1. **Introduction**

In the competitive landscape of college sports, particularly in the realm of Power 5 football, which encompasses the five most prominent college football conferences in the country (the Atlantic Coast Conference, the Big Ten Conference, the Big 12 Conference, the Pac-12 Conference, and the Southeastern Conference), questions arise regarding the influence of financial expenditure on team success. This prompts a critical examination of fairness and competitive advantage: How much does the power of spending/money affect team and player performance in college sports?[[1]](#footnote-1) Does it grant teams with higher spending an unfair advantage resulting in a pay-to-win system?

This project tackles these questions by analyzing college football financial data and performance statistics over a ten-year period from 2013 to 2023. These datasets contain a plethora of evaluation metrics, ranging from on-field performance to the financial resources allocated to maintain overall football operations. The analysis of this data aims to provide statistical clarity to the overarching question of whether spending has a discernible impact in shaping the performance of Power 5 college football teams.

This objective is intertwined with the overall research question: “Does higher spending on college football programs correlate with increased success in terms of performance metrics?” However, this question cannot simply be answered on its own. The attribution of success in college football stems from several arenas of cash flow, requiring an initial division of the research question into two distinct areas.

The first focus is on distinct categories of spending and how those variables correlate with on field performance, recognizing how financial investment yields various impacts on a college football team’s success. Additionally, the longitude of the research question and how the relationship between spending and team success has changed over time should be investigated. Through these two approaches, the aim is to provide a comprehensive analysis of the complex dynamic between financial investment and performance in college football.

1. **Motivation**

The original motivation for this project stemmed from a desire to explore how the newly introduced Name, Image, and Likeness (NIL) framework which allows college athletes to make money from their personal brand through endorsement, sponsorships, and social media monetization, impacts college football performance and recruiting rankings. It is believed that the NIL represented a shift in the traditional college sports landscape, resulting in a change in the overall focus of financial spending in college athletics. The integration of the NIL into the analysis could offer insights into the new relationship between financial investment and on-field performance and recruiting for Power 5 college football teams.

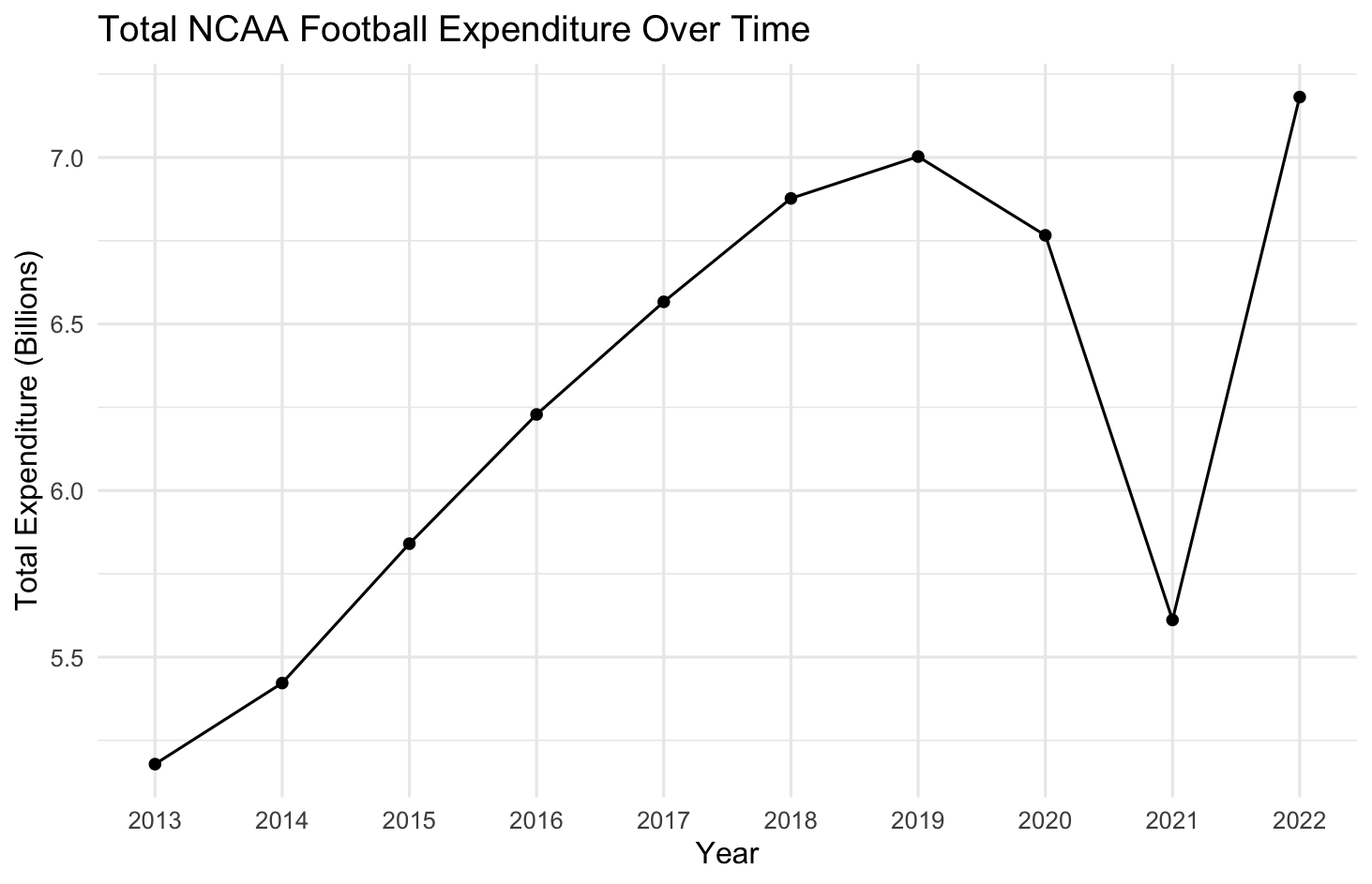
However, due to the limited access to comprehensive NIL data, the original topic was changed to examine other facets of college athletic expenditures, such as coaching salaries, recruitment efforts, and facility infrastructure investments. As a result, the discourse regarding the relationship between financial investment and college football program’s success changed from a singular focus on the NIL deal to a broader examination of college football program’s spending.2

Following the COVID-19 pandemic, there have been notable changes in the financial expenditures for collegiate athletics. As you can see in Figure 1, there has always been an increase in expenditure over the years until the drop due to COVID-19; however, after the pandemic, spending rose quickly to massive, unseen amounts. This highlights the heightened relevance of financial analysis in current times.

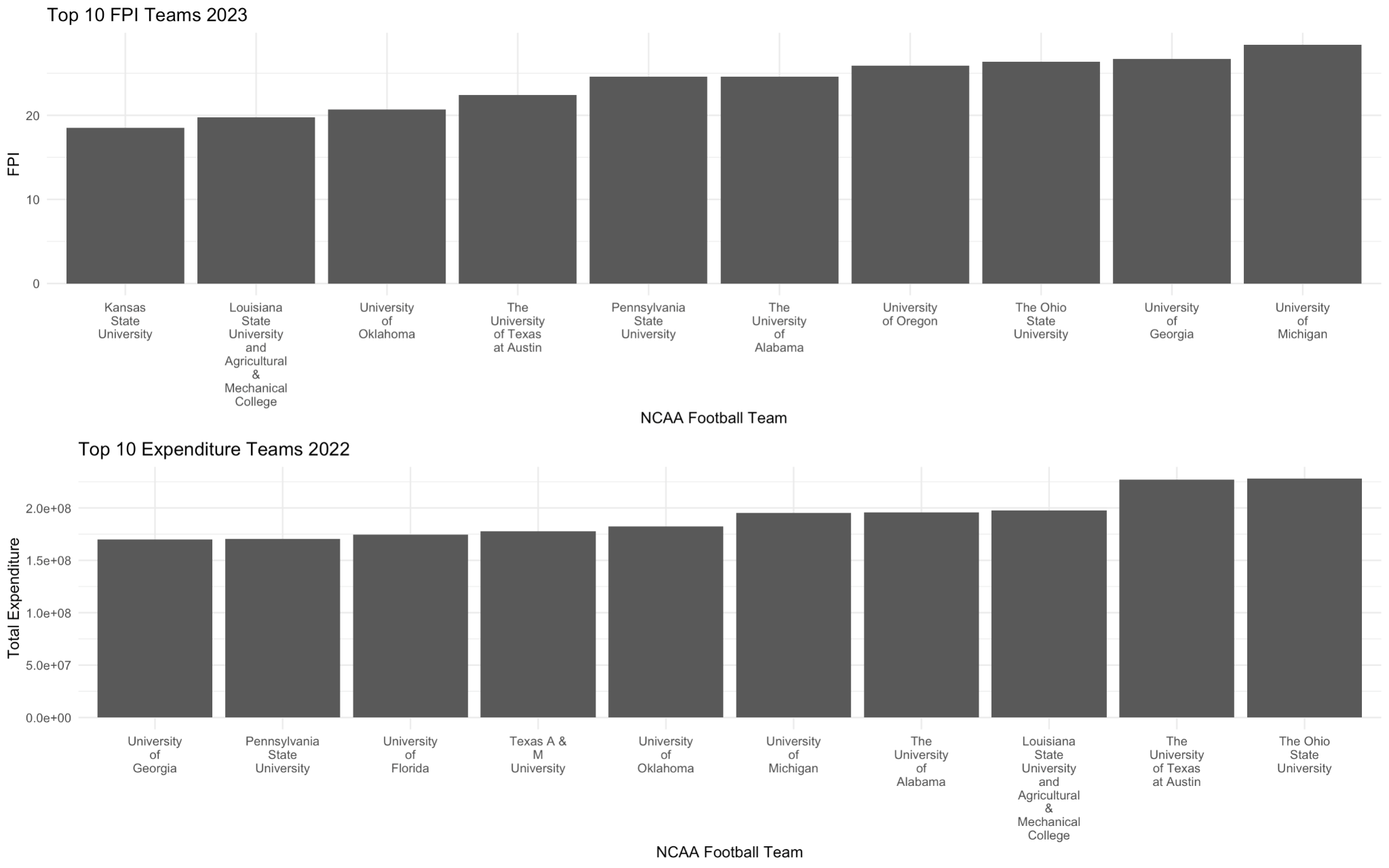
Moreover, the relationship between spending and performance is relevant not only to each individual program and their desire for success, but also holds broader implications regarding the equity of college athletics. This is evident in Figure 2, showing that eight of the top ten teams based on 2023 Football Power Index (FPI) were also a part of the top ten teams based on total expenditure in 2022. FPI is a measure of team strength that is meant to be the best predictor of a team's performance going forward for the rest of the season. FPI represents how many points above or below average a team is. Projected results are based on 20,000 simulations of the rest of the season using logical inputs. A team's initial FPI uses the previous four seasons with heavier emphasis towards more recent seasons, returning starters and head coach, and previous four recruiting classes as logical inputs. The FPI then adjusts based on in-season statistics including, opponent’s strength, game locations, final score margin, impact of offense and defense, strength of schedule, and days between games[[2]](#footnote-2). Ratings and projections are updated daily. The overlap apparent in Figure 2 suggests a potential link between financial investment and on-field success. Understanding how expenditure influences performance can inform strategic planning and resource allocation within football organizations, potentially altering financial planning strategies.

By analyzing these financial and performance metrics, the goal is to provide an informed and empirically based analysis of the relationship between finances and athletics in Power 5 college football. These findings have the potential to inform future NCAA regulations, shape financial planning within athletic departments, and unravel the complexities associated with the ‘pay-to-win’ philosophy in college sports.

*Figure 1*

****

*Figure 2*



1. **Discussion of Data**

The data that was gathered for the analysis encompasses two main categories: financial data and performance metrics, with each coming from two unique sources. The first dataset from Knight-Newhouse College Athletics Database focuses on revenues and expenses related to college athletics[[3]](#footnote-3). This database was created in collaboration with Syracuse University’s Newhouse School of Public Communications and Knight Commission on Intercollegiate Athletics. This online database contains self-reported data from institutions via NCAA financial reports. This dataset is measured to offer insights into the revenues and expenses in college athletics. The data collected covers 52 division one, power five conference teams. Each team's revenue breakdown sheds insight into the monetary influences from external sources. Investors such as corporate sponsors are an important influence on the budget along with that comes a say in where their money will be spent. Furthermore, the dataset breaks down the expenditures for each college to show the areas where programs spend the most money. Unlike the revenue variables, expenses show spending actions taken by the universities themselves. The expense variables account for on team costs such as the salaries of coaches and staff along with aid granted to the athletes themselves. Together, the revenue and expense variables from this dataset operate together to summarize monetary decision making for each collegiate program.

This fiscal data was originally collected to ensure effective oversight of college football across all Division 1 schools and to improve accountability. In 2010, the Knight Commission on Intercollegiate Athletics called for greater transparency of athletic finances by institutions, including better measures to compare trends in spending. The data was obtained from NCAA Division I institutions that are required to provide information based on state public records laws. The data are self-reported by institutions via NCAA Financial Report Forms, which are filed with the NCAA annually.

The second dataset includes performance metrics related to the on-field performance and team rankings and was gathered from ESPN NCAA Football rankings for the years 2014 to 2023 These metrics are recorded to best analyze NCAA football teams to understand team and player performance. This is utilized by ESPN in sports reporting and coverage of NCAA football games. Each year, ESPN evaluates every Division I football program for their efficiency on the field. This analysis is done through the Football Power Index (FPI)[[4]](#footnote-4) which is used as an evaluation tool for comparing the offensive and defensive performances of each team. FPI is a predictive tool developed by the company. Its goal is to use team strength to determine likelihood of results throughout the season. Thus, a higher FPI score would indicate that based on prior efficiency ratings a team is more likely to be successful on the field. A lower score doesn’t necessarily indicate a team is worse; instead, it would show that given recent efficiency data a team is performing at a lower level. Overall, the FPI tool developed by ESPN provides an evaluation method for the performances of the 52 teams listed in the revenues and expenses dataset.

To acquire this performance data, a web scraping technique was utilized which involved reading HTML content from web pages, selecting specific elements using CSS selectors, and converting the extracted data into a usable format for analysis in R, and scraping the information from each year into a separate CSV file. This method ensured that there was accurate and up-to-date data directly from ESPN’s website. In order to scrape it from the website, a function was built that takes two arguments: ‘season’ and ‘stat’. Season refers to the year of the season of interest and stat refers to one of the three tabs of ESPN’s data table, called ‘FPI’, ‘Resume’ and ‘Efficiencies’. The function will throw an error if the stat or season is unavailable (it is specified for inputs to be after 2010 although the data goes back to 2005). The function builds URLs based on the inputs, then scrapes from the provided URL. If stat is "FPI", it extracts team rankings and their corresponding FPI scores. If stat is "RESUME", it retrieves additional resume-related rankings. If stat is "EFF", it fetches efficiency-related rankings. The scraped data is then processed and formatted into data frames specific to each statistic type.

Subsequently, these datasets were joined to create a unified dataset that contained information regarding financial investment and athletic performance in Power 5 college football for each university. The process involved using a common identifier for each university: the IPEDS ID (Integrated Postsecondary Education Data System). Due to slight variations in university names between the datasets, schools were matched based on their IPEDS ID. This resulted in a complete dataset with university’s names and IDs matched with the twenty expenses and revenues variables and ten performance metrics for a span of ten total years. Several variables were created to further understand the data. First, a change in FPI and efficiency rank was added from year to year. Next, two ratios were created: total expenses over total revenues and total football spending over total expenses. Although ratios do not depict total amounts, they were created to normalize the two variables, allowing for a common financial metric to compare across universities. Additionally, changes in ratios can highlight the shifts in financial resource allocation, offering potential insight to what improves on-field performance.

While a considerable amount of data pertaining to financial and performance metrics was gathered for the analysis, there are specific data points that were not publicly available, but would have been valuable to include. Namely, access to NIL (Name, Image, and Likeness) data and recruitment rankings for certain universities was limited. The absence of NIL data can be attributed to the very recent implementation of regulations in this area as recent as 2021. Furthermore, much of the revenue generated by athletes often bypasses direct university channels, originating instead from external donors and sponsors. Regarding recruitment rankings, the dataset contains numerous missing values for this variable. This is primarily because the rankings typically range from 1 to 100, which may not encompass all Power 5 college football teams. Therefore, some universities' recruitment ranks were not available for inclusion in the analysis. Consequently, this data has not been tied into the final dataset, but is available if needed for further exploration.

By analyzing the available data, the potential question is whether spending can impact performance in college football. This idea leads into the question of does higher spending on college football programs correlate with increased success in terms of performance metrics? The attributions to success can stem from distinct sections, so this idea can be broken down further. First, focusing on distinct areas of spending and how those variables can correlate with a team’s on-field performance. Also, by investigating the longitudinally of the data and how the relationship between spending and team success has changed in each year.

1. **EDA/Analysis Discussion**

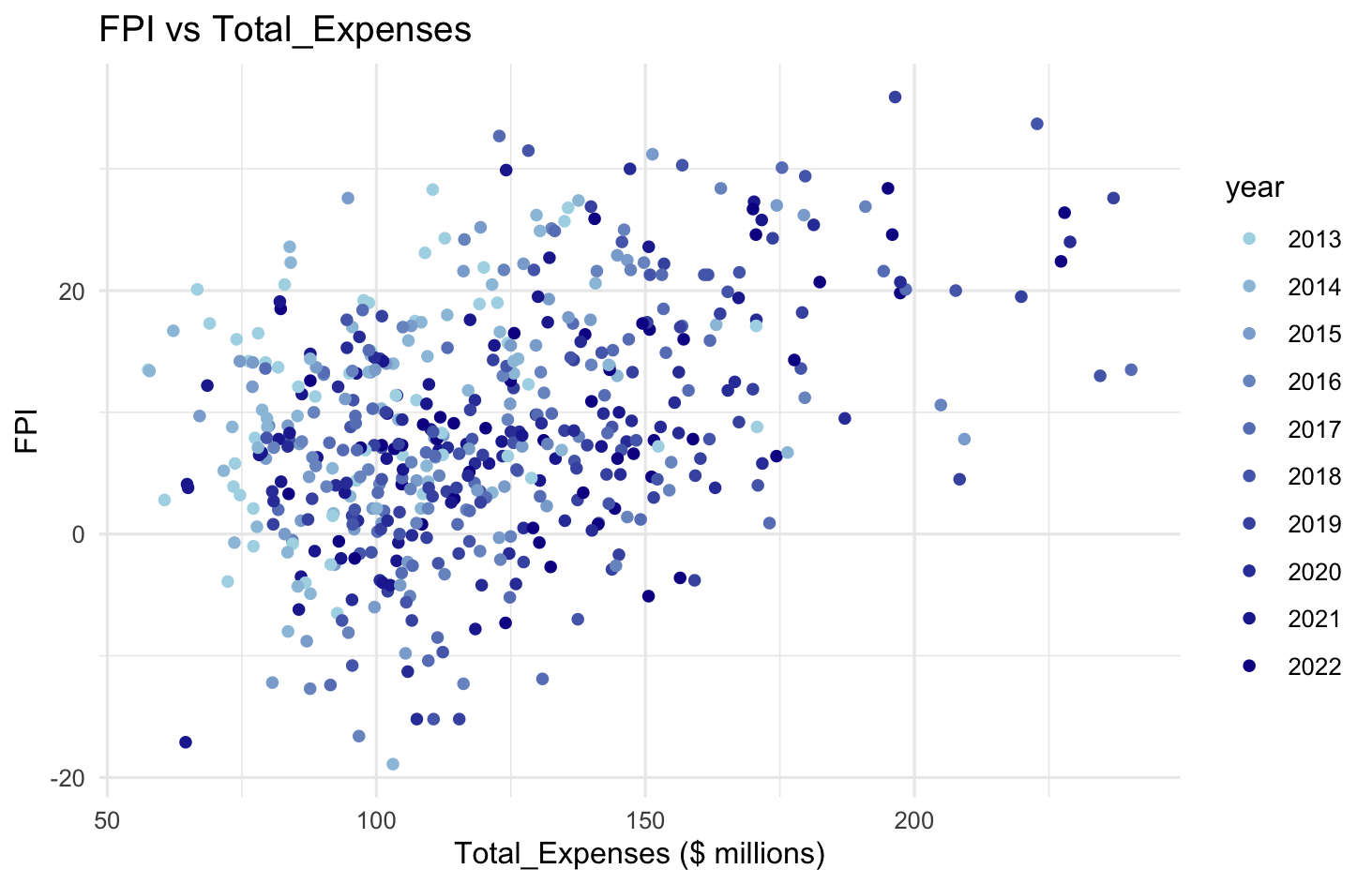
To answer the question of whether spending can impact performance in college football, a LASSO regression analysis was selected. LASSO Regression, or Least Absolute Shrinkage and Selection Operator, is a type of linear regression that uses shrinkage. LASSO regression shrinks coefficients towards zero. The LASSO procedure encourages simple, sparse models (models with fewer parameters). This regression is well-suited for the current models with 24 potential predictors and only 52 observations per year. The goal of this model is to predict whether the amount of money spent and the places where that money is spent indicates on-field success in power 5 division 1 football teams. LASSO regression avoids high variance by including a penalty term to balance the model complexity and goodness of fit, leading to more stable and interpretable models. LASSO allows aids in variable selection to narrow down the most relevant predictors while maintaining an interpretable model. These attributes are key to creating a successful model for the data.

During the beginning exploratory stage of the analysis, the FPI variable was deemed to be a good initial response variable for the data because it is a measure of team strength that is meant to be the best predictor of a team's performance for a season. FPI represents how many points above or below average a team is. The range of the FPI variable is -18.9 to 35.9 in this dataset. 0 is considered an average team. The data was transformed into a wide format to conduct multiple year wise analysis. Delta FPI and efficiency variables were created as well as total and total football expenses and revenues ratios as mentioned above. However, the new variables yielded no new information and were removed from the data frame. After manipulating the data, the modeling process started with a stepwise multiple linear regression for each individual year, spanning from 2014 to 2023 FPI. The yearly FPI was used as a response variable with numerous predictors including all other expense and revenue variables (omitting NA values) and the previous year’s FPI as a “baseline measure of FPI” for the 52 universities. This stepwise regression served to establish an initial relationship between FPI and the significant predictors of the dataset. However, the large number of possible predictors caused potential overfitting in the stepwise regression. All variables in the model are monetary and continuous and have the potential to be correlated with others. The multicollinearity is reduced by looking at each yearly model; however, it is still a potential area of concern with the data’s structure. The LASSO model mitigates this issue. LASSO handles multicollinearity more effectively than stepwise by directly penalizing the absolute size of coefficients. These yearly models and LASSO regression’s feature selection alleviate multicollinearity among the variables.

Ridge regression was also investigated as a possible analysis method. However, after further research, LASSO was chosen to maintain clear interpretability of the model and answer the research question effectively. Ridge regression does account for multicollinearity, but does not aid in needed variable selection for the data. Ridge regression keeps all predictors in the model and helps provide information about weakly correlated predictors; but, the goal for this analysis is to find where universities can get the biggest bang for their buck. Hence, the goal is to find the most influential predictors.

After better understanding the data through stepwise regressions, some exploratory data analysis was conducted to ensure that a LASSO regression is an appropriate model choice. The assumptions needed for LASSO are that the data must be linearly correlated, the errors need to be independent of each other, the model must be sparse, and an optimal lambda should be chosen to minimize the MSE. It is important to note that each assumption was checked in each yearly model to ensure that the data does meet these assumptions. The structure of the analysis guarantees the independence of errors, by creating yearly models that ensure independent records because the model analyzes 52 unique teams in one year. By creating the independent yearly models, it still allows for the investigation of significant trends over time, but ensures that assumptions are met. The ten years of data are not being included in one singular model; they are being used in ten separated models for each season. In terms of linearity, the predictors were examined to ensure that they exhibit linear relationships with the target variable, future year’s FPI. Scatter plots of each monetary variable were used for visualizing these relationships. A few notable example scatter plots are shown below. Figures 3 through 6, look at the scatter plots of several variables vs FPI. The color corresponds to the year. These four variables (total expenses, football spending, coaching salaries, and recruiting) appear to show a linear relationship with FPI and highlight how spending has increased over the years with the darker dots corresponding to more recent years. These few variables were noted as important during the variable selection process. One interesting finding in figure 4 is that recruiting spending drops drastically in 2022. This is most likely due to COVID-19.

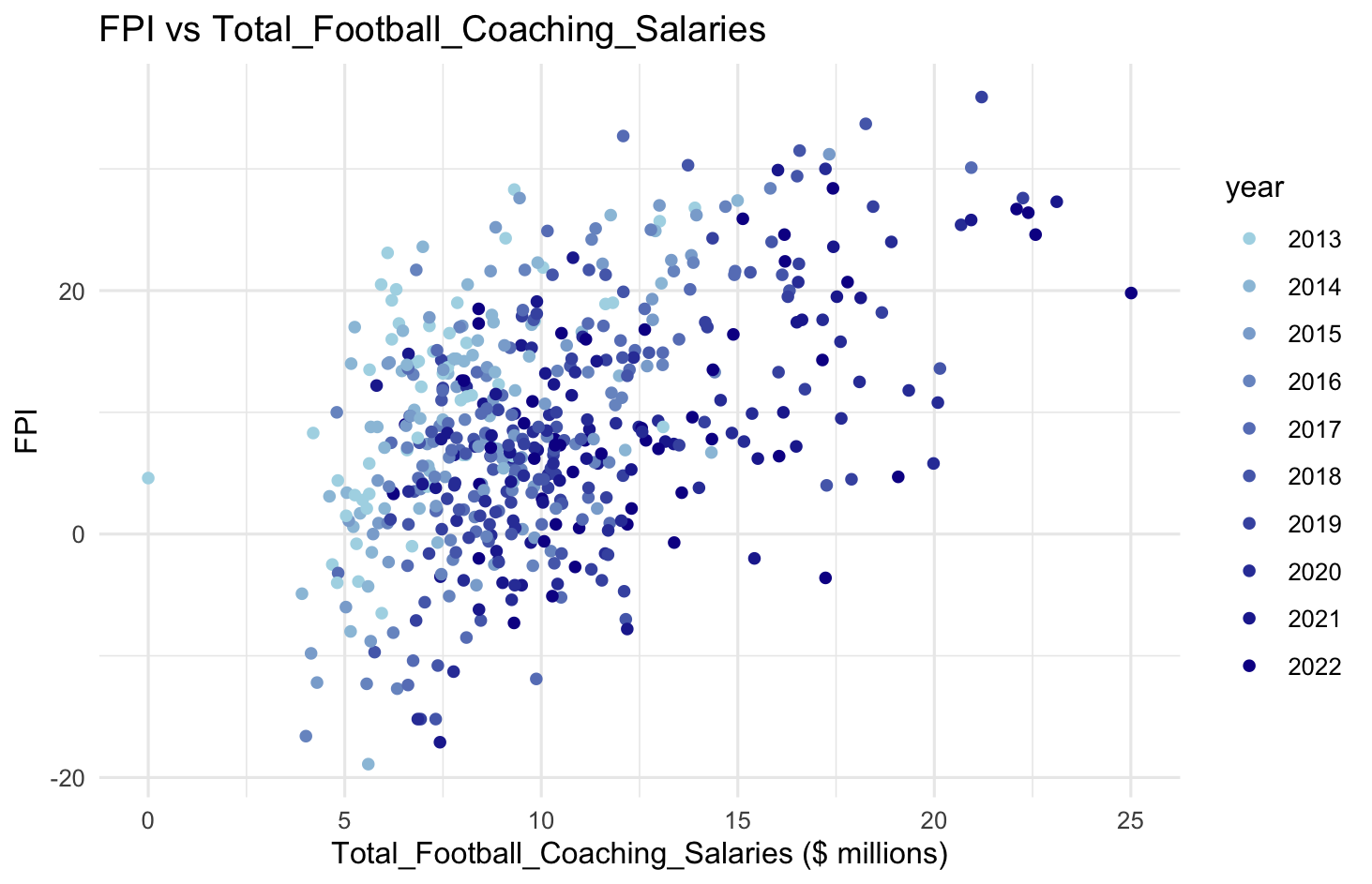
*Figure 3*

**

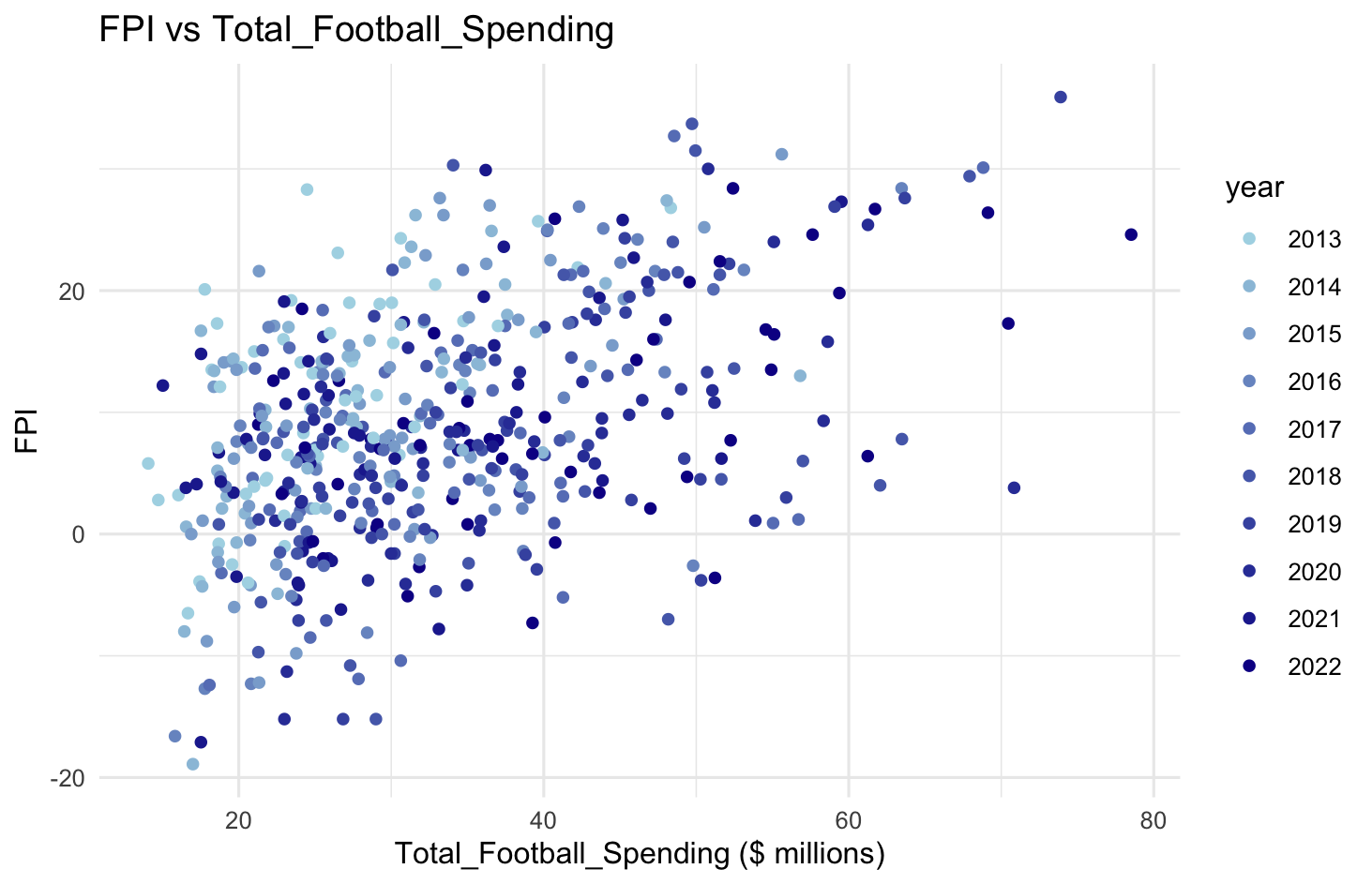
*Figure 4*

**

*Figure 5*

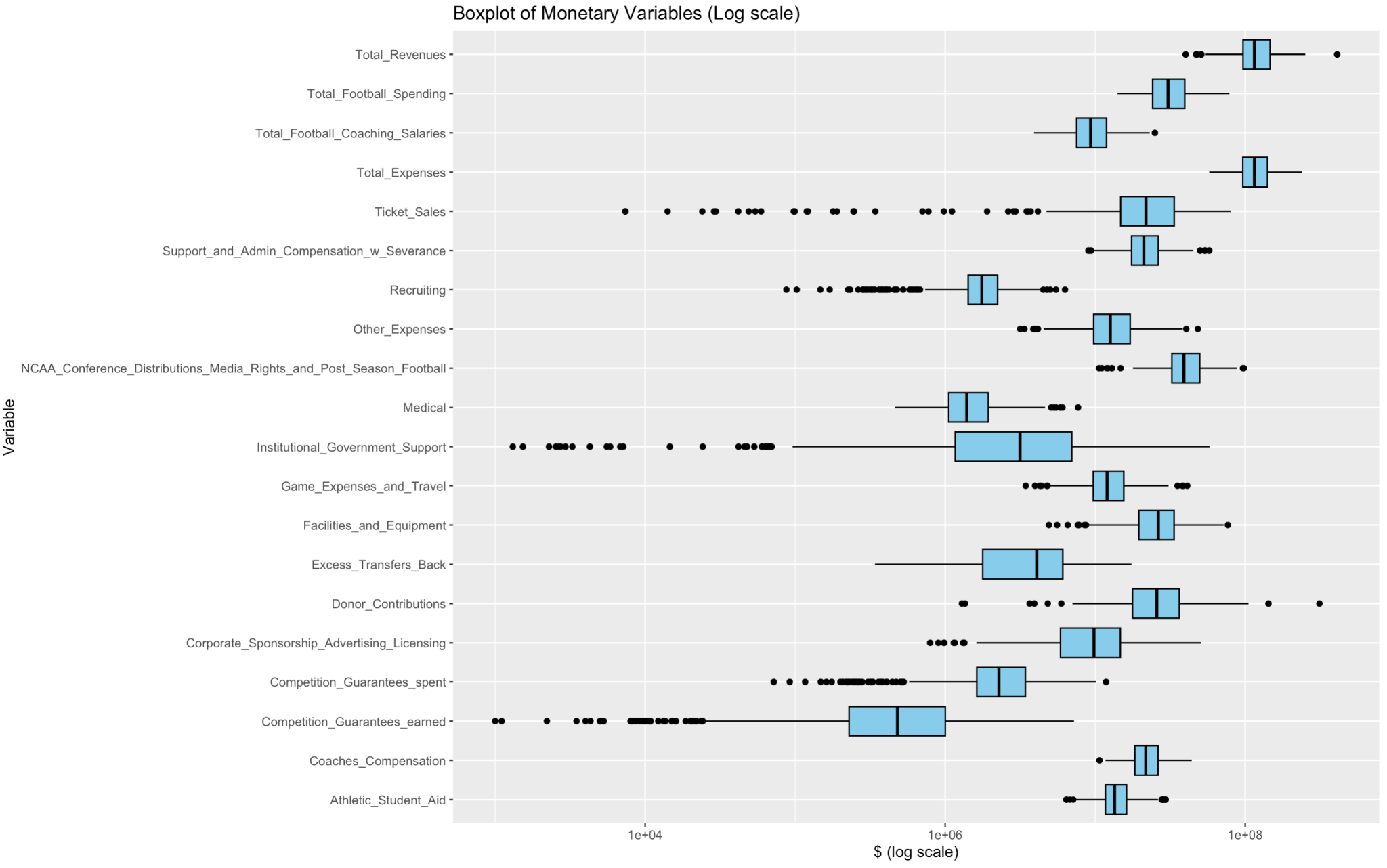
**

*Figure 6*

**

In terms of the distributions of each predictor variable, figure 7 shows boxplots for each variable on a log scale. Ticket sales have the most low-end outliers probably due to COVID-19 in the 2022 season. Institutional government support also has many low-end outliers because many schools received little to no government support over the years. Competition guarantees earned also have low-end outliers due the lack of travel in 2022 probability due again to COVID. Each variable's box plot shows little to no high-end outliers and concise spreads of each variable. Figure 7 is helpful in understanding variable distributions and assessing data integrity, ultimately contributing to the development of a robust and reliable LASSO regression model.

*Figure 7*



After the data was investigated and LASSO was deemed an appropriate analysis tool, the final model building process began. The wide formatted data was subset to each specific year. The model set up was similar to that of the previous stepwise regression. The future year FPI was the response vector and the previous year monetary variables along with previous FPI were turned into a matrix for the model R code. From there, cross validation was performed to find the optimal lambda that minimizes the MSE of the model. The yearly data was then fitted with the optimal lambda and the RMSE was calculated. This was repeated for each year from 2014 to 2023.

In general, LASSO regression can lose some information when variables are correlated because it retains only one of the variables and sets other correlated variables to zero. The interpretability of a LASSO regression is considered good when it selects a small number of features with non-zero coefficients, making it easier to identify the most important predictors.[[5]](#footnote-5) This is also the case with the ten year by year models. For each model, the number of predictors dropped from twenty-one to anywhere between three and ten predictors. Table 1 below shows the significant predictors and their coefficients for each LASSO model. The values in the table are the coefficients associated with each predictor for each model. The root mean square error (RMSE) is also reported in the table below to show model accuracy (The RMSE measures the average difference between the actual and predicted values of the target variable). The benefits in the LASSO regression models are variable selection, sparse modeling, and interpretation of the non-zero coefficients. In this dataset with many monetary predictors, LASSO regression highlights the most influential expenses that college football teams have, aiding in understanding the underlying relationships, as seen in the table below. Sparse models are easier to understand and communicate, especially in situations where transparency and clarity are crucial, such as in decision-making processes about university spending. Finally, the non-zero coefficients in a LASSO model provide direct information about the magnitude and direction of the relationships between the predictors and the response variable. Table 1 shows the previous year FPI has the largest effect on predicting the follow years FPI. This is not surprising as it is meant to be a baseline measure of how good a team was at the end of the previous season. Other predictors such as coaches competitions, total football coaching salary, competition guarantees spent, total football spending, and recruiting had significant impacts of predicting the following years FPI. Three of these variables (total football coaching salary, total football spending, and recruiting) showed up in the exploratory data analysis as areas of focus. However, total expenses never showed up as a significant predictor in any of the models. Finally, table 1 is a summary of the information used to aid in the general final model to find the best predictors for any FPI year.

*Table 1*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2013-2014** | **2014-2015** | **2015-2016** | **2016-2017** | **2017-2018** | **2018-2019** | **2019-2020** | **2020-2021** | **2021-2022** | **2022-2023** | **Appearances** |
| **Previous FPI** | 4.6 | 5.04 | 5.68 | 4.83 | 5.26 | 5.48 | 5.29 | 4.16 | 4.04 | 4.9 | 100% |
| **Competition Guarantees Spent** | 0.71 | 0.31 | 2.32 |  |  |  |  | 0.15 |  |  | 40% |
| **Total Football spending** | 1.68 | 1.07 | 1.97 |  |  |  |  | 0.45 | 0.28 |  | 50% |
| **Total Football Coaches Salary** | 0.49 |  |  | 2.42 | 0.71 | 0.74 | 2.3 | 0.57 |  |  | 60% |
| **Donor Contributions** |  |  |  |  | 0.21 |  |  |  |  |  | 10% |
| **Ticket sales** |  | 0.57 | 0.98 |  |  | 0.37 |  |  |  | 0.68 | 40% |
| **Government Support** |  | 0.67 | 0.96 |  |  |  |  | 1.2 | 1.15 |  | 40% |
| **Other Expenses** |  |  | 1.83 |  |  |  |  |  | 1.88 |  | 20% |
| **Medical** |  |  | 0.86 | 0.86 |  | 0.44 | 1.39 |  |  |  | 40% |
| **NCAA Media Rights** |  |  |  |  | 0.79 | 0.67 |  |  |  | 1.11 | 30% |
| **Athletic Student Aid** |  |  | 0.46 |  |  |  | 0.94 |  |  |  | 20% |
| **Recruiting** |  |  | 0.75 |  |  |  |  | 1.39 | 1.49 | 1.55 | 40% |
| **Coaches Compensation** |  |  |  |  |  | 2.56 |  |  | 0.45 |  | 20% |
| **Game and Travel Expenses** |  |  | 1.47 |  |  |  |  |  |  | 0.79 | 20% |
| **RMSE** | 4.94 | 5.52 | 5.12 | 6.36 | 6.30 | 6.21 | 5.37 | 4.89 | 5.33 | 5.12 |  |

1. **Results**

After looking at the significant predictors of next season’s FPI from each year, the year over year data of each of these predictors was averaged and used to predict 2023 FPI as a final example response variable. Once averaged, an initial LASSO model was created using every predictor that showed significance in any year, producing a MSE of 40.29 and an RMSE of 6.35, and the largest coefficient in this model was previous season’s FPI at around 2.4, with the rest around 1.0. In this context, the RMSE is used to interpret the spread of the errors of the model, meaning a range of 12.7 points would be captured by any school in this model. Next, a second model was built using only the predictors that appeared in year over year models that produced an MSE <30, which includes FPI, recruiting, total football spending, total coaching salaries, institutional government support, competition guarantees earned and spent, other expenses, coach’s compensations, excess transfers back, and game expenses and travel. In this model, the previous season’s FPI and recruiting were the largest of the LASSO coefficients at just under 3, while the other two were about 0.5. This model produced an MSE of 41.14 and an RMSE of 6.41, so it was decided that a LASSO would be done on the resulting original significant predictors from the first LASSO, which produced an MSE of 38.35 and an RMSE of 6.193, lower than any other model attempted. This final model includes the predictors, FPI, donor contributions, recruiting, other expenses, and ticket sales. The coefficients for this model were 2.6 for previous season’s FPI, 2.0 for recruiting, 1.4 donor contributions, and about 0.7 for both other expenses and ticket sales. These predictors were chosen as the most important monetary variables to predict next year’s FPI.

1. **Conclusion**

According to the models, if a team is working on prioritizing where program spend is allocated, it would be best to focus efforts on recruiting and donor contribution expenses, since they were most influential in predicting a greater subsequent FPI. However, considering that FPI is a game-point unit of measurement, and the best model produced an MSE of ~38 and an RMSE of ~6.2, it is quite evident that FPI is hard to predict, at least from data that is publicly available. To identify more accurate predictors, it would be most beneficial to gain access to data that may not be currently publicly available, such as NIL spend and team strategies. It is also worth noting that it is not only money that makes or breaks a college football team. There are a lot of variables that play into a successful college football season that are not easily quantifiable, such as team legacy and reputation, as well as the skill sets and efforts of many individual players, coaches, and coordinators.

Looking ahead, it is important to remember that the Power 5 conferences have been reduced to four, with the historically top two performers (Big10 and SEC) being bolstered by strong additions from the weaker conferences. This change effectively makes certain conferences stronger than others which affects scheduling and therefore strength of schedule measures as it factors into FPI.​​ It is also important to note that we are entering a new era of NIL expenditure in college athletics, and its effect on aggregate data and how it affects a team’s performance in any given year cannot yet be known.

**Appendix 1 – Data Dictionary**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Type** |
| Data | The name of the Division 1 institution | Categorical |
| IPEDS\_ID | A unique six-digit identification number assigned to U.S. postsecondary institutions in the U.S. Department of Education Integrated Postsecondary Education Data System (IPEDS) | Quantitative |
| Year | The year on which the data is collected | Quantitative |
| FBS\_Conference | The name of the Power 5 conference in which the university participates | Categorical |
| Total\_Expenses | Total athletic operating expenses, including athletic scholarship costs. | Quantitative |
| Excess\_Transfers\_Back | Positive net revenues generated by athletics and transferred to the institution for non-athletics purposes | Quantitative |
| Other\_Expenses | Sports equipment, uniforms and supplies, fundraising, marketing and promotion, sports camps, spirit groups, direct overhead and administrative expenses, indirect institutional support, membership and dues, student-athlete meals, and, other operating expenses | Quantitative |
| Medical | Medical expenses and medical insurance premiums. | Quantitative |
| Competition\_Guarantees\_spent | Amounts paid to visiting participating institutions, including per diems and/or travel and meals. | Quantitative |
| Competition\_Guarantees\_earned | Revenue received from participation in away or neutral-site games. | Quantitative |
| Recruiting | Spending on transportation, lodging, meals, and other personnel and administrative expenses relating to recruitment of prospective student-athletes | Quantitative |
| Game\_Expenses\_and\_Travel | Game expenses relate to competition expenses other than travel. | Quantitative |
| Facilities\_and\_Equipment | Facility expenses include debt service, leases, and rental fees for athletic facilities. This includes overhead and administrative expenses. Equipment expenses include spending for items provided to teams, including in-kind equipment. | Quantitative |
| Coaches\_Compensation | Coaches compensation includes bonuses and benefits, but not severance payments. | Quantitative |
| Support\_and\_Admin\_Compensation\_w\_Severance | Support and administrative staff compensation includes bonuses and benefits paid to all administrative and support staff | Quantitative |
| Athletic\_Student\_Aid | Total expenses for athletic student aid, including tuition and fees, room and board, books, summer school, tuition discounts, and waivers, including aid given to student-athletes who have exhausted their eligibility or who are inactive due to medical reasons. | Quantitative |
| Total\_Revenues | Total revenues for the athletics program minus “Less Transfers to the Institution. | Quantitative |
| Other\_Revenue | Compensation and benefits provided by a third party; game program, novelty, parking and concession sales; sports camps and clinics; athletics restricted endowment and investments income; and, other operating revenue. | Quantitative |
| Corporate\_Sponsorship\_Advertising\_Licensing | Revenue generated by the institution from royalties, licensing, advertisements and sponsorships. | Quantitative |
| Donor\_Contributions Competition\_Guarantees\_earned | Funds contributed from individuals, corporations, associations, foundations, clubs or other organizations external to the athletics program above the face value for tickets | Quantitative |
| NCAA\_Conference\_Distributions\_Media\_Rights\_and\_Post-Season\_Football | Revenue received from the NCAA (including championships) and athletics conferences, media rights, and postseason football bowl games. | Quantitative |
| Ticket\_Sales | Revenue received from ticket sales for all NCAA-sponsored sports at an institution | Quantitative |
| Institutional\_Government\_Support | Revenue received from governments, direct funds from the institution for athletics operations, and costs covered and services provided by the institution to athletics (and for athletics debt) but not charged to athletics | Quantitative |
| Student\_Fees | Fees paid by students and allocated for the restricted use of the athletics department. | Quantitative |
| Total\_Institutional\_Government\_Support\_and\_Student\_Fees | Combination of Institutional/Government Support and Student Fees. | Quantitative |
| Total\_Football\_Spending | Total football operating expenses, including the cost of athletics student aid. | Quantitative |
| Total\_Football\_Coaching\_Salaries | Total compensation reported for all football coaches, including salaries, benefits and bonuses paid by the university, and contractually-guaranteed amounts paid by third parties. | Quantitative |
| fpi | Football Power Index that measures team's true strength on net points scale; expected point margin vs average opponent on neutral field | Quantitative |
| rk | Football Power Index Rank vs all FBS teams. | Quantitative |
| eff\_ove | Net efficiency on 0-100 scale; incorporates offense, defense and special teams efficiencies into a single schedule-adjusted measure of per-play efficiency. | Quantitative |
| rnk\_ove | Team's overall efficiency rank among all FBS teams. | Quantitative |
| eff\_off | Offensive efficiency on 0-100 scale; based on offense's contribution to scoring margin on per-play basis, adjusted for strength of opposing defenses faced. | Quantitative |
| rnk\_off | Team's offensive efficiency rank among all FBS teams. | Quantitative |
| eff\_def | Defensive efficiency on 0-100 scale; based on defense's contribution to scoring margin on per-play basis, adjusted for strength of opposing offenses faced. | Quantitative |
| rnk\_def | Team's defensive efficiency rank among all FBS teams. | Quantitative |
| eff\_spe | Special teams efficiency on 0-100 scale; based on special teams' contribution to scoring margin on per-play basis, adjusted for strength of opposing special teams faced. | Quantitative |
| rnk\_spe | Team's special teams efficiency ranks among all FBS teams. | Quantitative |
| ExpRev | Total\_Expenses / Total\_Revenues | Quantitative |
| FBExp | Total\_Football\_Spending / Total\_Expenses | Quantitative |

1. Haworth, Christian. “[Spending Smart: A Comparative Look at Areas of University Spending within Football Programs and Their Effect on Wins from 2005-2022](https://www.samford.edu/sports-analytics/fans/2023/Spending-Smart-A-Comparative-Look-at-Areas-of-University-Spending-Within-Football-Programs-and-Their-Effect-on-Wins-From-2005-2022#:~:text=The%20analysis%20reveals%20a%20statistically,in%20the%20number%20of%20Wins.).” Samford University, November 7, 2023.

   2 “[College Football’s Bottom-Line Impact: Exploring the Relationship of Football Performance on Athletic Finances for Division I Institutions Today](https://thesportjournal.org/article/college-footballs-bottom-line-impact-exploring-the-relationship-of-football-performance-on-athletic-finances-for-division-i-institutions-today/).” The Sport Journal, August 20, 2021. [↑](#footnote-ref-1)
2. Video explaining FPI: <https://www.espn.ph/video/clip?id=23307329> [↑](#footnote-ref-2)
3. Precise definitions of the individual revenue and expense sources from the Knight-Newhouse College Athletics Database are given in the data dictionary found in Appendix 1 [↑](#footnote-ref-3)
4. Precise variables used from the ESPN FPI Database are given in the data dictionary found in Appendix 1 [↑](#footnote-ref-4)
5. <https://www.sciencedirect.com/topics/psychology/lasso-regression> [↑](#footnote-ref-5)