Laboratory Exercise 1

Objective

- This case study provides an opportunity to demonstrate our ability to combine datasets and produce meaningful analysis. Specifically, we would like to provide a decision maker with more than just data—we want to provide insights, understanding, and wisdom. This exercise allows the student an opportunity to demonstrate progress (or mastery) of learning objectives 1, 2, 3, 4, and 5.

1) Obtain data and understand data structures and data elements.
2) Scrub data using scripting methods, to include debugging, for data manipulation in R and other tools.
3) Explore data using essential qualitative analysis techniques, including descriptive statistics.
4) Model relationships between data using the appropriate analytical methodologies matched to the information and the needs of clients and users.
5) Interpret the data, model, analysis, and findings, and communicate the results in a meaningful way.

Instructions

- The research question is, how can we recommend the best salary (total compensation, minus bonus) for our next head football coach?
- Start with the data **Coaches**.
- Review the data—clean as appropriate.
- Consider the base worksheet and the additional data.
    - Stadium size
    - Graduation rate:
        - Available from http://fs.ncaa.org/Docs/newmedia/public/rates/index.html
        - Use the 2006 cohort and include both GSR and FGR
    - Annual donations to program (if available)
- Develop an additional vector for each school using last year's record.
- Build a data frame for your analysis.
- Conduct an initial data analysis.
- Fit a regression model with the salary as the response and the relevant predictors (i.e., you will need more than one predictor).
- Answer the following questions in your report:
    - What is the recommended salary for the Syracuse football coach?
    - What would his salary be if we were still in the Big East? What if we went to the Big Ten?
    - What schools did we drop from our data, and why?
    - What effect does graduation rate have on the projected salary?

    - How good is our model?
    - What is the single biggest impact on salary size?
- Bonus:
    - Develop a geographic visualization that in your view best depicts the conferences' median salary.
    - Fit a hierarchical model based on conference.
- Double bonus:
    - Using some additional materials—https://apsportal.ibm.com/analytics—recreate your code for the basic regression model with a training and test set, but do not use conference as a predictor.

Additional Instructions

- Don't forget what you learned in your previous courses; do your own work, document any assistance, and use comments for clarity.

Submission Items

- Case study report with data analysis, graphics, and answers to specific questions
- Supporting notebook for the report—be sure to include comments regarding how you combined additional data into your analysis