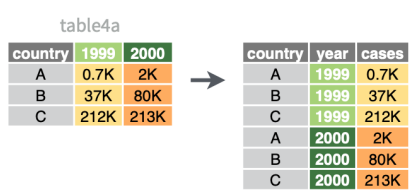
COMP 112: DATA WRANGLING SUMMARY SHEET
Sam Kennedy

- Pipe operator |>
  - Passes object to function
  - Tidyverse syntax: %>%

- Wrangling verbs
  - Select
    - Select a subset of columns
  - Mutate
    - Mutate existing columns or add new ones
  - Filter
    - Filter subset of rows based on values of 1+ columns
    - Boolean condition
  - Arrange
    - Arrange rows based on values of 1+ columns
  - Summarize
    - Calculate numerical summary of a column
    - Ex: mean, median, max, min
  - Group_by
    - Group rows based on values of 1+ columns

- Reshaping
  - Pivot longer
    - Join several columns into two columns

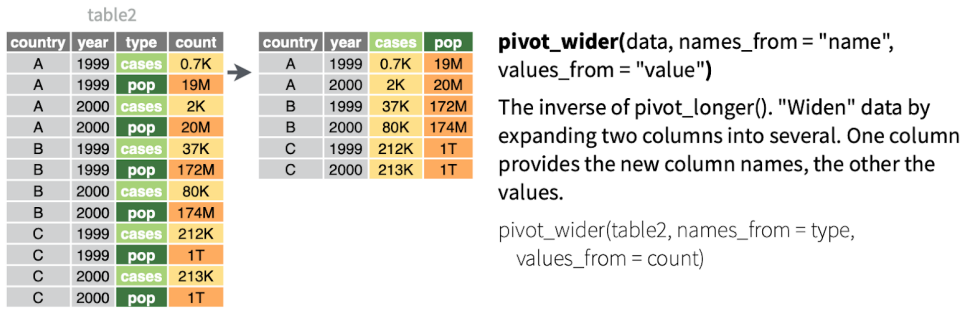


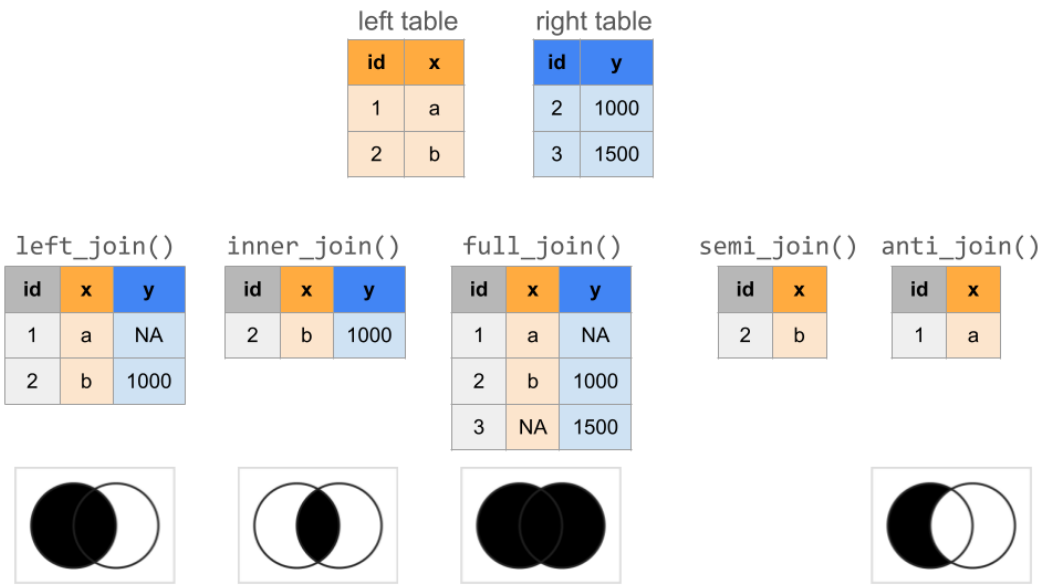
pivot_longer(data, cols, names_to = "name", values_to = "value", values_drop_na = FALSE**)**

"Lengthen" data by collapsing several columns into two. Column names move to a new names_to column and values to a new values_to column.

pivot_longer(table4a, cols = 2:3, names_to ="year", values_to = "cases")

    -
  - Pivot wider
    - Expand two columns into several columns

- 

- Joining datasets
  - Left_join()
    - Keeps all observations from the left, drops those on the right without a match in the left
  - Inner_join()
    - Keeps only the observations from the left with a match in the right.
  - Full_join()
    - Keeps all observations from both the left and the right
  - 

- Factors
  - Fct_recode
    - Change labels of factor
  - Fct_relevel
    - Change char variables to factors
    - Specify meaningful order for factor variables
    - Reorder factor variables
  - Fct_reorder

        - Reorder factor variables using another variable

- Strings
    - str_replace()
        - finds the first part of x that matches the pattern and replaces it with replacement
    - str_replace_all()
        - finds all instances in x that matches the pattern and replaces it with replacement
    - str_to_lower()
        - Converts uppercase characters to lowercase
    - str_sub()
        - Only keeps subset of characters in string from 'start' value to 'end' value
    - str_length()
        - Returns number of characters in string
    - str_detect()
        - Returns TRUE if string contains given pattern, otherwise returns FALSE

Essential Functions

The `stringr` package within `tidyverse` contains lots of functions to help process strings. We'll focus on the most common. Letting `x` be a string variable...

function	arguments	returns
`str_replace()`	`x, pattern, replacement`	a modified string
`str_replace_all()`	`x, pattern, replacement`	a modified string
`str_to_lower()`	`x`	a modified string
`str_sub()`	`x, start, end`	a modified string
`str_length()`	`x`	a number
`str_detect()`	`x, pattern`	TRUE/FALSE

    -