

Project Proposal

CS 6630 - Visualization for Data Science

Basic Information

Project Title: Impact of Money in College Football

Nick Stephenson	u1015157	nijastleo15@gmail.com
Jeff Kelso	u1240886	jeff.kelso3@gmail.com
Andrew Potts	u1052633	andrewcarlospotts@gmail.com

Link to project: <https://github.com/nijastleo15/dataviscourse-pr-collegefootballviz>

Background & 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.

Project Objectives

We intend to make a visualization that enables exploration of our college football dataset. To do this, the visualization will allow 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 will also allow the user to select subsets of the datasets to see how different conferences compare to the whole.

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.

Once we had the fields populated for each year, we aggregated the spreadsheets into a master CSV to be used in the visualization.

Links to data sources:

<https://www.sports-reference.com/cfb/schools/>

<https://web3.ncaa.org/lstdbi/search?types=major&q=>

<https://ope.ed.gov/athletics/#/customdata/search>

etc.

<http://sports.usatoday.com/ncaa/salaries/>

Wins/Losses, Conferences data

NCAA infractions information

Expenses / Revenues / # of undergrads,

Head Coach salaries

Data Processing

The data should not require cleanup since the team created the CSV file from multiple sources. As far as derived data, we can show different ratios between our variables. For example, we could divide the revenue by the number of undergraduates (revenue per student) for each school to see if schools with more students make more money. We could also derive the cost per win for each year along with the conference averages so direct comparisons can be made between conferences in addition to teams. The format of the data will be a row for each year (2003 - 2016) for each school. This means that one school will have up to 14 rows of data. Some schools joined the highest level of college football (Division I-A / FBS) later, so they will have fewer rows. Over time, the conference that team plays in can change, and of course revenues, expenses, coach salaries, wins, etc. will be different each year. To read the data in our code, we will use a year slider to select the year, which will filter the data to use rows with that year.

Visualization Design

The plan is that our visualization will include a map of the United States that shows the location of the schools for a given year. The year will be chosen with a slider that goes from 2003 to 2016. The map will be further filtered by conference, private or public, and any other filters we decide to incorporate. The schools themselves will either be represented by a simple circle or the school's logo. When hovering over the circles/logos, a tooltip will show the schools name, their record that year as well as rest of the data for that school. The map will take up the left side of the screen initially. The right side of the screen will include a scatter plot comparing two variables in the X/Y axis and a third through circle size. These variables will be selected using dropdown menus below the scatter plot similar to one of the homework assignments. Variables include number of undergraduates, revenue, expenses, head coach salary, wins and any derived combinations we decide to use. A check box will apply the same filters to the

scatter plot as are used in the map. Clicking on a school in the map will highlight that school's data point in the scatter plot.

Additionally, when a school is selected, two more visualizations will appear below the map and scatter plot. On the left there will be a table showing a team's "profile". This will include a table of data across all years to reveal patterns for that school across time. On the right, there will be a line graph comparing chosen variables over time. The x-axis will be years (from 2003 to 2016). Dropdown menus below the graph will select two y-variables to be applied on both sides of the graph. For example, the left y-axis might show revenue while the one on the right shows the number of wins. This not only shows how these two variables change over time, but also provides a way to show how the two variables may interact with each other in addition to the scatter plot above.

Must-Have Features

In order for our project to enable data exploration, we must implement the US map with university locations, the modular X/Y Scatter, and school profile. The map needs to be interactive to allow for team selection. The scatter plot needs to be linked to the map to allow for highlighting and data filtering. The plot must have drop down menus below so that the axes can be changed by the user.

Optional Features

It would be neat to allow the user to select multiple conferences to highlight on the scatterplot. This could be implemented through multiple dropdown menus that change the color of the selected conference nodes in the scatterplot while the schools outside of the conferences of interest are left gray. A toggle switch that filters the scatterplot data to just the selected conference and re-scales the axes could also be useful since some users may only be interested in specific conferences.

The school profile could be modified to take in data for an entire conference and display the sum of each field for every year. Initially this seemed like it would simply be a matter of filtering the data, but since the conference members have changed over time it may not lead to an accurate representation of the data.

Project Schedule

- Proposal due **Sunday, October 28**
- Work to be done by **Friday, November 2**
 - Lat/Long for schools,
 - Setup HTML files with divs
 - Directory structure
 - Render map with tooltips
 - Implement year slider
- Project Milestone due **Friday, November 9**
 - Scatter plot renders and reacts to year slider
 - Complete skeleton of team profile table below map/scatter plot
 - Create a “release” in GitHub
- Work to be done by **Friday, November 16**
 - Render team profile table
 - Render team line plot
- Work to be done by **Friday, November 23**
 - If finished with the “musts” by November 16, implement the optional components
 - Complete two-minute screen-cast with narration.
- Final Project due **Friday, November 30**
 - Debug, wrap-up things