

Week-9

Tan Ee Xuan

2023-10-18

Slide 8

```
library(tidyverse)
```

```
## — Attaching packages — tidyverse 1.3.2 —
## ✓ ggplot2 3.3.6      ✓ purrr 0.3.4
## ✓ tibble 3.1.8       ✓ dplyr 1.0.9
## ✓ tidyr 1.2.0        ✓ stringr 1.4.0
## ✓ readr 2.1.2        ✓ forcats 0.5.1
## — Conflicts — tidyverse_conflicts() —
## * dplyr::filter() masks stats::filter()
## * dplyr::lag()     masks stats::lag()
```

```
tidydata <- tribble(
  ~country, ~year, ~cases, ~population,
  "Afghanistan", 1999, 745, 19987071,
  "Afghanistan", 2000, 2666, 20595360,
  "Brazil", 1999, 37737, 172006362,
  "Brazil", 2000, 80488, 174504898,
  "China", 1999, 212258, 1272915272,
  "China", 2000, 213766, 1280428583
)
```

```
tidydata
```

```
## # A tibble: 6 × 4
##   country      year cases population
##   <chr>      <dbl> <dbl>      <dbl>
## 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
```

```
nontidydata <- tribble(
  ~country, ~year, ~rate,
  "Afghanistan", 1999, "745/19987071",
  "Afghanistan", 2000, "2666/20595360",
  "Brazil", 1999, "37737/172006362",
  "Brazil", 2000, "80488/174504898",
  "China", 1999, "212258/1272915272",
  "China", 2000, "213766/1280428583")
```

```
nontidydata
```

```
## # A tibble: 6 × 3
##   country      year rate
##   <chr>      <dbl> <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
```

Slide 11

```
nontidydata
```

```
## # A tibble: 6 × 3
##   country      year rate
##   <chr>      <dbl> <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
```

```
tidieddata <- nontidydata %>%
  separate(rate, into = c("cases",
                        "population"),
           sep = "/")
```

```
tidieddata
```

```
## # A tibble: 6 × 4
##   country      year cases population
##   <chr>      <dbl> <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
```

Slide 12

```
newtidieddata <- tidieddata %>%
  pivot_longer(
    cols = cases:population,
    names_to = "measurement",
    values_to = "value"
  )
```

```
newtidieddata
```

```
## # A tibble: 12 × 4
##   country      year measurement value
##   <chr>      <dbl> <chr>      <chr>
## 1 Afghanistan 1999 cases      745
## 2 Afghanistan 1999 population 19987071
## 3 Afghanistan 2000 cases      2666
## 4 Afghanistan 2000 population 20595360
## 5 Brazil      1999 cases      37737
## 6 Brazil      1999 population 172006362
## 7 Brazil      2000 cases      80488
## 8 Brazil      2000 population 174504898
## 9 China       1999 cases      212258
## 10 China      1999 population 1272915272
## 11 China      2000 cases      213766
## 12 China      2000 population 1280428583
```

Slide 14

```
df <- tribble(
  ~id, ~bp1, ~bp2,
  "A", 100, 120,
  "B", 140, 115,
  "C", 120, 125
)
```

```
df
```

```
## # A tibble: 3 × 3
##   id      bp1  bp2
##   <chr> <dbl> <dbl>
## 1 A      100  120
## 2 B      140  115
## 3 C      120  125
```

```
df %>%
  pivot_longer(
    cols = bp1:bp2,
    names_to = "measurement",
    values_to = "value"
  )
```

```
## # A tibble: 6 × 3
##   id      measurement value
##   <chr> <chr>      <dbl>
## 1 A      bp1          100
## 2 A      bp2          120
## 3 B      bp1          140
## 4 B      bp2          115
## 5 C      bp1          120
## 6 C      bp2          125
```

Slide 18

```
newtidieddata
```

```
## # A tibble: 12 × 4
##   country      year measurement value
##   <chr>      <dbl> <chr>      <chr>
## 1 Afghanistan 1999 cases      745
## 2 Afghanistan 1999 population 19987071
## 3 Afghanistan 2000 cases      2666
## 4 Afghanistan 2000 population 20595360
## 5 Brazil       1999 cases      37737
## 6 Brazil       1999 population 172006362
## 7 Brazil       2000 cases      80488
## 8 Brazil       2000 population 174504898
## 9 China        1999 cases      212258
## 10 China       1999 population 1272915272
## 11 China       2000 cases      213766
## 12 China       2000 population 1280428583
```

```
newtidieddata %>%
  pivot_wider(names_from = "measurement",
              values_from = "value")
```

```
## # A tibble: 6 × 4
##   country      year cases population
##   <chr>      <dbl> <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
```

Slide 19

```
df <- tribble(
  ~id, ~measurement, ~value,
  "A", "bp1", 100,
  "B", "bp1", 140,
  "B", "bp2", 115,
  "A", "bp2", 120,
  "A", "bp3", 105
)

df
```

```
## # A tibble: 5 × 3
##   id      measurement value
##   <chr> <chr>         <dbl>
## 1 A      bp1           100
## 2 B      bp1           140
## 3 B      bp2           115
## 4 A      bp2           120
## 5 A      bp3           105
```

```
df %>%
  pivot_wider(
    names_from = measurement,
    values_from = value
  )
```

```
## # A tibble: 2 × 4
##   id      bp1      bp2      bp3
##   <chr> <dbl> <dbl> <dbl>
## 1 A      100    120    105
## 2 B      140    115     NA
```