R Ladies 5 Minute Fun

Luda

5/10/2018

The idea behind this mini-talk is to build on other R-Ladies’s talks by continuing the idea of Emily Robinson’s “Lesser Kown ⭐s of the Tidyverse” and using the janitor package (thank you, Erin!) Problem: survey data with a bunch of factor variables that are not reading in correctly 🙄 Pattern: several of the factor variables have the same levels 🧐 Solution: change all those variables at once! 🤩 Specific 📦s featured:

* forcats()
* janitor() - specifically tabyl()
* dplyr() - specifically mutate\_at()

### Load these libraries

suppressMessages(library(tidyverse))  
suppressMessages(library(here))  
suppressMessages(library(janitor))  
# suppressMessages(devtools::install\_github("hadley/emo"))  
suppressMessages(library(emo))

### Load the data

pre <- suppressMessages(read\_csv("/Users/ljanda/Documents/misc/r\_ladies\_pre\_data.csv"))

### Inspect the problem

(Note: I already set the variables with the same factor levels to have the same prepend)

pre %>%   
 select\_if(is.character) %>%   
 select(contains("rye\_"),   
 contains("hk\_"),   
 contains("years")) %>%   
 map(~tabyl(.))

## $rye\_readscitext  
## . n percent valid\_percent  
## 1 Emerging 34 0.14345992 0.16585366  
## 2 Experienced 139 0.58649789 0.67804878  
## 3 Expert 29 0.12236287 0.14146341  
## 4 Very new 3 0.01265823 0.01463415  
## 5 <NA> 32 0.13502110 NA  
##   
## $rye\_learnscivocab  
## . n percent valid\_percent  
## 1 Emerging 40 0.168776371 0.195121951  
## 2 Experienced 137 0.578059072 0.668292683  
## 3 Expert 26 0.109704641 0.126829268  
## 4 Very new 2 0.008438819 0.009756098  
## 5 <NA> 32 0.135021097 NA  
##   
## $rye\_scitext  
## . n percent valid\_percent  
## 1 Emerging 67 0.28270042 0.32682927  
## 2 Experienced 111 0.46835443 0.54146341  
## 3 Expert 17 0.07172996 0.08292683  
## 4 Very new 10 0.04219409 0.04878049  
## 5 <NA> 32 0.13502110 NA  
##   
## $hk\_ellparticipateinsci  
## . n percent valid\_percent  
## 1 Knowledgeable 136 0.57383966 0.66341463  
## 2 Not at all knowledgeable 8 0.03375527 0.03902439  
## 3 Not knowledgeable 20 0.08438819 0.09756098  
## 4 Very knowledgeable 41 0.17299578 0.20000000  
## 5 <NA> 32 0.13502110 NA  
##   
## $hk\_swdparticipateinsci  
## . n percent valid\_percent  
## 1 Knowledgeable 138 0.58227848 0.67317073  
## 2 Not at all knowledgeable 5 0.02109705 0.02439024  
## 3 Not knowledgeable 27 0.11392405 0.13170732  
## 4 Very knowledgeable 35 0.14767932 0.17073171  
## 5 <NA> 32 0.13502110 NA  
##   
## $hk\_teaching\_science  
## . n percent valid\_percent  
## 1 Knowledgeable 136 0.573839662 0.666666667  
## 2 Not at all knowledgeable 2 0.008438819 0.009803922  
## 3 Not very knowledgeable 32 0.135021097 0.156862745  
## 4 Very knowledgeable 34 0.143459916 0.166666667  
## 5 <NA> 33 0.139240506 NA  
##   
## $hk\_teaching\_sci\_field\_trial\_units  
## . n percent valid\_percent  
## 1 Knowledgeable 134 0.565400844 0.656862745  
## 2 Not at all knowledgeable 1 0.004219409 0.004901961  
## 3 Not knowledgeable 46 0.194092827 0.225490196  
## 4 Very knowledgeable 23 0.097046414 0.112745098  
## 5 <NA> 33 0.139240506 NA  
##   
## $years\_teaching  
## . n percent valid\_percent  
## 1 11 to 15 33 0.13924051 0.15137615  
## 2 16 to 20 35 0.14767932 0.16055046  
## 3 2 10 0.04219409 0.04587156  
## 4 3 16 0.06751055 0.07339450  
## 5 4 8 0.03375527 0.03669725  
## 6 5 16 0.06751055 0.07339450  
## 7 6 to 10 47 0.19831224 0.21559633  
## 8 more than 20 53 0.22362869 0.24311927  
## 9 <NA> 19 0.08016878 NA  
##   
## $years\_teaching\_current\_grade  
## . n percent valid\_percent  
## 1 1 23 0.09704641 0.10599078  
## 2 11 to 15 24 0.10126582 0.11059908  
## 3 16 to 19 10 0.04219409 0.04608295  
## 4 2 30 0.12658228 0.13824885  
## 5 20 or more 13 0.05485232 0.05990783  
## 6 3 33 0.13924051 0.15207373  
## 7 4 16 0.06751055 0.07373272  
## 8 5 10 0.04219409 0.04608295  
## 9 6 to 10 58 0.24472574 0.26728111  
## 10 <NA> 20 0.08438819 NA

### Set correct levels

rye\_levels <- c("Very new", "Emerging", "Experienced", "Expert")  
hk\_levels <- c("Not at all knowledgeable", "Not knowledgeable", "Knowledgeable", "Very knowledgeable")  
years\_levels <- c("1", "2", "3", "4", "5", "6 to 10", "11 to 15", "16 to 20", "more than 20")

### Apply levels

pre\_f <-   
 pre %>%   
 mutate\_at(vars(contains("rye\_")), ~ factor(., levels = rye\_levels)) %>%   
 mutate\_at(vars(contains("hk\_")), ~ factor(., levels = hk\_levels)) %>%   
 mutate\_at(vars(contains("years\_")), ~ factor(., levels = years\_levels))

### Check levels

pre\_f %>%   
 select\_if(is.factor) %>%   
 select(contains("rye\_"), contains("hk\_"), contains("years")) %>%   
 map(~tabyl(.))

## $rye\_readscitext  
## . n percent valid\_percent  
## 1 Very new 3 0.01265823 0.01463415  
## 2 Emerging 34 0.14345992 0.16585366  
## 3 Experienced 139 0.58649789 0.67804878  
## 4 Expert 29 0.12236287 0.14146341  
## 5 <NA> 32 0.13502110 NA  
##   
## $rye\_learnscivocab  
## . n percent valid\_percent  
## 1 Very new 2 0.008438819 0.009756098  
## 2 Emerging 40 0.168776371 0.195121951  
## 3 Experienced 137 0.578059072 0.668292683  
## 4 Expert 26 0.109704641 0.126829268  
## 5 <NA> 32 0.135021097 NA  
##   
## $rye\_scitext  
## . n percent valid\_percent  
## 1 Very new 10 0.04219409 0.04878049  
## 2 Emerging 67 0.28270042 0.32682927  
## 3 Experienced 111 0.46835443 0.54146341  
## 4 Expert 17 0.07172996 0.08292683  
## 5 <NA> 32 0.13502110 NA  
##   
## $hk\_ellparticipateinsci  
## . n percent valid\_percent  
## 1 Not at all knowledgeable 8 0.03375527 0.03902439  
## 2 Not knowledgeable 20 0.08438819 0.09756098  
## 3 Knowledgeable 136 0.57383966 0.66341463  
## 4 Very knowledgeable 41 0.17299578 0.20000000  
## 5 <NA> 32 0.13502110 NA  
##   
## $hk\_swdparticipateinsci  
## . n percent valid\_percent  
## 1 Not at all knowledgeable 5 0.02109705 0.02439024  
## 2 Not knowledgeable 27 0.11392405 0.13170732  
## 3 Knowledgeable 138 0.58227848 0.67317073  
## 4 Very knowledgeable 35 0.14767932 0.17073171  
## 5 <NA> 32 0.13502110 NA  
##   
## $hk\_teaching\_science  
## . n percent valid\_percent  
## 1 Not at all knowledgeable 2 0.008438819 0.01162791  
## 2 Not knowledgeable 0 0.000000000 0.00000000  
## 3 Knowledgeable 136 0.573839662 0.79069767  
## 4 Very knowledgeable 34 0.143459916 0.19767442  
## 5 <NA> 65 0.274261603 NA  
##   
## $hk\_teaching\_sci\_field\_trial\_units  
## . n percent valid\_percent  
## 1 Not at all knowledgeable 1 0.004219409 0.004901961  
## 2 Not knowledgeable 46 0.194092827 0.225490196  
## 3 Knowledgeable 134 0.565400844 0.656862745  
## 4 Very knowledgeable 23 0.097046414 0.112745098  
## 5 <NA> 33 0.139240506 NA  
##   
## $years\_teaching  
## . n percent valid\_percent  
## 1 1 0 0.00000000 0.00000000  
## 2 2 10 0.04219409 0.04587156  
## 3 3 16 0.06751055 0.07339450  
## 4 4 8 0.03375527 0.03669725  
## 5 5 16 0.06751055 0.07339450  
## 6 6 to 10 47 0.19831224 0.21559633  
## 7 11 to 15 33 0.13924051 0.15137615  
## 8 16 to 20 35 0.14767932 0.16055046  
## 9 more than 20 53 0.22362869 0.24311927  
## 10 <NA> 19 0.08016878 NA  
##   
## $years\_teaching\_current\_grade  
## . n percent valid\_percent  
## 1 1 23 0.09704641 0.11855670  
## 2 2 30 0.12658228 0.15463918  
## 3 3 33 0.13924051 0.17010309  
## 4 4 16 0.06751055 0.08247423  
## 5 5 10 0.04219409 0.05154639  
## 6 6 to 10 58 0.24472574 0.29896907  
## 7 11 to 15 24 0.10126582 0.12371134  
## 8 16 to 20 0 0.00000000 0.00000000  
## 9 more than 20 0 0.00000000 0.00000000  
## 10 <NA> 43 0.18143460 NA

🎉 🎉 🎉