- (5%) Coding Budget: You are allowed at most one eleven steps to complete this assignment. They can be any combination of `DATA` and `PROC` steps. If you feel you need a higher budget, you can request additional ones in person by demonstrating a need for a higher budget.

Specifications:

- (0%) There is no requirement to work in a team for this project. However, if you do choose to form a team it can have no more than three people in it. Teams of four or more people will receive a zero on this project.
- (5%) Setting up paths: Programmatically change your working directory as needed to set up any `FILENAME` and `LIBNAME` statements. Do not use absolute paths anywhere in your code other than in these X commands at the top of your program!
- (10%) This project makes use of our most recent skills - web-scraping and quantile regression. As a change of pace, this project does not provide a fixed data source. Your team (even if it is a team of one!) must identify a website that contains data you would like to scrape. The goal is for you to find a topic you're interested in learning more about, find some data regarding that topic, read it in, and then fit some regression models. The data and website must conform to the following rules:
  - The data is stored at a website that does not require any login information.
  - The website either does not have a robots.txt file or its robots.txt file gives you permission to scrape data from the site.
  - The data you want is not stored on a single page (e.g. it uses pagination to store data on different pages).
  - The data you scrape is contains at least three numeric variables suitable for use in a regression model.
- (50%) Read in the data you plan to use in your regression model. Remember, it is supposed to have data stored across multiple pages. Write a macro that automatically reads in the data from each of the pages. You can assume that if the website adds pages that they follow the same structure they use now. You have the flexibility to design a macro that meets your needs; however, here are some items to consider when you write it:
  - Almost everyone who uses macros hates positional parameters.
  - The user should never be asked to specify the names of all the pages, or any of the pages, actually. Remember, web scraping depends heavily on hardcoding, so this macro would likely not be used to scrape another site unless the two sites used the same design for their websites. Once the website is specified (which you can do in a %LET statement if you'd like) everything else can be as automatic as possible within the site.
  - You've also taken ST 445 - you might find some of the tasks are better handled with `DATA` steps while others are better handled with `SQL`. That's perfectly fine.
  - Think about efficiency. The budget is looser than it has ever been before, but you'll still want your code to be pretty efficient.
- (10%) Once your website is correctly scraped, do the following. You're going to want to look up the names of some of these options in the SAS documentation.
  - State a specific question (or set of questions) you're interested in answering with your regression model by placing them in your code as comments.
  - Fit a quantile regression model to at least 5 quantiles to support your investigation.
  - Save the predicted values,
  - Have `QUANTREG` make every diagnostic plot that it can make. (That's five plots.) Make sure leverage and outlier points are labeled using some meaningful variable from your data set.
  - Save the following diagnostic measures all in the same data set: leverage indicator, Mahalanobis distance, outlier indicator, predicted values, robust distance, standardized residuals, raw residuals.
- (15%) Reproduce at least one of the diagnostic plots using the saved diagnostic measures. (Note that I'm not considering the plot of the original data with the fitted models on it to be a diagnostic plot.)
- (5%) What do the regression models tell you about the question(s) you asked earlier? Include your conclusions as a comment.

BONUS Reproduce all the plots - diagnostic and fitted - using the saved data.