

In_Class_9/15

Austin Dollar

9/15/2021

Tidy Data Assignment

12.2.1 Q1

Using prose, describe how the variables and observations are organised in each of the sample tables.

Variables in tables, are similar to the names of columns when thinking of an excel spreadsheet, or in other words, the categories of measured data. Observations, meanwhile, are the rows of a table, where each observation marks a specific instance of data.

12.3.3 Q1

Why are `pivot_longer()` and `pivot_wider()` not perfectly symmetrical? Carefully consider the following example:

```
stocks <- tibble(
  year   = c(2015, 2015, 2016, 2016),
  half   = c( 1,    2,    1,    2),
  return = c(1.88, 0.59, 0.92, 0.17)
)
stocks %>%
  pivot_wider(names_from = year, values_from = return) %>%
  pivot_longer(`2015`:`2016`, names_to = "year", values_to = "return")
```

```
## # A tibble: 4 x 3
##   half year return
##   <dbl> <chr> <dbl>
## 1     1 2015   1.88
## 2     1 2016   0.92
## 3     2 2015   0.59
## 4     2 2016   0.17
```

`pivot_longer()` has a `names_ptypes` argument, e.g. `names_ptypes = list(year = double())`. What does it do?

These two are not perfectly symmetrical as the data strongly leans towards one side and throws off the whole table. The `names` variable for `pivot_longer()` is to name the newly created variables.

12.3.3 Q2

Why does this code fail?

This code fails because you are trying to edit and pivot sections of data that do not exist, so obviously, it cannot to what is requested.

A working Version of the code is included below:

```
table4a %>%  
  pivot_longer(c(2, 3), names_to = "year", values_to = "cases")
```

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

12.3.3 Q4

Tidy the simple tibble below. Do you need to make it wider or longer? What are the variables?

```
preg <- tribble(  
  ~pregnant, ~male, ~female,  
  "yes",      NA,    10,  
  "no",       20,    12  
)
```

The variables here are pregnant, male, and female. You could change this data and make it longer by combining the male and female data sets to a categorical variable named “sex” or “gender”