Step 1\_TidyTuesday Project\_Tutorial 9109 Group 1

Khadeeja Naseer

Michael Skotar

Ruoqi Sun

Justin Teng

Zehan Liu

Guanlin Wu

The data we selected is from the Friends R package on the TidyTuesday website. Friends is an American television sitcom which aired on NBC from September 22, 1994, to May 6, 2004, lasting 10 seasons. It revolves around a group of 6 friends who live in New York City, and follows their adventures with love, romance, and family. This dataset contains three data files: friends.csv, which is about the show’s transcripts, friends\_emotions.csv and friends\_info.csv, which introduce some additional information about the show. The 6 variables in this file that we focus on include: season (season number), episode (episode number), written\_by (Writer of the episode), air\_date (date of release), us\_views\_million (number of viewers in USA upon release), and imdb\_rating (Rating of episode on IMDB). There are 4 key relationships among the variables that we would like to investigate. Comparing season vs us\_views\_millions will allow us to determine how viewership changed through the course of the show. The next relationship is episode and imdb\_rating, which will display the quality of the show over time. We can compare us\_views\_millions and imdb\_rating to explore the relation among them and examine if a larger audience implies a higher imdb rating. Analyzing air\_date and us\_views\_millions would reveal if particular dates saw a higher audience. Finally, exploring the relationship between imdb\_rating and written\_by will display what writer(s) content appealed to the viewers.