Assignment #3: Pythagorean Formula

Your task is to evaluate the effectiveness of the Pythagorean formula in the sports league of your choice.

Why are you being asked to do this?

Fundamentally, the question we are trying to answer is: *How much do we want to pay attention to win-loss record v. score differential?* The answer to this question varies by sport and by league, and it will drive the modeling decisions we make on future assignments.

What (exactly) are you being asked to do?

Find a dataset of game scores from a completed league or competition of your choice. Use these data to estimate the optimal exponent α within the Pythagorean formula for predicting winning percentage in this league. Report the value of α you obtained. Split your dataset into two equal-sized subsets, and then answer the following questions:

- How well does first-half **actual** winning percentage predict second-half **actual** winning percentage? Report the correlation and create a plot to illustrate the relationship between these two variables.
- How well does first-half **Pythag** winning percentage predict second-half **actual** winning percentage? Report the correlation and create a plot to illustrate the relationship between these two variables.
- How well does first-half **residual** winning percentage predict second-half **residual** winning percentage? Report the correlation and create a plot to illustrate the relationship between these two variables. *Residual winning percentage is actual winning percentage minus Pythag winning percentage. We are investigating how much signal there is in a team's over-/under-performance of its Pythag record.*

Finally, **escape from model land** by interpreting your results in the context of the real world. You need NOT address all of these, but here are examples of questions you may consider addressing: *Why might this model be an inadequate representation of reality? What decision in sport management might be affected by this analysis? How might this analysis change the way fans think about the sport?*

Submission Requirements

- A PDF report (max 2 pages) summarizing your findings, including at minimum the following:
  - a description of the league and dataset you chose
  - the value of α you estimated for the Pythagorean formula using your data
  - the correlations between the requested variables
  - three plots illustrating the relationships between the requested variables
  - an interpretation of your results in the context of the real world
- An R script that contains all of the code you used to perform the analysis

Reminders

- Prepare your report as if your audience is a front office executive who has not seen the assignment prompt. Write clearly and concisely, and format your report in a way that makes it easy to read.
- In this class we value exercising **creativity** on homework assignments! Look for opportunities to put your own personal touch on your work—try to do more than parrot what you've been taught.
- Please **anonymize** your submission by removing any personally identifiable information (including file paths in your R script that contain things like a username!).

How will your grade be determined?

You will get feedback on your work product based on several criteria. Within each of those criteria, the feedback will be: Missing (0%), Needs Improvement (70%), Good (85%) or Exceeds Expectations (100%). Your grade on the assignment will be the average of the grades across criteria. The criteria are:

1. **Data modeling.** Did you correctly estimate the optimal exponent α in the Pythagorean formula? Did you correctly report the correlations between the requested variables?

2. **Data visualization.** Did you create plots to visualize the relationships between the requested variables? Are the plots aesthetically designed and easy to read?

3. **Creative thinking.** Did you bring your own ideas from outside of this class to bear on the assignment?

4. **Critical thinking.** Did you escape from model land? Did you weigh evidence from multiple perspectives in forming your conclusion? Did you provide a thoughtful interpretation of your results?

5. **Written communication.** Did you write clearly and concisely? Did you organize your key ideas with the evidence supporting them? Did you format your report in a way that makes it easy to read?