

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

Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

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.

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.

Related Work

Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

The greatest inspiration for our visualization comes from the gap plot visualization we implemented in homework assignment #4. With hundreds of teams that each have multiple datasets to visualize, we believed this gap plot layout would provide the best insight into our data. We will also implement a table similar to the one produced in homework assignment #5.

Questions

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

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:

- How have coaches' salaries changed over the years?
- Is there a relationship between the size of the school and how much money the football program receives?
- Does more money equate to more wins?

Data

Source, scraping method, cleanup, etc.

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

What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

TBD

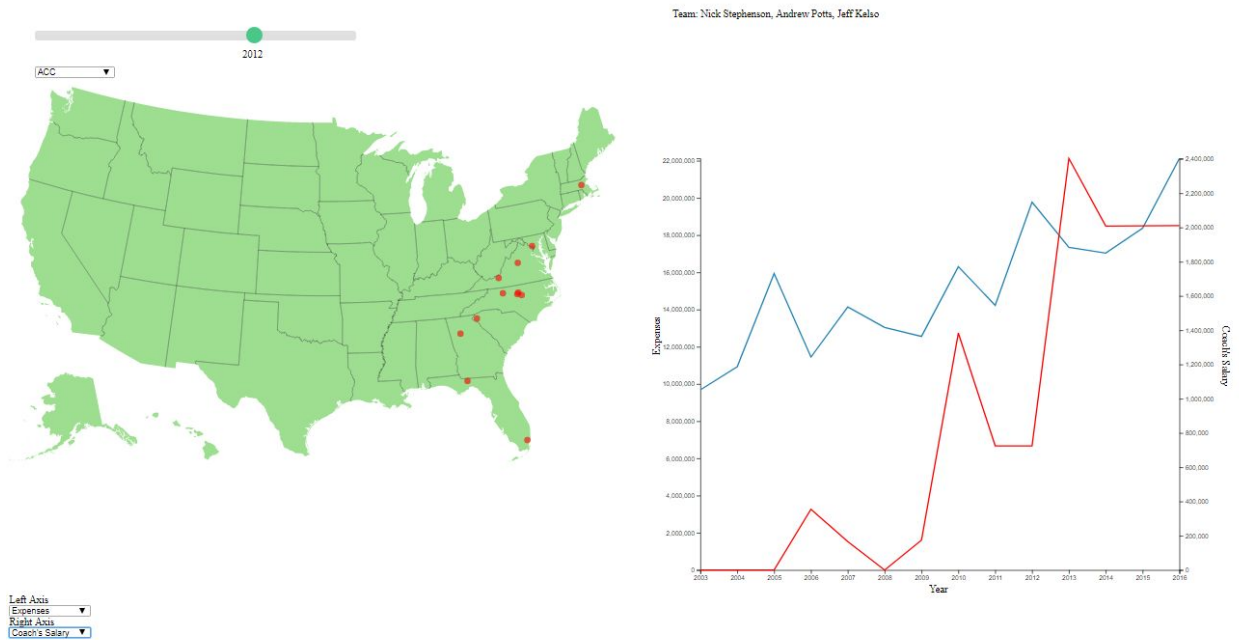
Design Evolution

What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

Initially the plan was a map including the locations of the schools, a scatter plot similar to HW 4, 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. So far we have not deviated from the proposal.

Implementation

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.



The screenshot above shows the current implementation of a dual-axis line chart. This code has not yet been merged with the master branch. The current project includes the map and scatter plot. Eventually, the line chart will be merged and the table will be implemented. Once all the major components are in place, the next step will be to link everything together.

Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

TBD