```
(base) Vasu's MacBook Pro:~ vasugoel$ r

R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> path <- system.file('extdata', package='dslabs')
> path
[1] "/Library/Frameworks/R.framework/Versions/3.6/Resources/library/dslabs/extdata"
> list.files(path)
[1] "2010_bigfive_regents.xls"
[2] "carbon_emissions.csv"
[3] "fertility-two-countries-example.csv"
[4] "HRlist2.txt"
[5] "life-expectancy-and-fertility-two-countries-example.csv"
[6] "murders.csv"
[7] "olive.csv"
[8] "RD-Mortality-Report_2015-18-180531.pdf"
[9] "ssa-death-probability.csv"
> filename <- 'life-expectancy-and-fertility-two-countries-example.csv'
> filepath <- file.path(path, filename)
> filepath
[1] "/Library/Frameworks/R.framework/Versions/3.6/Resources/library/dslabs/extdata/life-expectancy-and-fertility-two-countries-example.csv"
> library(tidyverse)
── Attaching packages ─────────────────────────────────────── tidyverse 1.2.1 ──
✔ ggplot2 3.2.1     ✔ purrr   0.3.2
✔ tibble  2.1.3     ✔ dplyr   0.8.3
✔ tidyr   0.8.3     ✔ stringr 1.4.0
✔ readr   1.3.1     ✔ forcats 0.4.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
> raw_data <- read_csv(filepath)
Parsed with column specification:
cols(
  .default = col_double(),
  country = col_character()
)
See spec(...) for full column specifications.
> raw_data
# A tibble: 2 x 113
  country `1960_fertility` `1960_life_expe… `1961_fertility` `1961_life_expe…
  <chr>              <dbl>            <dbl>            <dbl>            <dbl>
1 Germany             2.41             69.3             2.44             69.8
2 South …             6.16             53.0             5.99             53.8
# … with 108 more variables: `1962_fertility` <dbl>,
#   `1962_life_expectancy` <dbl>, `1963_fertility` <dbl>,
#   `1963_life_expectancy` <dbl>, `1964_fertility` <dbl>,
#   `1964_life_expectancy` <dbl>, `1965_fertility` <dbl>,
#   `1965_life_expectancy` <dbl>, `1966_fertility` <dbl>,
#   `1966_life_expectancy` <dbl>, `1967_fertility` <dbl>,
#   `1967_life_expectancy` <dbl>, `1968_fertility` <dbl>,
#   `1968_life_expectancy` <dbl>, `1969_fertility` <dbl>,
#   `1969_life_expectancy` <dbl>, `1970_fertility` <dbl>,
#   `1970_life_expectancy` <dbl>, `1971_fertility` <dbl>,
#   `1971_life_expectancy` <dbl>, `1972_fertility` <dbl>,
#   `1972_life_expectancy` <dbl>, `1973_fertility` <dbl>,
#   `1973_life_expectancy` <dbl>, `1974_fertility` <dbl>,
#   `1974_life_expectancy` <dbl>, `1975_fertility` <dbl>,
#   `1975_life_expectancy` <dbl>, `1976_fertility` <dbl>,
#   `1976_life_expectancy` <dbl>, `1977_fertility` <dbl>,
#   `1977_life_expectancy` <dbl>, `1978_fertility` <dbl>,
#   `1978_life_expectancy` <dbl>, `1979_fertility` <dbl>,
#   `1979_life_expectancy` <dbl>, `1980_fertility` <dbl>,
#   `1980_life_expectancy` <dbl>, `1981_fertility` <dbl>,
#   `1981_life_expectancy` <dbl>, `1982_fertility` <dbl>,
#   `1982_life_expectancy` <dbl>, `1983_fertility` <dbl>,
#   `1983_life_expectancy` <dbl>, `1984_fertility` <dbl>,
#   `1984_life_expectancy` <dbl>, `1985_fertility` <dbl>,
#   `1985_life_expectancy` <dbl>, `1986_fertility` <dbl>,
#   `1986_life_expectancy` <dbl>, `1987_fertility` <dbl>,
#   `1987_life_expectancy` <dbl>, `1988_fertility` <dbl>,
#   `1988_life_expectancy` <dbl>, `1989_fertility` <dbl>,
#   `1989_life_expectancy` <dbl>, `1990_fertility` <dbl>,
#   `1990_life_expectancy` <dbl>, `1991_fertility` <dbl>,
#   `1991_life_expectancy` <dbl>, `1992_fertility` <dbl>,
#   `1992_life_expectancy` <dbl>, `1993_fertility` <dbl>,
#   `1993_life_expectancy` <dbl>, `1994_fertility` <dbl>,
#   `1994_life_expectancy` <dbl>, `1995_fertility` <dbl>,
#   `1995_life_expectancy` <dbl>, `1996_fertility` <dbl>,
#   `1996_life_expectancy` <dbl>, `1997_fertility` <dbl>,
#   `1997_life_expectancy` <dbl>, `1998_fertility` <dbl>,
#   `1998_life_expectancy` <dbl>, `1999_fertility` <dbl>,
#   `1999_life_expectancy` <dbl>, `2000_fertility` <dbl>,
```

```
#   `2000_life_expectancy` <dbl>, `2001_fertility` <dbl>,
#   `2001_life_expectancy` <dbl>, `2002_fertility` <dbl>,
#   `2002_life_expectancy` <dbl>, `2003_fertility` <dbl>,
#   `2003_life_expectancy` <dbl>, `2004_fertility` <dbl>,
#   `2004_life_expectancy` <dbl>, `2005_fertility` <dbl>,
#   `2005_life_expectancy` <dbl>, `2006_fertility` <dbl>,
#   `2006_life_expectancy` <dbl>, `2007_fertility` <dbl>,
#   `2007_life_expectancy` <dbl>, `2008_fertility` <dbl>,
#   `2008_life_expectancy` <dbl>, `2009_fertility` <dbl>,
#   `2009_life_expectancy` <dbl>, `2010_fertility` <dbl>,
#   `2010_life_expectancy` <dbl>, `2011_fertility` <dbl>,
#   `2011_life_expectancy` <dbl>, …
> select(raw_data, 1:5 )
# A tibble: 2 x 5
  country  `1960_fertility` `1960_life_expec… `1961_fertility` `1961_life_expec…
  <chr>               <dbl>              <dbl>            <dbl>             <dbl>
1 Germany              2.41               69.3             2.44              69.8
2 South K…             6.16               53.0             5.99              53.8
> dat <- raw_data %>% gather(key, value, -country)
> dat
# A tibble: 224 x 3
   country     key                  value
   <chr>       <chr>                <dbl>
 1 Germany     1960_fertility        2.41
 2 South Korea 1960_fertility        6.16
 3 Germany     1960_life_expectancy 69.3
 4 South Korea 1960_life_expectancy 53.0
 5 Germany     1961_fertility        2.44
 6 South Korea 1961_fertility        5.99
 7 Germany     1961_life_expectancy 69.8
 8 South Korea 1961_life_expectancy 53.8
 9 Germany     1962_fertility        2.47
10 South Korea 1962_fertility        5.79
# … with 214 more rows
> dat %>% separate(key, c('year', 'variable_name'), sep='_')
# A tibble: 224 x 4
   country     year  variable_name value
   <chr>       <chr> <chr>         <dbl>
 1 Germany     1960  fertility      2.41
 2 South Korea 1960  fertility      6.16
 3 Germany     1960  life          69.3
 4 South Korea 1960  life          53.0
 5 Germany     1961  fertility      2.44
 6 South Korea 1961  fertility      5.99
 7 Germany     1961  life          69.8
 8 South Korea 1961  life          53.8
 9 Germany     1962  fertility      2.47
10 South Korea 1962  fertility      5.79
# … with 214 more rows
Warning message:
Expected 2 pieces. Additional pieces discarded in 112 rows [3, 4, 7, 8, 11, 12, 15, 16, 19, 20, 23, 24, 27, 28, 31, 32, 35, 36, 39, 40, ...].
> dat %>% separate(key, c('year', 'variable_name'), sep='_', extra='merge')
# A tibble: 224 x 4
   country     year  variable_name   value
   <chr>       <chr> <chr>           <dbl>
 1 Germany     1960  fertility        2.41
 2 South Korea 1960  fertility        6.16
 3 Germany     1960  life_expectancy 69.3
 4 South Korea 1960  life_expectancy 53.0
 5 Germany     1961  fertility        2.44
 6 South Korea 1961  fertility        5.99
 7 Germany     1961  life_expectancy 69.8
 8 South Korea 1961  life_expectancy 53.8
 9 Germany     1962  fertility        2.47
10 South Korea 1962  fertility        5.79
# … with 214 more rows
> dat %>% separate(key, c('year', 'variable_name'), sep='_', extra='merge') %>% spread(variable_name, value)
# A tibble: 112 x 4
   country year  fertility life_expectancy
   <chr>   <chr>     <dbl>           <dbl>
 1 Germany 1960       2.41            69.3
 2 Germany 1961       2.44            69.8
 3 Germany 1962       2.47            70.0
 4 Germany 1963       2.49            70.1
 5 Germany 1964       2.49            70.7
 6 Germany 1965       2.48            70.6
 7 Germany 1966       2.44            70.8
 8 Germany 1967       2.37            71.0
 9 Germany 1968       2.28            70.6
10 Germany 1969       2.17            70.5
# … with 102 more rows
>
>
>
> # example of unite() function
> dat %>% separate(key, c('year', 'variable_name'), sep='_')
# A tibble: 224 x 4
   country     year  variable_name value
   <chr>       <chr> <chr>         <dbl>
 1 Germany     1960  fertility      2.41
 2 South Korea 1960  fertility      6.16
 3 Germany     1960  life          69.3
 4 South Korea 1960  life          53.0
 5 Germany     1961  fertility      2.44
 6 South Korea 1961  fertility      5.99
 7 Germany     1961  life          69.8
 8 South Korea 1961  life          53.8
 9 Germany     1962  fertility      2.47
```

```
10 South Korea 1962  fertility      5.79
# … with 214 more rows
Warning message:
Expected 2 pieces. Additional pieces discarded in 112 rows [3, 4, 7, 8, 11, 12, 15, 16, 19, 20, 23, 24, 27, 28, 31, 32, 35, 36, 39, 40, ...].
> dat %>% separate(key, c('year', 'first_variable_name', 'second_variable_name'), sep='_', fill='right')
# A tibble: 224 x 5
   country     year  first_variable_name second_variable_name value
   <chr>       <chr> <chr>               <chr>                <dbl>
 1 Germany     1960  fertility           NA                    2.41
 2 South Korea 1960  fertility           NA                    6.16
 3 Germany     1960  life                expectancy           69.3
 4 South Korea 1960  life                expectancy           53.0
 5 Germany     1961  fertility           NA                    2.44
 6 South Korea 1961  fertility           NA                    5.99
 7 Germany     1961  life                expectancy           69.8
 8 South Korea 1961  life                expectancy           53.8
 9 Germany     1962  fertility           NA                    2.47
10 South Korea 1962  fertility           NA                    5.79
# … with 214 more rows
> dat %>% separate(key, c('year', 'first_variable_name', 'second_variable_name'), sep='_', fill='right') %>% unite(variable_name, first_variable_name, second_
variable_name, sep='_') %>% spread(variable_name, value) %>% rename(fertility=fertility_NA)
# A tibble: 112 x 4
   country year  fertility life_expectancy
   <chr>   <chr>     <dbl>           <dbl>
 1 Germany 1960       2.41            69.3
 2 Germany 1961       2.44            69.8
 3 Germany 1962       2.47            70.0
 4 Germany 1963       2.49            70.1
 5 Germany 1964       2.49            70.7
 6 Germany 1965       2.48            70.6
 7 Germany 1966       2.44            70.8
 8 Germany 1967       2.37            71.0
 9 Germany 1968       2.28            70.6
10 Germany 1969       2.17            70.5
# … with 102 more rows
>
```