

## A1: Data Query, Subsetting, and Sampling

In this exercise, you will work with a data set of movies created by my brother. The data are stored in a CSV file (**matts\_movies.csv**) and consists of 1,852 movies with the following attributes:

“Movie.Name” = Name of movie

“Director” = Director of movie

“Release.Year” = Year the movie was released

“My.Rating” = My brother’s rating of the movie

“Genre” = The movie genre

“Own” = Whether my brother owns a copy of this movie (please don’t judge him here).

Create code to obtain the result and/or answer the following questions. Submit your code and the answers in an R file (.R) or text file (.txt). Include all answers below the code as comments.

- T1.** How many genres are differentiated in the dataset? (2 Points)
- T2.** Which genre has the most entries? (2 Points)
- T3.** What percentage of the movies did Matt rate above 8.5? (2 Points)
- T4.** What percentage of the movies are dramas? (2 Points)
- T5.** What percentage of the movies are dramas with a rating higher than 9.5? (2 Points)
- T6.** How many movies were directed by Wes Anderson? (2 Points)
- T7.** What percentage of the movies were made after 2010, are comedies, and have a rating above 9? (4 Points)
- T8.** How many romance or independent films were released during or after 2005 and have a rating above 9? (2 Points)
- T9.** What percentage of the independent films have a rating greater than 8.5? (2 Points)
- T10.** What percentage of the movies on the list does Matt own? (2 Points)
- T11.** Generate code to subset out only the “Movie.Name” and “Director” columns. (2 Points)
- T12.** Generate code to obtain a data frame of the mean rating by genre. (2 Points)
- T13.** Which genre has the highest mean rating? (2 Point)
- T14.** Which director, Wes Anderson or David Fincher, has the highest average rating for his films included in the data set? (4 Points)
- T15.** Generate code to randomly sample 25 movies from the data set. (2 Points)
- T16.** Generate code to sample 1 movie per Genre. (2 Points)
- T17.** Which director has the largest range of ratings? (4 Point)