This project asks you to download and analyze user rating information from IMDb for three TV shows of your choice\*.   
(\* Newer episodes for current TV shows may not be available. Additionally, less well-known shows or shows that aired prior to 1980 may not be available.)

**Step 1: Collecting Your Data**

* You will use a Web App that allows you to “scrape” data from a TV show’s IMDb site. Go to <https://stlawu.shinyapps.io/imdb_tv/> to load the App.
* Choose three TV shows to analyze. Type in the names of each TV show to scrape the episode data. (If your show doesn’t appear, double check that you spelled it correctly. If it still doesn’t appear, pick a different show.)
* After the data have been “scraped”, you should see a preview of the data on the screen. If you like what you see, click on the “Download” button to save the data as a Comma Separated Values file (basically a generic version of an Excel spreadsheet). Save it somewhere you’ll be able to access it from a campus computer (e.g., your **P**ersonal drive)
* To load your dataset into Minitab
  + Open the folder in which your data are stored
    - Right click on file name, and choose “Open With” and then choose “Minitab”
* Since you need to download your data and then analyze it, I do not recommend working on the remote desktop (a campus lab PC would be easiest)

**Step 2: Required Analyses**

Note this is just a list of what you need to analyze. See Step 3 for how to present it.

*Graphical and Descriptive Analyses*

1. Univariate Analyses
   1. An analysis of the average Episode Ratings (ignoring seasons)
      1. For each TV show, report descriptive statistics of average Episode Rating (including sample sizes, means, and five number summaries)
      2. For each TV show, provide an appropriate graphical display of average Episode Ratings
      3. For each TV Show, discuss whether or not a Normal distribution is appropriate for modeling the average Episode Ratings
      4. For each TV show, discuss how the average show rating (from the IMDb website for the TV show) compares to average Episode Ratings
      5. Include a paragraph that summarizes similarities and differences in the trends for the three shows
2. Bivariate Analyses
   1. An analysis of the relationship between average Episode Ratings and Overall Episode Number
      1. For each TV show, provide an appropriate graphical display (with a regression line) that allows you to predict average Episode Rating using Overall Episode Number
      2. For each TV show, report the regression equation and for the plot described above
      3. For **one** TV show, provide an interpretation of the slope of the regression equation
      4. For each TV show, identify the episode that appears to have the largest residual (either positive or negative), and provide some commentary on that episode (from a subject matter expert perspective)
      5. Include a paragraph that summarizes similarities and differences in the trends for the three shows

*Inferential Analysis*

1. Use an appropriate statistical method to determine if there are any statistically significant differences in the mean of the average Episode Ratings for the three shows. Provide the appropriate Minitab output with a paragraph addressing conditions (aside from randomness) that may have failed and any conclusions that can be made from the output.

**Step 3: Written Report**

**The written report is the only thing you will turn in it should include the following sections**

1. Introduction
   * The Introduction should provide some basic information about your TV shows. For each TV show, please include such information as when it started, how many seasons it has been running (and how many episodes there have been), what network it is on, and the *overall series rating from the TV Show’s IMDb webpage*. Please include a few sentences about why you chose the TV shows that you did.
2. Analysis
   * For each of the analyses listed above, you must provide the graphical displays of the data and those displays must be fully described and referenced in the paper.
     1. Your description should include context and be supported with appropriate numerical values and units. For each analysis, be sure to discuss the same features we did in class. (e.g., Center, Spread, Shape, and Outliers for single numerical variables)
     2. Each graphical display must appear in the text (near where it is discussed) and have a figure number and informative (but brief) caption (see, for example, the captions on Figures 4.1 and 4.2 of your textbook, pp. 134-135).
3. Discussion
   * A paragraph or two highlighting the interesting findings from your show. This may include a discussion on outliers (and a brief description of what happened in the episode to “make it unusual”) or other trends that you observe that have some “real world connection.” This will likely require you understanding the show you choose and being willing to dig around online (if need be) to learn some specifics about your show.
4. References – the URL for the IMDb pages for your TV Shows (in case I need to double-check anything)

**Other Requirements and Tips**

1. You must write in complete sentences and paragraphs (one paragraph per analysis subpart). **No bulleted lists!**
2. The project must be typed and graphics must be resized and placed within the report. *Do not simply place them at the end of the project.*
3. **This is an individual project. If you choose to discuss the project with a classmate/friend, please be sure that you use different TV shows.**
4. **This project is worth 60 project points:** 50 points for analyses (including descriptions) and 10 points for quality of writing and proper formatting.