

Day 5 - Data manipulation with the tidyverse

Introduction

The purpose of today's activity is to build some confidence using the tools in the `tidyverse` R package to manipulate data. To guide our exploration today, we will use a dataset¹ containing all movies and TV shows available on Netflix (as of 2022).

At the end of today's notes, you should be familiar with the following functions:

- `filter`
 - `select`
 - `mutate/case_when`
 - `group_by`
 - `summarize`
-

¹[Link](#) to the data set on Kaggle.

Challenge prompt

With so many streaming platforms available today, it is hard to choose the best one! The same data set exists for Amazon prime and Hulu. Load in both data sets, and determine the degree of overlap in titles available in the US between each of the services. (There are a lot of ways to do this! Have fun with it.)

```
# load other data
netflix <- read_csv("netflix_titles.csv")
hulu <- read_csv("hulu_titles.csv")
prime <- read_csv("amazon_prime_titles.csv")
```