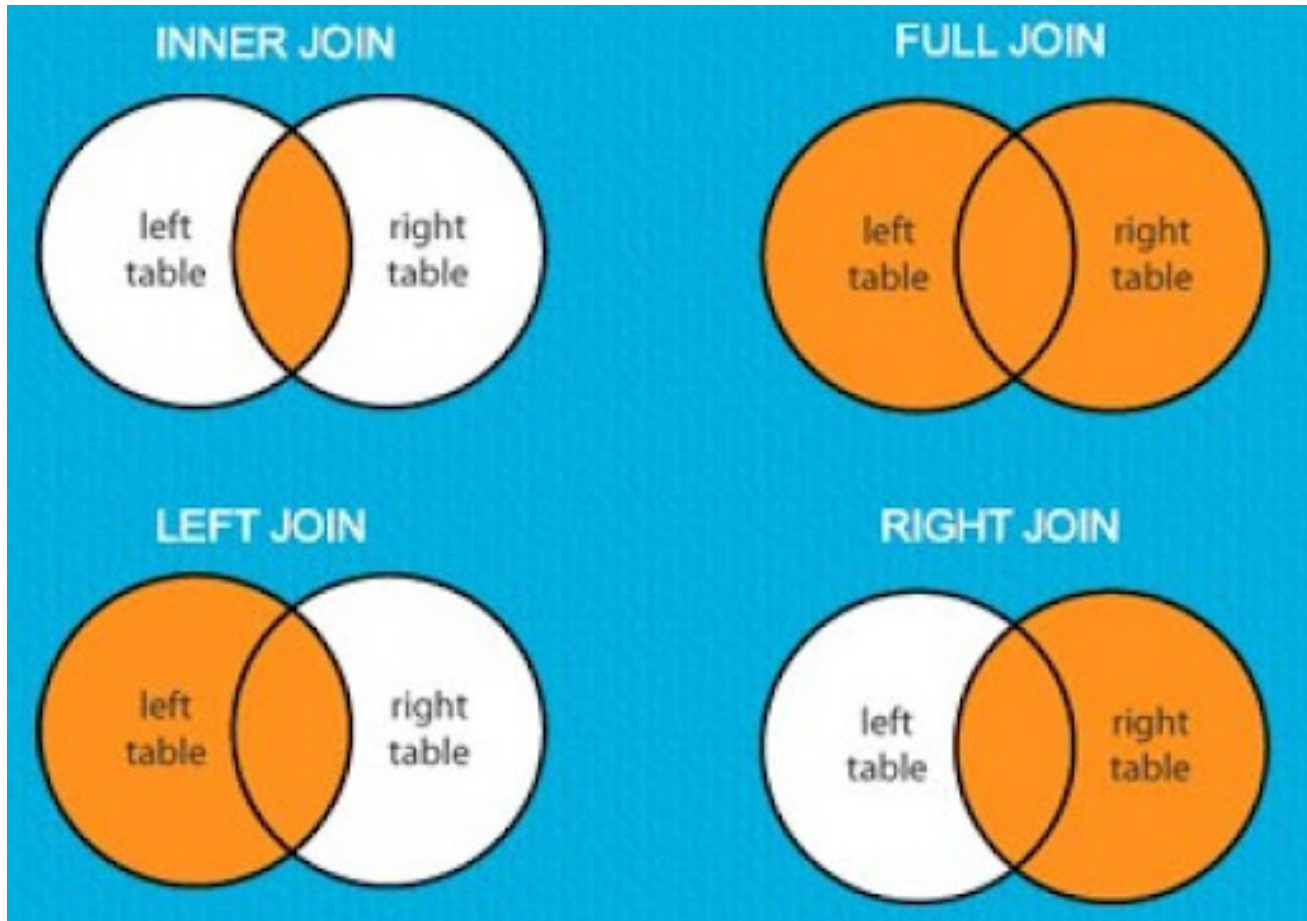
```
counties_unite <- counties %>%
  select(state:population) %>%
  unite("region_metro",c(region,metro),sep = "-")
counties_unite
```

```
## # A tibble: 3,138 x 4
##   state   county region_metro population
##   <chr>   <chr>   <chr>         <dbl>
## 1 Alabama Autauga South-Metro      55221
## 2 Alabama Baldwin South-Metro     195121
## 3 Alabama Barbour South-Nonmetro    26932
## 4 Alabama Bibb South-Metro      22604
## 5 Alabama Blount South-Metro      57710
## 6 Alabama Bullock South-Nonmetro    10678
## 7 Alabama Butler South-Nonmetro    20354
## 8 Alabama Calhoun South-Metro     116648
## 9 Alabama Chambers South-Nonmetro    34079
## 10 Alabama Cherokee South-Nonmetro    26008
## # ... with 3,128 more rows
```

```
# separate
counties_separate <- counties_unite %>%
  separate(region_metro,c("region2","metro2"),sep = "-")
counties_separate
```

```
## # A tibble: 3,138 x 5
##   state   county region2 metro2 population
##   <chr>   <chr>   <chr>   <chr>         <dbl>
## 1 Alabama Autauga South Metro      55221
## 2 Alabama Baldwin South Metro     195121
## 3 Alabama Barbour South Nonmetro    26932
## 4 Alabama Bibb South Metro      22604
## 5 Alabama Blount South Metro      57710
## 6 Alabama Bullock South Nonmetro    10678
## 7 Alabama Butler South Nonmetro    20354
## 8 Alabama Calhoun South Metro     116648
## 9 Alabama Chambers South Nonmetro    34079
## 10 Alabama Cherokee South Nonmetro    26008
## # ... with 3,128 more rows
```

12 join



- Inner join: `merge(df1, df2)`
- Outer join: `merge(x = df1, y = df2, by = "no", all = TRUE)`
- Left outer: `merge(x = df1, y = df2, by = "no", all.x = TRUE)`
- Right outer: `merge(x = df1, y = df2, by = "no", all.y = TRUE)`

```
set.seed(1234)
# data frame 1

df_1 <- data.frame(no = c(1:10),
                   math = round(rnorm(10,70,20)),
                   stat = round(rnorm(10,60,30)))

df_1
```

```
##      no math stat
```

```
## 1 1 46 46
## 2 2 76 30
## 3 3 92 37
## 4 4 23 62
## 5 5 79 89
## 6 6 80 57
## 7 7 59 45
## 8 8 59 33
## 9 9 59 35
## 10 10 52 132
```

```
# data frame 2
```

```
df_2 <- data.frame(no = c(2, 4, 6, 7, 11,12),
                    city = c("Ankara","İzmir","İzmir","İstanbul","Adana","Trabzon"),
                    age =c(20,18,22,19,25,21))
```

```
df_2
```

```
## no city age
## 1 2 Ankara 20
## 2 4 İzmir 18
## 3 6 İzmir 22
## 4 7 İstanbul 19
## 5 11 Adana 25
## 6 12 Trabzon 21
```

```
# Inner Join - her iki tablonun ortak kayıtlarını eşleştirir
```

```
merge(x=df_1,y=df_2) # merge ile
```

```
## no math stat city age
## 1 2 76 30 Ankara 20
## 2 4 23 62 İzmir 18
## 3 6 80 57 İzmir 22
## 4 7 59 45 İstanbul 19
```

```
merge(x=df_1,y=df_2,by="no") # merge ile
```

```
## no math stat city age
## 1 2 76 30 Ankara 20
## 2 4 23 62 İzmir 18
## 3 6 80 57 İzmir 22
## 4 7 59 45 İstanbul 19
```

```
merge(x=df_1,y=df_2,by=c("no"="no")) # merge ile
```

```
##   no math stat      city age
## 1  2   76   30   Ankara  20
## 2  4   23   62     İzmir  18
## 3  6   80   57     İzmir  22
## 4  7   59   45 İstanbul  19
```

```
df_1 %>% inner_join(df_2) # dplyr-inner_join ile
```

```
##   no math stat      city age
## 1  2   76   30   Ankara  20
## 2  4   23   62     İzmir  18
## 3  6   80   57     İzmir  22
## 4  7   59   45 İstanbul  19
```

```
df_1 %>% inner_join(df_2,by=c("no"="no")) #dplyr-inner_join ile
```

```
##   no math stat      city age
## 1  2   76   30   Ankara  20
## 2  4   23   62     İzmir  18
## 3  6   80   57     İzmir  22
## 4  7   59   45 İstanbul  19
```

Left Join - ilk tablonun tamamını ve ikinci tablodan eşleşenleri getirir

```
merge(x=df_1,y=df_2,by="no",all.x = TRUE) # merge ile
```

```
##   no math stat      city age
## 1  1   46   46     <NA>  NA
## 2  2   76   30   Ankara  20
## 3  3   92   37     <NA>  NA
## 4  4   23   62     İzmir  18
## 5  5   79   89     <NA>  NA
## 6  6   80   57     İzmir  22
## 7  7   59   45 İstanbul  19
## 8  8   59   33     <NA>  NA
## 9  9   59   35     <NA>  NA
## 10 10   52  132     <NA>  NA
```

```
df_1 %>% left_join(df_2,by=c("no"="no")) # dplyr-left_join ile
```

```
##      no math stat      city age
## 1    1   46   46      <NA>  NA
## 2    2   76   30    Ankara  20
## 3    3   92   37      <NA>  NA
## 4    4   23   62     İzmir  18
## 5    5   79   89      <NA>  NA
## 6    6   80   57     İzmir  22
## 7    7   59   45 İstanbul  19
## 8    8   59   33      <NA>  NA
## 9    9   59   35      <NA>  NA
## 10  10   52  132      <NA>  NA
```

Right Join -ikinci tablonun tamamını ve ilk tablodan eşleşenleri getirir

```
merge(x=df_1,y=df_2,by="no",all.y = TRUE) # merge ile
```

```
##      no math stat      city age
## 1    2   76   30    Ankara  20
## 2    4   23   62     İzmir  18
## 3    6   80   57     İzmir  22
## 4    7   59   45 İstanbul  19
## 5   11   NA   NA     Adana  25
## 6   12   NA   NA    Trabzon  21
```

```
df_1 %>% right_join(df_2,by=c("no"="no")) # dplyr-right_join ile
```

```
##      no math stat      city age
## 1    2   76   30    Ankara  20
## 2    4   23   62     İzmir  18
## 3    6   80   57     İzmir  22
## 4    7   59   45 İstanbul  19
## 5   11   NA   NA     Adana  25
## 6   12   NA   NA    Trabzon  21
```

Full Join -her iki tablodan bütün kayıtları eşleştirir

```
merge(x=df_1,y=df_2,by="no",all = TRUE) # merge ile
```

```
##      no math stat      city age
```

```
## 1 1 46 46 <NA> NA
## 2 2 76 30 Ankara 20
## 3 3 92 37 <NA> NA
## 4 4 23 62 İzmir 18
## 5 5 79 89 <NA> NA
## 6 6 80 57 İzmir 22
## 7 7 59 45 İstanbul 19
## 8 8 59 33 <NA> NA
## 9 9 59 35 <NA> NA
## 10 10 52 132 <NA> NA
## 11 11 NA NA Adana 25
## 12 12 NA NA Trabzon 21
```

```
df_1 %>% full_join(df_2,by=c("no"="no")) # dplyr-right_join ile
```

```
##      no math stat      city age
## 1 1 46 46 <NA> NA
## 2 2 76 30 Ankara 20
## 3 3 92 37 <NA> NA
## 4 4 23 62 İzmir 18
## 5 5 79 89 <NA> NA
## 6 6 80 57 İzmir 22
## 7 7 59 45 İstanbul 19
## 8 8 59 33 <NA> NA
## 9 9 59 35 <NA> NA
## 10 10 52 132 <NA> NA
## 11 11 NA NA Adana 25
## 12 12 NA NA Trabzon 21
```