CS 6630 Visualization for Data Science - Project: Impact of Money in College Football
Group Members: Nick Stephenson, Andrew Potts, Jeff Kelso

Process Book

Overview & Motivation

The NCAA generates over $1 billion in revenue annually. In addition, the highest-paid public employee in the majority of states is either a college football or basketball coach. Therefore, in spite of being a non-profit organization, the NCAA is clearly out to be profitable. Thus, the question becomes, to what extent does money impact NCAA sports? As it is football season, we decided to narrow the scope of this question and instead focus on the impact of money on college football. Furthermore, many universities rely on their football team's revenue to cover the costs of other athletic programs, and it is often the case that the football program is the only profitable athletic program. Although none of us work with athletics data in our research scopes, we all enjoy watching college football and lament the perceived impact money has on the sport.

To analyze money's impact, we made a visualization that enables exploration of our college football dataset. The visualization allows the user to select which variables are displayed in the scatter plot so that different correlations can be spotted and analyzed further. This includes exploring the relationship between money and a team's success, and overall analyzing trends of money in college football. The visualization also allows the user to select subsets of the datasets to see how different conferences compare to the whole.

Related Work

The greatest inspiration for our visualization comes from the gap plot visualization we implemented in homework assignment #4. With over 100 teams that each have multiple variables to visualize, we believed this scatter plot layout would provide the best insight into our data.

Questions

The key question our visualization centers on is, how has the influence of money in college football changed over the years? To answer this question, we decided we also ought to answer the following questions:

- Has the infusion of money into college football changed over time?
- Does more revenues equate to more wins?
- Have coaches' salaries changed over the years?

- Is there a relationship between the size of the school and how much revenues the football program brings in?
- Is there a relationship between the number of infractions a team commits and its revenues?

Data

The majority of our data came from the U.S. Department of Education Office of Postsecondary Education. We pulled Excel files for each school that outlined the institution's college football revenue and expenses for 2003 through 2016. The Sports-Reference website had Win/Loss records. The coaches' salaries were the most difficult data to find. The USA Today website had data for 2006-2016 excluding 2008.

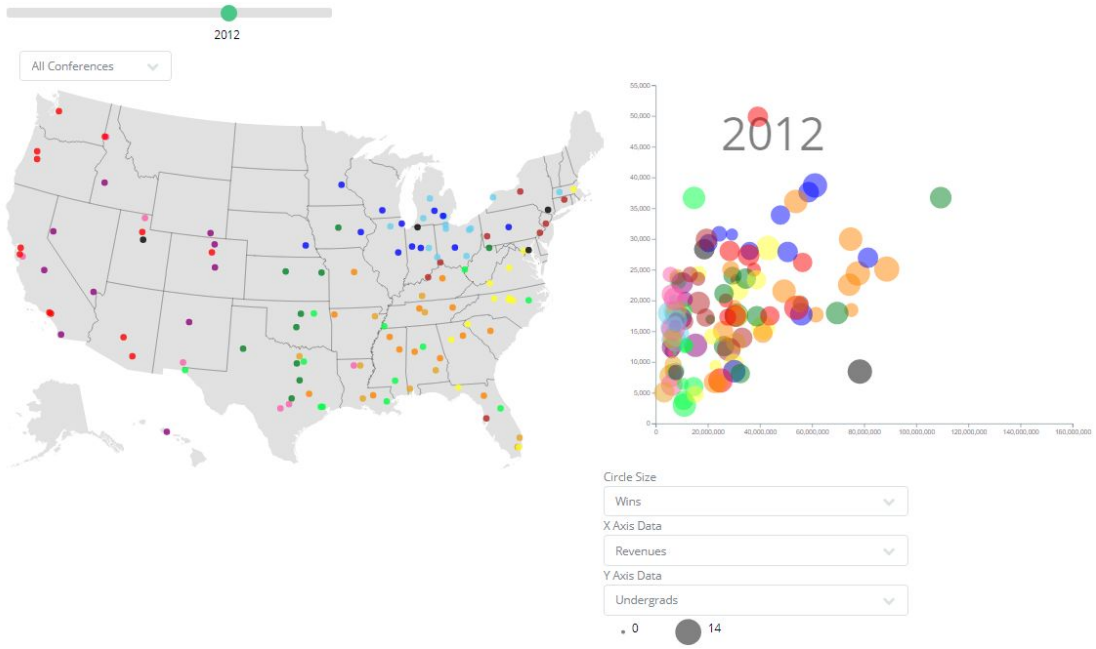
Exploratory Data Analysis

We did not use any "pre-visualizations" to look at our data. Our goal was to let our visualization provide the insights we were looking for.

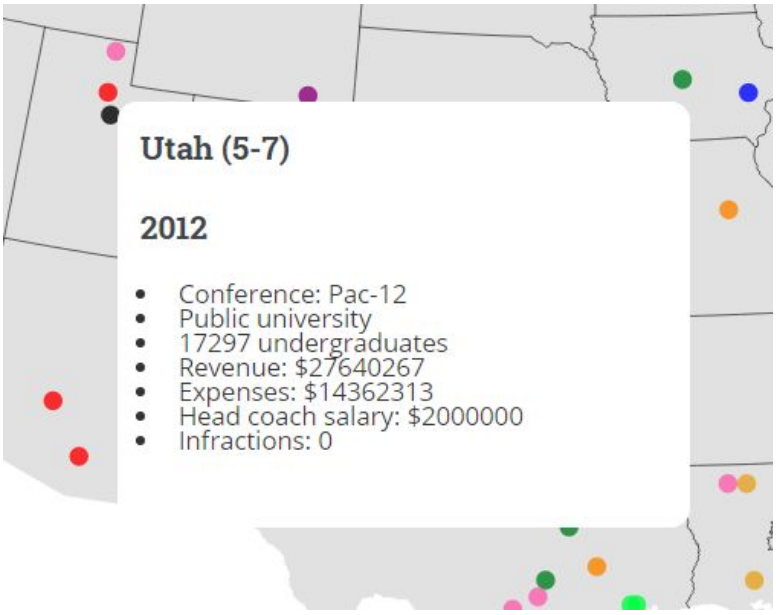
Design Evolution

Initially, the planned visualizations were a map that included the locations of the schools, a scatter plot similar to HW 4 that displayed interactions between variables we selected for our analyses, a table showing all the data for a school that was clicked on the map, and a dual-axis line chart showing trends over time. After our meeting with Sam, however, we decided to remove the table from the visualization. As Sam pointed out, the majority of the data in the table is already displayed elsewhere in the visualization. Furthermore, although a table may be a useful reference to look at specific data, it does little to add to the visualization as it is extremely difficult to detect patterns from a table. For these reasons, we decided to drop the table from our visualization.

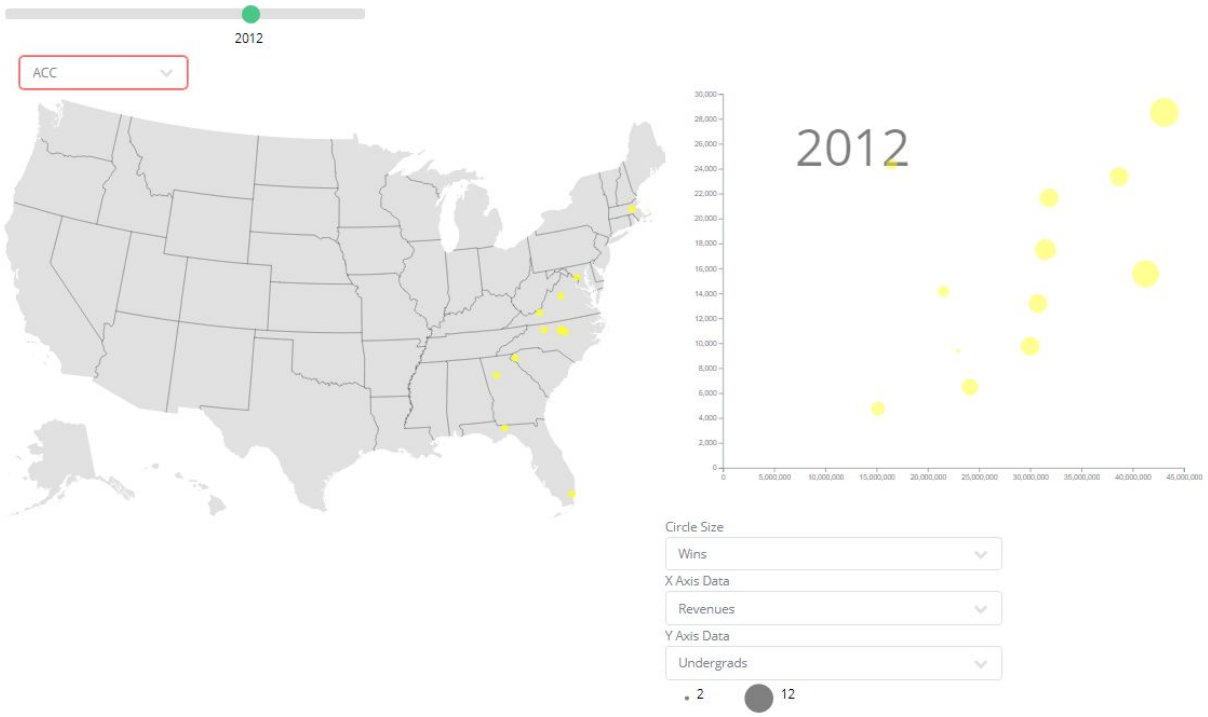
Implementation



Upon loading the web page, the user will see the two visualizations shown above. On the left is a year slider and a map of the US with each FBS football team encoded as a colored circle. The color represents the conference to which the team belongs. When the user hovers a specific circle, a tooltip appears that informs the user of the school, the conference, whether it is a public or private school, revenue for that year, expenses for that year, the head coach salary for that year, and the number of NCAA infractions for that year.



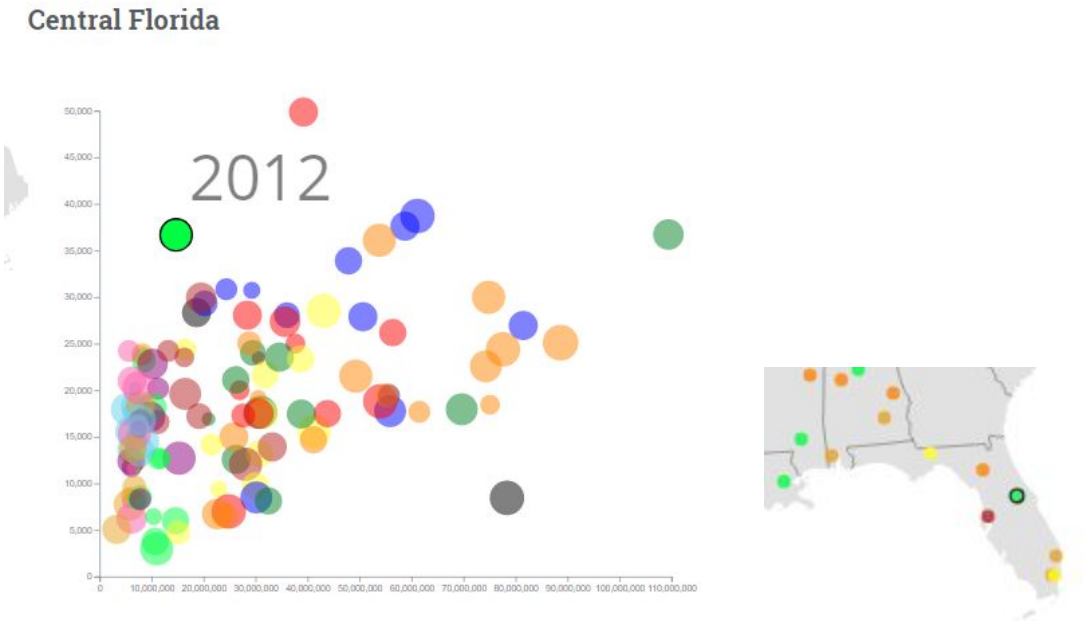
The drop down menu above the map allows the user to display the teams for a certain conference on both the map and the scatter plot, as indicated below:



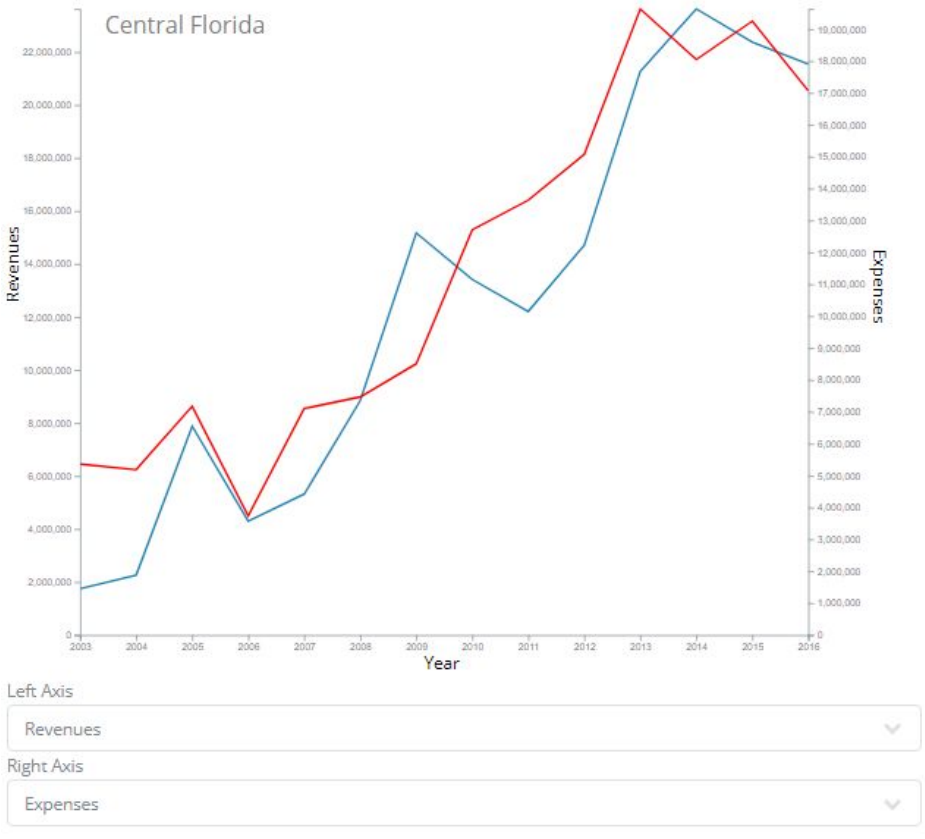
The scatter plot displays the year along with three dropdown menus that allow the user to select which data they want on the x and y axes and which data they want to encode as circle size. Again, the circles in the scatter plot are colored according to which conference the team belongs.

When the year slide is adjusted, the data in the scatterplot and the map is adjusted accordingly. Changes in the map were the result of some teams changing conferences or joining the FBS. This adjustment in the scatter plot is the key interaction that will inform the user of major changes over time. Through this interaction, the user can detect major patterns in each one of the variables included in the drop down menus.

If a user hovers over a circle on the scatter plot, the team appears in the top left corner. If the user clicks on a school on the map or on a team in the scatter plot, several things happen. The circle is highlighted with a black border on the scatter plot and on the map to let the user know where the school is located.



Also, a third visualization appears at the bottom of the screen.

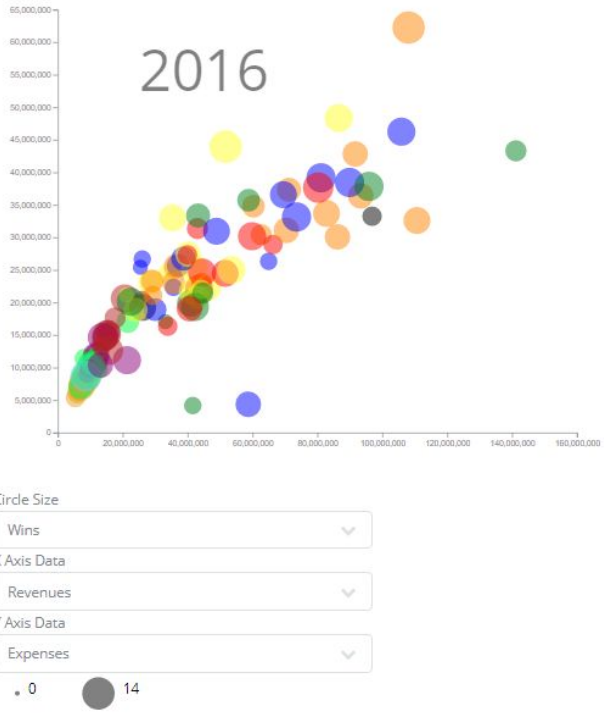


This visualization provides detailed time-series data for that specific school in the form of line graphs. In this visualization, the user can only update the two y axes; the x-axis reflects years. The scatter plot provides a macro-level visualization for what is happening across the country, and the double-axes line graph shows a micro-level visualization for each individual school. The user should take caution as the scales for the two-axes are different. Therefore, even though one line may appear to have a greater value, this could be a result of the two different scales.

Evaluation

Revenues and Expenses

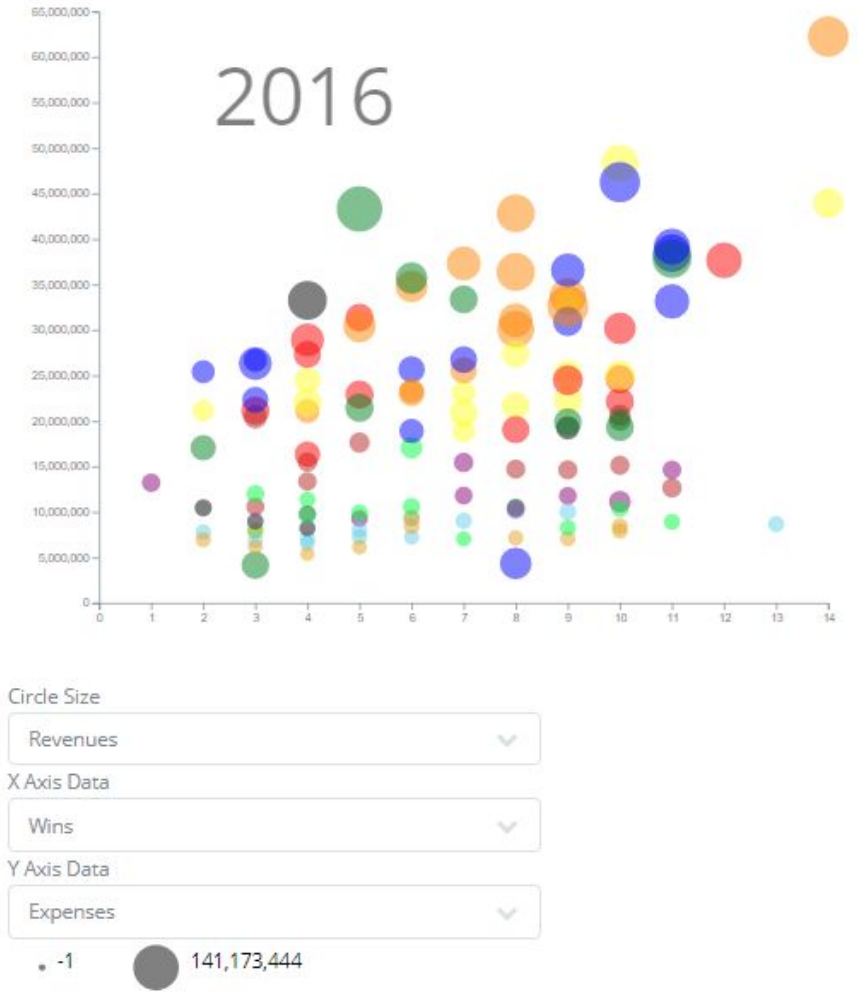
Revenues and expenses have consistently increased over time. The screenshots below show a side-by-side comparison of revenues and expenses for 2003 and 2016.



Not only have revenues and expenses increased, but the the revenue and expense differences between teams has increased dramatically. While most teams had comparable revenues and expenses in the past as indicated by the cluster in 2003, this is no longer the case by 2016. There is far greater dispersion in 2016, and we see a couple teams (Alabama - orange, Texas - green) pulling away from the main body, indicating their revenues and/or expenses greatly exceed that of their peers.

Revenues and Wins

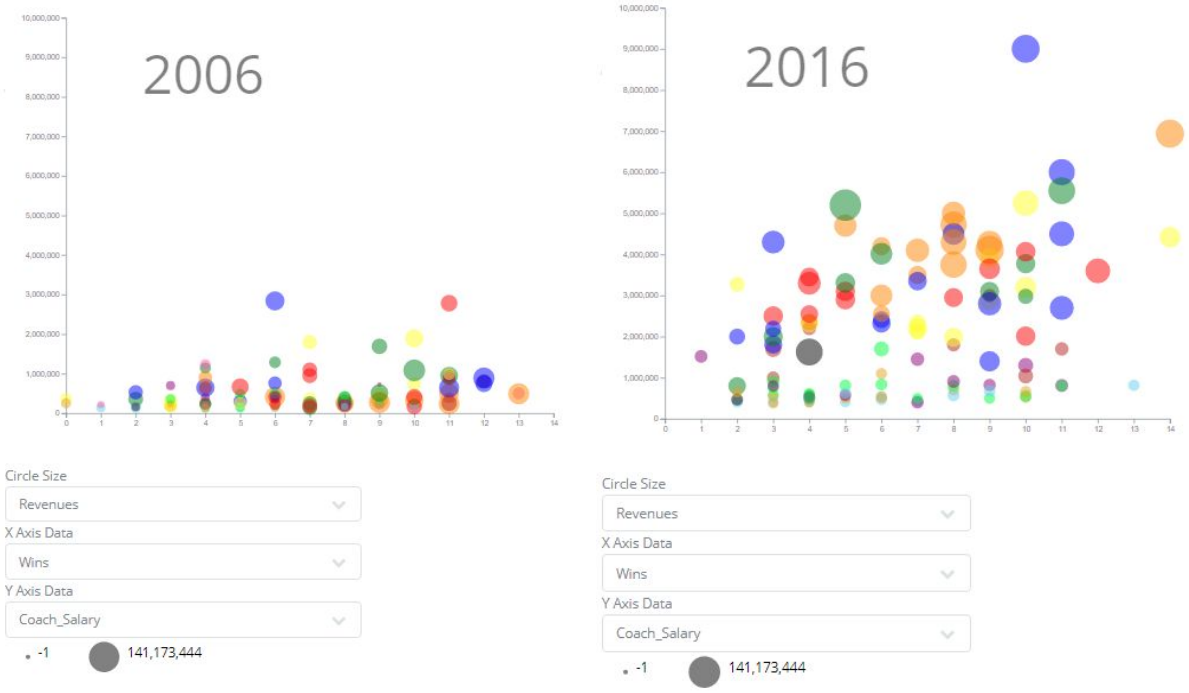
It appears there is a relationship between revenues and wins. For example, the picture below has wins along the x-axis, expenses on the y-axis, and revenues as circle size. We can see as the number of wins increases, so too does the circle size. While the correlation is not perfect, there seems to be a strong linear relationship. Although the image below is just a snapshot from 2016, the pattern holds true across the years we looked at - teams with larger revenues and expenses seemed to have more wins than the teams with less revenues and expenses.



Coaches' salaries

Although we are missing multiple years, we are still able to see how coaches' salaries have changed. If we look at 2006, which is the first year for which data are available, we can see that salaries are pretty similar, regardless of record or revenues -- even the two highest-paid coaches make about $3 million. When we compare 2006 salaries with 2016 salaries, again we

see how the dispersion has increased. There also appears to be a linear trend -- as the number of wins increases, so too does the salary amount. In addition, in 2006, only two salaries were around $3 million. In 2016, roughly 40 to 50% of salaries are at $3 million or above, with the highest around $9 million. In just ten years, the highest salary increased by about 200%, while the average salary saw an increase of well over $1 million.

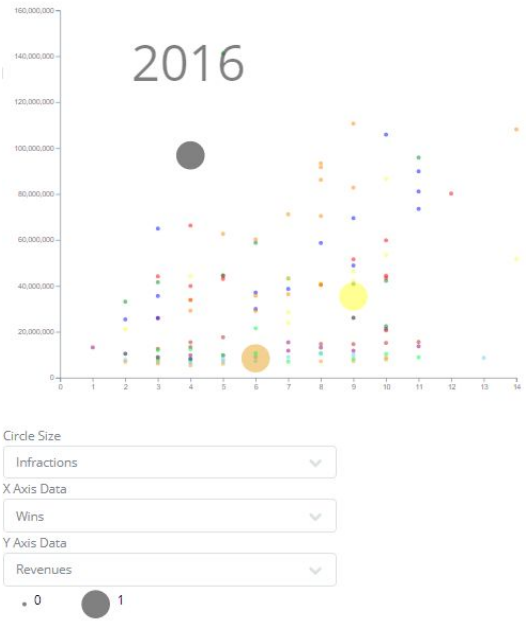


Undergraduate Population and revenues

There appears to be a weak linear relationship between a team's revenues and the size of the school, though this relationship appears to strengthen as time progresses. In 2003, there appears to be a very small relationship as teams with higher revenues appear to have the most undergraduates. However, we still observe teams with smaller revenues with undergraduate populations comparable to those teams with the highest revenues. In 2016, we observe a general upward trend -- as revenues increase, the undergraduate population tends to increase as well.

Revenues and Infractions

The relationship between wins, revenues, and infractions (see below) was nonexistent. We had hypothesized that teams with greater revenues would have a higher risk of violating NCAA rules than teams with less revenues. Based on this data, it appears there is no relationship. This does not mean we should outright reject our hypothesis. After all, the database we looked at only contained data for major NCAA infractions. It is plausible that teams with greater revenues violate rules more often, but these violations may not be big enough to be included in the database we used. Further investigations are required to draw any conclusions on this matter.



Further Improvements

One metric that that would improve our visualization is the estimated profit per game. Some teams may have successful seasons, but they may have greater expenses through marketing and advertising, coaches' salaries, recruitment fees, etc.. The estimated profit per game would allow for a more standardized comparison between teams. For example, one team may bring in $100 million in revenues for the season, but their expenses could be $75 million, resulting in a profit of $2.08 million per game (assuming 12 games per season). Meanwhile, another team may have $75 million in revenues for the season, but their expenses are $45 million, resulting in a profit of $2.5 million per game. As our current visualization stands, it would appear the first team is more successful simply because their revenues are greater. Not only would this metric provide more accurate insight into which programs are the most profitable, it would also show whether football teams have become more profitable over time, and this could account for why schools are willing to consistently invest more and more money into football.

In addition, it would have been more insightful if we designed a visualization that allowed users to look specifically at how conferences changed over time. As it stands, once the user updates the year with the slider, the visualization resets to show all the conferences. The main reason we opted to reset the visualization is because conferences are not the same across all the years as some conferences are dissolved, received more teams, or lose teams.