

# TIDYR Pivot Wider and Pivot Longer Examples

Fan Wang

2020-04-01

## Contents

Long to Wide . . . . .	1
------------------------	---

### Long to Wide

Go back to [fan's REconTools Package](#), [R4Econ Repository](#), or [Intro Stats with R Repository](#).

Using the [pivot\\_wider](#) function in tidyR to reshape panel or other data structures

**Panel Long Attendance Roster to Wide** There are  $N$  students in class, but only a subset of them attend class each day. If student  $id_i$  is in class on day  $Q$ , the teacher records on a sheet the date and the student ID. So if the student has been in class 10 times, the teacher has ten rows of recorded data for the student with two columns: column one is the student ID, and column two is the date on which the student was in class. Suppose there were 50 students, who on average attended exactly 10 classes each during the semester, this means we have  $10 \cdot 50$  rows of data, with differing numbers of rows for each student. This is shown as `df_panel_attend_date` generated below.

Now we want to generate a new dataframe, where each row is a date, and each column is a student. The values in the new dataframe shows, at the  $Q^{th}$  day, how many classes student  $i$  has attended so far. The following results is also in a REconTools Function. This is shown as `df_attend_cumu_by_day` generated below.

**First**, generate the raw data structure, `df_panel_attend_date`:

```
# Define
it_N <- 3
it_M <- 5
svr_id <- 'student_id'

# from : support/rand/fs_rand_draws.Rmd
set.seed(222)
df_panel_attend_date <- as_tibble(matrix(it_M, nrow=it_N, ncol=1)) %>%
  rowid_to_column(var = svr_id) %>%
  uncount(V1) %>%
  group_by(!sym(svr_id)) %>% mutate(date = row_number()) %>%
  ungroup() %>% mutate(in_class = case_when(rnorm(n(),mean=0,sd=1) < 0 ~ 1, TRUE ~ 0)) %>%
  filter(in_class == 1) %>% select(!sym(svr_id), date) %>%
  rename(date_in_class = date)

# Print
kable(df_panel_attend_date) %>%
  kable_styling_fc_wide()
```

**Second**, generate wider data structure, `df_attend_cumu_by_day`:

student_id	date_in_class
1	2
1	4
2	1
2	2
2	5
3	2
3	3
3	5

```

# Define
svr_id <- 'student_id'
svr_date <- 'date_in_class'
st_idcol_prefix <- 'sid_'

# Generate cumulative enrollment counts by date
df_panel_attend_date_addone <- df_panel_attend_date %>% mutate(attended = 1)
kable(df_panel_attend_date_addone) %>%
  kable_styling_fc_wide()

# Pivot Wide
df_panel_attend_date_wider <- df_panel_attend_date_addone %>%
  pivot_wider(names_from = svr_id,
              values_from = attended)
kable(df_panel_attend_date_wider) %>%
  kable_styling_fc_wide()

# Sort and rename
# rename see: https://fanwangecon.github.io/R4Econ/amto/tibble/fs\_tib\_basics.html
ar_unique_ids <- sort(unique(df_panel_attend_date %>% pull(!sym(svr_id))))
df_panel_attend_date_wider_sort <- df_panel_attend_date_wider %>%

```

student_id	date_in_class	attended
1	2	1
1	4	1
2	1	1
2	2	1
2	5	1
3	2	1
3	3	1
3	5	1

date_in_class	1	2	3
2	1	1	1
4	1	NA	NA
1	NA	1	NA
5	NA	1	1
3	NA	NA	1

```

arrange(!sym(svr_date)) %>%
  rename_at(vars(num_range(' ', ar_unique_ids))
            , list(~paste0(st_idcol_prefix, . , ' '))
            )
kable(df_panel_attend_date_wider_sort) %>%
  kable_styling_fc_wide()

```

*# replace NA and cumusum again*  
*# see: R4Econ/support/function/fs\_func\_multivar for renaming and replacing*

date_in_class	sid_1	sid_2	sid_3
1	NA	1	NA
2	1	1	1
3	NA	NA	1
4	1	NA	NA
5	NA	1	1

```
df_attend_cumu_by_day <- df_panel_attend_date_wider_sort %>%
  mutate_at(vars(contains(st_idcol_prefix)), list(~replace_na(., 0))) %>%
  mutate_at(vars(contains(st_idcol_prefix)), list(~cumsum(.)))

kable(df_attend_cumu_by_day) %>%
  kable_styling_fc_wide()
```

date_in_class	sid_1	sid_2	sid_3
1	0	1	0
2	1	2	1
3	1	2	2
4	2	2	2
5	2	3	3

The structure above is also a function in Fan's [REconTools](#) Package, here the function is tested:

```
# Parameters
df <- df_panel_attend_date
svr_id_i <- 'student_id'
svr_id_t <- 'date_in_class'
st_idcol_prefix <- 'sid_'

# Invoke Function
ls_df_rosterwide <- ff_panel_expand_longrosterwide(df, svr_id_t, svr_id_i, st_idcol_prefix)
df_rosterwide_func <- ls_df_rosterwide$df_rosterwide
df_rosterwide_cumu_func <- ls_df_rosterwide$df_rosterwide_cumu
```

```

# Print
print(df_roster_wide_func)

## # A tibble: 5 x 4
##   date_in_class sid_1 sid_2 sid_3
##         <int> <dbl> <dbl> <dbl>
## 1           1    NA     1    NA
## 2           2     1     1     1
## 3           3    NA    NA     1
## 4           4     1    NA    NA
## 5           5    NA     1     1

print(df_roster_wide_cumu_func)

```

```

## # A tibble: 5 x 4
##   date_in_class sid_1 sid_2 sid_3
##         <int> <dbl> <dbl> <dbl>
## 1           1     0     1     0
## 2           2     1     2     1
## 3           3     1     2     2
## 4           4     2     2     2
## 5           5     2     3     3

```