Laboratory Exercise - 1

Objective

- This exercise provides an opportunity to demonstrate our ability to combine data sets 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) **O**btain data and understand data structures and data elements.
2) **S**crub data using scripting methods, to include debugging, for data manipulation in R and other tools.
3) **E**xplore data using essential qualitative analysis techniques including descriptive statistics.
4) **M**odel relationships between data using the appropriate analytical methodologies matched to the information and the needs of clients and users.
5) I**N**terpret the data, model, analysis, and findings. Communicate the results in a meaningful way.

Instructions

- The research question is how can we recommend the best salary (totalpay or schoolpay – your choice) for our next head football coach?
- Start with the data **Coaches –https://github.com/2SUBDA/IST_718**
- Review the data – clean as appropriate
- Consider the base worksheet and additional data such as:
    - Stadium size – available via internet search
    - Graduation rate :

      Available from https://www.icpsr.umich.edu/web/ICPSR/studies/26801/summary or
      From here https://web3.ncaa.org/aprsearch/gsrsearch
      Use the most recent cohort available and include both GSR and FGR

    - Annual donations to program – if available via internet search
    - School's win/loss record from last available year
    - Other data as you determine might be applicable
- Build a data frame for your analysis
- Conduct an initial data analysis – exploratory data analysis – develop appropriate visualizations
- 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

Additional Instructions

- Don't forget what you learned in your previous courses; do your own work, document any assistance, 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