library(tidyverse)

## -- Attaching packages ------------------------------------------- tidyverse 1.3.0 --

## ggplot2 3.2.1 purrr 0.3.3  
## tibble 2.1.3 dplyr 0.8.3  
## tidyr 1.0.0 stringr 1.4.0  
## readr 1.3.1 forcats 0.4.0

## -- Conflicts ---------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

longreads2019 <- read\_csv("~/Downloads/Blogging A to Z/reads2019\_long.csv")

## Parsed with column specification:  
## cols(  
## Book.ID = col\_double(),  
## Title = col\_character(),  
## genre = col\_character()  
## )

I can use the dummy\_cols functions to create the genres flags, that I can aggregate down and merge into the reads2019 file .For this function, you’ll need the fastDummies package (so add install.packages(“fastDummies”) before the rest of the code). Also, since the number of dummy code variables typically are equal to the number of categories minus 1, the function automatically removes the first dummy variable from the final file. Since I’m using these as flags rather than dummy variables, I want to overide that default, which I do with remove\_first\_dummy = FALSE.

library(fastDummies)  
  
genres <- longreads2019 %>%  
 dummy\_cols(select\_columns = "genre", remove\_first\_dummy = FALSE)  
  
genres <- genres %>%  
 group\_by(Book.ID) %>%  
 summarise(Fiction = max(genre\_Fiction),  
 Childrens = max(genre\_Childrens),  
 Fantasy = max(genre\_Fantasy),  
 SciFi = max(genre\_SciFi),  
 Mystery = max(genre\_Mystery),  
 SelfHelp = max(genre\_SelfHelp))  
  
reads2019 <- read\_csv("~/Downloads/Blogging A to Z/ReadsNoGenre.csv",  
 col\_names = TRUE)

## Parsed with column specification:  
## cols(  
## Title = col\_character(),  
## Pages = col\_double(),  
## date\_started = col\_character(),  
## date\_read = col\_character(),  
## Book.ID = col\_double(),  
## Author = col\_character(),  
## AdditionalAuthors = col\_character(),  
## AverageRating = col\_double(),  
## OriginalPublicationYear = col\_double(),  
## read\_time = col\_double(),  
## MyRating = col\_double(),  
## Gender = col\_double(),  
## NewRating = col\_double(),  
## FinalRating = col\_double()  
## )

reads2019 <- reads2019 %>%  
 left\_join(genres, by = "Book.ID")

I know I’ve sprinkled in other tidyverse functions in these posts, such as group\_by and summarise. Don’t worry! I’ll post more about those functions in this series – stay tuned!