R Example DPLYR Counting

Fan Wang

Go back to fan's REconTools Package, R4Econ Repository, or Intro Stats with R Repository.

```
rm(list = ls(all.names = TRUE))
options(knitr.duplicate.label = 'allow')
```

```
library(tidyverse)
library(knitr)
library(kableExtra)
library(REconTools)
# file name
st_file_name = 'fs_count_basics'
# Generate R File
try(purl(paste0(st_file_name, ".Rmd"), output=paste0(st_file_name, ".R"), documentation = 2))
# Generate PDF and HTML
# rmarkdown::render("C:/Users/fan/R4Econ/summarize/count/fs_count_basics.Rmd", "pdf_document")
# rmarkdown::render("C:/Users/fan/R4Econ/summarize/count/fs_count_basics.Rmd", "html_document")
```

Uncount

In some panel, there are N individuals, each observed for Y_i years. Given a dataset with two variables, the individual index, and the Y_i variable, expand the dataframe so that there is a row for each individual index's each unique year in the survey.

Search:

- r duplicate row by variable

Links:

- see: Create duplicate rows based on a variable

Algorithm:

1. generate testing frame, the individual attribute dataset with invariant information over panel
2. uncount, duplicate rows by years in survey
3. group and generate sorted index
4. add indiviual specific stat year to index

```
# 1. Array of Years in the Survey
ar_years_in_survey <- c(2,3,1,10,2,5)
ar_start_yaer <- c(1,2,3,1,1,1)
ar_end_year <- c(2,4,3,10,2,5)
mt_combine <- cbind(ar_years_in_survey, ar_start_yaer, ar_end_year)

# This is the individual attribute dataset, attributes that are invariant acrosss years
tb_indi_attributes <- as_tibble(mt_combine) %>% rowid_to_column(var = "ID")

# 2. Sort and generate variable equal to sorted index
```

```r
tb_indi_panel <- tb_indi_attributes %>% uncount(ar_years_in_survey)

# 3. Panel now construct exactly which year in survey, note that all needed is sort index
# Note sorting not needed, all rows identical now
tb_indi_panel <- tb_indi_panel %>%
                    group_by(ID) %>%
                    mutate(yr_in_survey = row_number())

tb_indi_panel <- tb_indi_panel %>%
                    mutate(calendar_year = yr_in_survey + ar_start_yaer - 1)

# Show results Head 10
tb_indi_panel %>% head(10) %>%
  kable() %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"))
```

ID	ar_start_yaer	ar_end_year	yr_in_survey	calendar_year
1	1	2	1	1
1	1	2	2	2
1	1	2	3	3
1	1	2	4	4
2	1	4	1	1
2	1	4	2	2
2	1	4	3	3
2	1	4	4	4
4	1	2	1	1
1	2	3	2	3

2

2

4

3

4

3

3

3

1

3

4

1

10

1

1

4

1

10

2

2

4

1

10

3

3

4

1

10

4

4