Date Tidying with tidyr

Adejumo Ridwan Suleiman

2022-09-26

Installing and loading tidyr

```
#install.package("tidyr)
library(tidyr)
```

Reshaping with Pivot longer

```
head(relig_income)
```

```
## # A tibble: 6 x 11
##   religion       `<$10k` `$10-2~1 $20-3~2 $30-4~3 $40-5~4 $50-7~5 $75-1~6 $100-~7
##   <chr>            <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 Agnostic            27      34      60      81      76     137     122     109
## 2 Atheist             12      27      37      52      35      70      73      59
## 3 Buddhist            27      21      30      34      33      58      62      39
## 4 Catholic           418     617     732     670     638    1116     949     792
## 5 Don't know/re~      15      14      15      11      10      35      21      17
## 6 Evangelical P~     575     869    1064     982     881    1486     949     723
## # ... with 2 more variables: `>150k` <dbl>, `Don't know/refused` <dbl>, and
## #   abbreviated variable names 1: `$10-20k`, 2: `$20-30k`, 3: `$30-40k`,
## #   4: `$40-50k`, 5: `$50-75k`, 6: `$75-100k`, 7: `$100-150k`
```

```
relig_income %>%
  pivot_longer(!religion,
               names_to = "income",
               values_to = "count")
```

```
## # A tibble: 180 x 3
##    religion income             count
##    <chr>    <chr>              <dbl>
##  1 Agnostic <$10k                 27
##  2 Agnostic $10-20k               34
##  3 Agnostic $20-30k               60
##  4 Agnostic $30-40k               81
##  5 Agnostic $40-50k               76
##  6 Agnostic $50-75k              137
```

```
##  7 Agnostic $75-100k             122
##  8 Agnostic $100-150k            109
##  9 Agnostic >150k                 84
## 10 Agnostic Don't know/refused    96
## # ... with 170 more rows
```

Reshaping with Pivot wider

```
head(us_rent_income)
```

```
## # A tibble: 6 x 5
##   GEOID NAME    variable estimate   moe
##   <chr> <chr>   <chr>       <dbl> <dbl>
## 1 01    Alabama income      24476   136
## 2 01    Alabama rent          747     3
## 3 02    Alaska  income      32940   508
## 4 02    Alaska  rent         1200    13
## 5 04    Arizona income      27517   148
## 6 04    Arizona rent          972     4
```

```
us_rent_income %>%
  pivot_wider(
    names_from = variable,
    values_from = estimate,
    values_fill = 0
  )
```

```
## # A tibble: 104 x 5
##    GEOID NAME         moe income  rent
##    <chr> <chr>      <dbl>  <dbl> <dbl>
##  1 01    Alabama      136  24476     0
##  2 01    Alabama        3      0   747
##  3 02    Alaska       508  32940     0
##  4 02    Alaska        13      0  1200
##  5 04    Arizona      148  27517     0
##  6 04    Arizona        4      0   972
##  7 05    Arkansas     165  23789     0
##  8 05    Arkansas       5      0   709
##  9 06    California   109  29454     0
## 10 06    California     3      0  1358
## # ... with 94 more rows
```

#Separating

```
head(table3)
```

```
## # A tibble: 6 x 3
##   country      year rate
##   <chr>       <int> <chr>
```

```
## 1 Afghanistan  1999 745/19987071
## 2 Afghanistan  2000 2666/20595360
## 3 Brazil       1999 37737/172006362
## 4 Brazil       2000 80488/174504898
## 5 China        1999 212258/1272915272
## 6 China        2000 213766/1280428583
```

```
table3 %>%
  separate(col = rate,
           into = c("cases", "population"),
           sep = "/")
```

```
## # A tibble: 6 x 4
##   country      year cases  population
##   <chr>       <int> <chr>  <chr>
## 1 Afghanistan  1999 745    19987071
## 2 Afghanistan  2000 2666   20595360
## 3 Brazil       1999 37737  172006362
## 4 Brazil       2000 80488  174504898
## 5 China        1999 212258 1272915272
## 6 China        2000 213766 1280428583
```

#Uniting

```
head(table5)
```

```
## # A tibble: 6 x 4
##   country     century year  rate
##   <chr>       <chr>   <chr> <chr>
## 1 Afghanistan 19      99    745/19987071
## 2 Afghanistan 20      00    2666/20595360
## 3 Brazil      19      99    37737/172006362
## 4 Brazil      20      00    80488/174504898
## 5 China       19      99    212258/1272915272
## 6 China       20      00    213766/1280428583
```

```
table5 %>%
  unite(col = new,
        century, year)
```

```
## # A tibble: 6 x 3
##   country     new   rate
##   <chr>       <chr> <chr>
## 1 Afghanistan 19_99 745/19987071
## 2 Afghanistan 20_00 2666/20595360
## 3 Brazil      19_99 37737/172006362
## 4 Brazil      20_00 80488/174504898
## 5 China       19_99 212258/1272915272
## 6 China       20_00 213766/1280428583
```

Exercises

Question 1

Which of the follwing is an argument in the function `pivot_longer`?
1. names_to - ANS
2. value_to
3. names_from
4. values_from
5. name_to

Question 2

Which of the following is an argument in the function `pivot_wider`?
1. names_to
2. value_to
3. names_from - ANS
4. value_from
5. name_to

Qustion 3

Which of the following arguments will fill missing values after using a `pivot_wider` on a data set?
1. value_fil
2. values_fill - ANS
3. fill
4. fill_na
5. na_fill