This project asks you to download user rating information from IMDb for a TV show of your choice (with a few restrictions) and conduct a descriptive analysis (numerically and graphically) of the data.

**Step 1: Collecting Your Data**

* Recently a student and Dr. Ramler developed a Web App that allows you to “scrape” data from a TV show’s IMDb site. Go to <http://shiny.stlawu.edu:3838/sample-apps/tvseries/> to load the App. (FYI, it takes a few seconds to load…and if you get a connection error just try to reload it.)
* Type in the name of your favorite TV show (with the restrictions listed below) to scrape the episode data. (If your show doesn’t appear, double check that you spelled it correctly. If it still doesn’t appear, just pick a different show.)
* After the data has been “scraped”, you should see some side-by-side boxplots (*hint: one of the plots you make for this project should look like this one*) and a preview of the data on the screen. If you like what you see, click on the “Download Data as CSV” 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). Note that you will need to log in to VPN in order to save the file to your P drive.
* **Restrictions on your TV show**
  + Make sure it is a TV show and not a movie.
  + It needs to have at least three seasons. (Please note that the data was last collected in Summer 2018 so some new seasons may not be available for all shows.)
  + It can’t be *The Simpsons* (as I am using that as an example in class)
  + It can’t be *Stranger Things* or *The Marvelous Mrs. Maisel* (as I don’t want to deal with spoilers related to the most recent seasons!)
  + Failure to abide by these restrictions will result in an automatic zero for this project.

**Step 2: Required Analyses**

For each subpart, you should create the appropriate graphical summary and calculate any numerical summaries that are appropriate. (Note this is just a list of what you need to analyze. See Step 3 for how to present it.)

1. Univariate (Single Variable) Analyses
   1. An analysis of the Episode Ratings (ignoring seasons)
   2. An analysis of User Votes (ignoring seasons)
2. Bivariate (Two Variable) Analyses
   1. An analysis of the Episode Ratings across the seasons (treat Season as a categorical variable)
   2. An analysis of the relationship between Episode Ratings and User Votes

**To Analyze your Data:**

* Logon to the **Virtual Desktop**
* To load your dataset into Minitab
  + Open Minitab (Click on the Spy Glass icon and type Minitab)
    - In top left corner, click File > Open
    - Navigate to the directory where you stored the data. It should appear as a file you can open. Click on the file name and then click ok.
  + Create the four graphs described above

**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 show. Be sure to 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*.
2. Analysis
   * For each of the analyses listed above, you must provide the graphical display of the data and that display 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. Any scatterplots you make should include a **smoother** to help identify any trends.
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 page for you TV Show (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 30 project points:** 25 points for analyses (including descriptions) and 5 points for quality of writing and proper formatting.

**Project 2 Variable Descriptions** The variables you get from the IMDb Web App are:

* Season: The season of the episode
* episodeNumber: The episode number within the season (not needed for this project, but still somewhat interesting to look at in a scatterplot vs User Rating)
* averageRating: The average user rating for the episode according to IMDb methodology.
* numVotes: The number of IMDb users providing votes for the episode
* Title: The name of the episode (not used to create any graphical displays, but still interesting to look at when discussing such things as outliers)