Grant Starnes

Dr. Liz Wanless

TESM-S501: Advanced Sports Analytics

12 October 2025

Sprint Challenge 2

What does it mean to create a 360 degree view of a problem? In class we reviewed several examples of data collection that mattered in sport (variables creating a 360 degree view for March Madness predictions and beyond). Discuss two of these examples in depth and choose another MIT SLOAN video in terms of 360 degree view achievement. Be sure to reference the specific data collection strategy. What additional data collection strategies may have also helped?

First off, what it means to create a 360-degree view of a problem is to begin by collecting and working with data or information from a variety of different perspectives in order to achieve a complete understanding of the problem, whatever the problem may be. Creating a 360-degree view of a problem allows the individual to get a glimpse of the problem as a whole from every possible angle, rather than just looking at the problem from one side or point of view. In order to get a 360-degree view, one can begin to gain the view from multiple perspectives by considering the opinions and ideas of stakeholders or various sources that are directly involved with the problem. Then, one can analyze any kind of information that is relevant to the problem, whether it's internal, external, or historical data/information. Combining the various perspectives from all of the stakeholders involved with the information present, one can gain a far greater understanding to make better decisions moving forward, as well as pinpoint new opportunities that may not have been entertained before.

A great example of creating a 360-degree view of a problem is attempting to predict March Madness tournament play. Like stated prior, to predict March Madness tournament play, we need to consider a variety of perspectives and different data to understand the situation as a whole, rather than from one point of view. For example, we can first look internally at March Madness teams and look at

their team performance, whether that's their offensive statistics, defensive statistics, their record, or strength of schedule. Diving further internally, we can even analyze statistics at the player level, such as their points, rebounds, assists, and efficiency. Externally, we can look at things that are out of the team's control, such as the opponent and their statistics, where the game will be played as it's normally in a neutral site for the tournament, fan engagement, sports betting odds, and the opinions of sport analysts. There is also the historical aspect of the game, as many like to say "history repeats itself." We can look at the team's history in the tournament overall, such as coaching in previous years, how the team plays at neutral sites, and how the team does while on the road. In this context, we can also incorporate outside metrics, in this case, KenPom ranking, which looks at offensive and defensive efficiency. Putting all of these together, one can gain a 360-degree view of each team playing in March Madness and understand from a well-rounded perspective rather than a fixed perspective and looking at one metric, such as seeding, which we know isn't the greatest predictor as every year there are upsets.

Next, we can look at a good example of creating a 360-degree view of a problem by surveying stakeholders in the SEC. When surveying stakeholders, gaining a 360-degree view could mean a variety of things. First, we want to get information about the stakeholders' background, such as which team they support, home games attended, gender, age, approximation to the stadium (travel time), and where the stakeholder sits in the given stadium. Next, we want to determine how satisfied the stakeholders are in various aspects regarding the team, like first impressions (parking, accessibility, cleanliness, efficiency, comfort, and security). Then, we can break things down into categories like food services, restrooms, crowd, band, and cheer participation, sighs and sounds at the stadium, and connectivity regarding wifi/mobile device service providers. With all of these in mind, gaining a 360-degree view of stakeholders means gaining a well-rounded insight about their expectations, needs, and experiences from a point of view that can vary at a high rate, as anyone can fill out the survey despite their background. In theory, this can help round out and limit biases and skewed data and ultimately provide the team/organization with data that is accurate and useful and can help them prescribe and predict future outcomes.

The MIT SLOAN video I elected to watch and look at how they use the 360-degree view of a problem was the video titled "SSAC25: How AI and This One Metric Can Reshape NFL Roster Decisions," presented by Joel Shapiro. The metric Shapiro uses is called "PCW" which is Percent Cash Wasted, and is the percentage of a team's payroll that is spent on players that are injured in a given year. Shapiro states that he uses data to find stories that help unlock performance and fulfill potential, very similar to what we've talked about in class recently regarding Cole Nussbaumer Knafflic's effective storytelling guidelines. He then goes on to state that for the data to be important or useful, there needs to be an action others can take after looking at the data, which could be considered the final step in the 360-degree view of a problem as the main goal is to help others and prescribe and predict future outcomes. His data collection is for the duration of the 2012-2022 NFL seasons, and hones in on the percent cash wasted as mentioned before. His data collection strategy was a company called Probility AI who simply provided the data to him for him to use and analyze. This Probility AI dataset allows him to mine through the data that was very unique to him. The next metric he uses is win percentage, and there is nearly a 1 to 1 ratio between win percentage and percent cash wasted. So, for every 1% cash wasted, the win percentage decreases by 1%. As for additional data collection strategies that may have also helped Shapiro, one of these could be looking at the player availability and the depth chart. Shapiro could've looked closer at the number of games missed by the player, recovery time, and who was replacing the injured player via the depth chart. This data collection strategy can look further internally to determine how much a team is actually affected by injuries. If the team has a talented, deep roster, they may be able to sustain a higher PCW compared to other teams. W compared to other teams, there's a bit of variability here. Another data collection strategy that may have benefitted Shapiro would be looking at the injury type and severity of the injury. Injuries can be and are categorized very well in sport today, and we know rather quickly how long a player is out for. In his current strategy, he treats all injuries the same, but they simply aren't. It also matters which player gets injured, for example, the starting quarterback, or a player blocking on the front line on kickoff return.

Generate your own metric (equation) to rank players or teams in a sport that you choose over a time period that you choose. Choose 15 players or teams and rank them according to the metric. Display the ranked teams and players in a compelling visual. Be sure to address the following:

a. **Utilize positive and negative performance metrics to create the overall metric. Display your metric and break down each of its components.**
b. **Explain the logic and reasoning behind the equation with a narrative explaining your approach.**
c. **Discuss the ranked players or teams - did your metric rank in the same way it should or the way that you thought it would? What strengths and weaknesses of your metric were revealed as a result of your analysis? Were there any surprises?**
d. **Show how your ranking correlates with various performance outcome measures (e.g., team wins). Choose at least 5 performance outcome measures, maybe even some within your equation. What does this tell you about your metric?**

Metric: College Football Performance Index

$$CFBPI = (0.4 * OS) + (0.3 * DS) + (0.2 * SOS) + (0.1 * TM)$$

This is my college football performance index metric (CFBPI). The various performance metrics are broken down as follows:

- OS → Offensive Strength (40%)
    - Points scored by the team per game. Offensive Strength is given a weight of 40% because this is the main source of a team scoring points, which leads to winning games, and is essentially the most important factor.
- DS → Defensive Strength (30%)
    - Points scored by opponents per game. Defensive Strength is given a weight of 30% because defense is what prevents an opposing team from scoring, and could even score themselves, but not as likely as offensive strength.

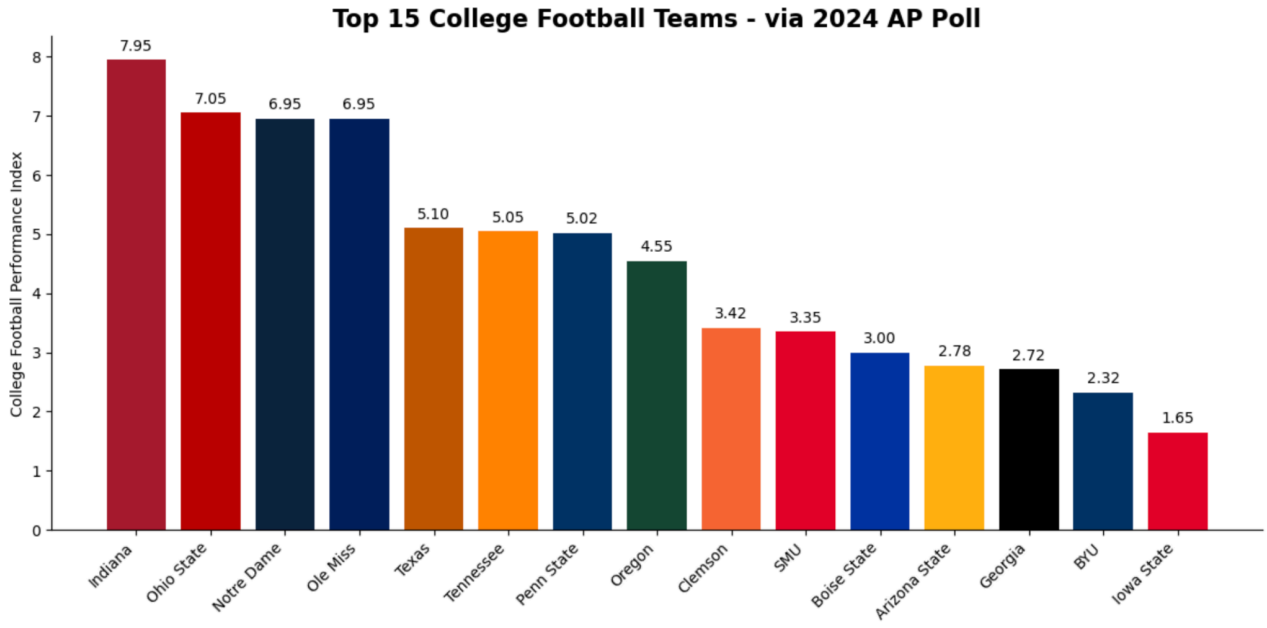
- ➢ SOS → Strength of Schedule (20%)
    - Strength of opposing teams on the team's schedule. Strength of Schedule is given a weight of 20% because it's an important factor in determining how good a team actually is. Some teams play more competitive, tougher schedules, whereas others play less competitive, easier schedules. This is a better performance metric compared to simply using wins and losses, and rewards the teams who play those tougher schedules.
- ➢ TM → Turnover Margin (10%)
    - Number of turnovers per game. Turnover Margin is given a weight of 10% because although important and can change the course of games, turnovers are less impactful than the strength of offense and defense, the main phases of scoring. Turnovers can also vary from game to game more than any other performance metric, so a lower weight is necessary.

Explanation of the College Football Performance Index

The College Football Performance Index (CFBPI) is a metric made up of multiple weighted factors based on their importance of game outcomes and determines the strength of a team overall. The CFBPI consists of offensive strength, defensive strength, strength of schedule, and turnover margin. As for offensive strength, it holds the most weight because offense is the main driver in scoring points for a team, leading to games being won. Defensive strength holds a little less weight than offensive due to decrease in point scoring. While defenses can still score, it is less likely, but they are the main entity standing in the way of an offense scoring more points. Next, strength of schedule is tied into this metric mainly to evaluate the quality of opponents a team faces. Some teams will play a far tougher schedule, for example, SEC and BIG 10 teams, whereas some teams will play a bit of a weaker schedule, which takes away just looking at wins and losses and rather looks at what teams a team has played. Lastly, turnover margin is still important, but holds the least weight in this overall metric. Turnovers are difficult to predict and vary more than any other performance metric within this overall metric. With the help of weights, it allows the

CFBPI to be more balanced and take into account different performance metrics and give them a greater influence than others based on their significance throughout each game.

CFBPI Ranking: Expectations, Strengths, and Weaknesses



As you can see, this is the bar chart that visualizes the CFBPI metric statistics for what were the top 15 teams at the close of the 2024 college football season. Regarding the CFBPI metric and the ranking of the teams, I believe the metric ranks them as expected and the way I thought it would. Based on the weights for different performance metrics and the data from the 2024 season, this ranking is in order. For example, the Indiana Hoosiers scored a total of 537 points, whereas teams scored a total of 203. They significantly outscored opponents all season, and the highest weighted performance metric is offensive strength, skyrocketing them to the first rank. Following Indiana are the 2024 champions, Ohio State, and 2024 runner up, Notre Dame, who also completed the feat Indiana did, but not scoring as many points per game offensively and defensively, also taking into account strength of schedule and turnover margin. As for the strengths of the metric, it closely aligns with the actual outcome of the college football playoff in 2024, with Ohio State winning and Notre Dame coming up just short. The remaining teams also made deeper runs in the playoffs, such as Penn State and Texas, and are also very close to the teams who made it to the

national championship. The performance metrics within the CFBPI seem to pretty accurately represent the actual outcome of the previous season, but not completely. Moving forward to weaknesses, the CFBPI metric could be rather misleading to those analyzing the visualization. They may think Indiana had an amazing season and even went on to win the national championship, but that isn't the case. Georgia, with a 2.72 CFBPI score actually made a far deeper run in the playoffs than Indiana did, even though they have a 7.95 CFBPI score. This goes to show the weaknesses that the statistics and visualization can mislead observers and the overall statistics from the season don't necessarily align with what the outcome was once playoffs came around. Lastly, the only surprise to me based on the visualization is Georgia's CFBPI, as they had a great season but are down near the end of the top 15 teams, and I expected them to be much higher, roughly grouped around Texas and Tennessee.

How the CFBPI Ranking Correlates with other Performance Outcome Measures

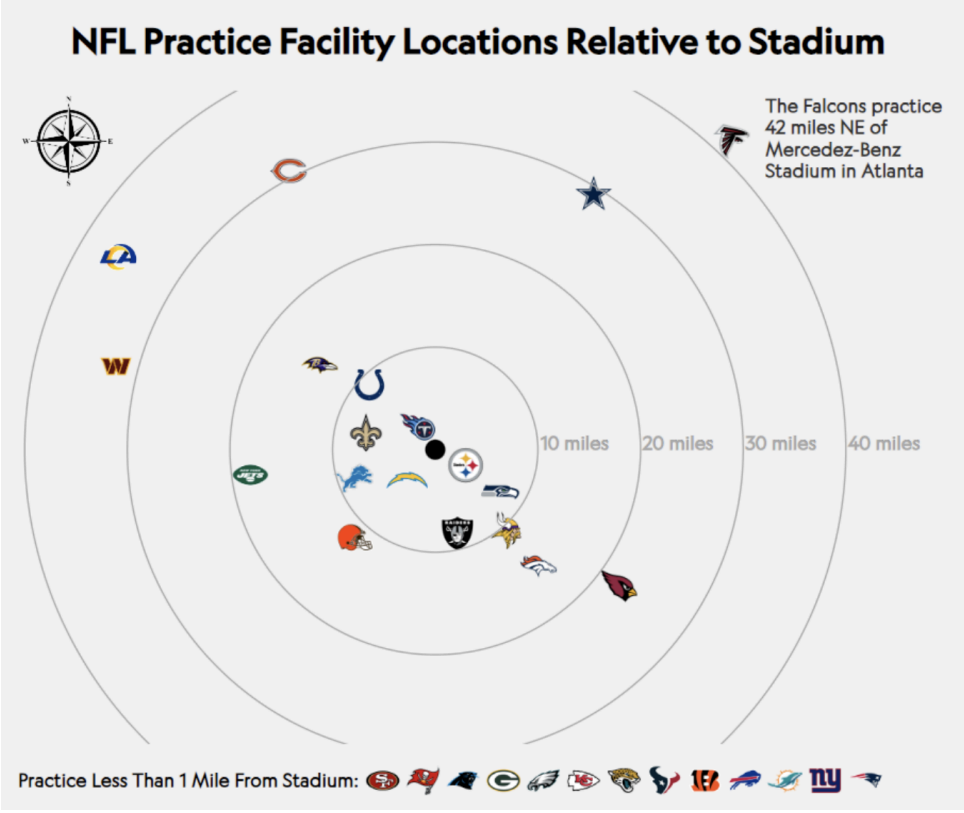
	Points For	**Points Against**	**Turnover Margin**	**Wins**	**Losses**
Indiana	537	203	+1.0	11	2
Ohio State	571	206	+0.3	14	2
Notre Dame	578	248	+1.1	14	2
Ole Miss	502	187	+0.8	10	3
Texas	528	245	+0.3	13	3
Tennessee	464	209	+0.5	10	3
Penn State	530	264	+0.7	13	3
Oregon	488	272	+0.4	13	1
Clemson	486	328	+1.2	10	4
SMU	511	309	-0.3	11	3
Boise State	522	316	+0.4	12	2

Arizona State	461	316	+1.0	11	3
Georgia	441	288	-0.2	11	3
BYU	405	255	+0.6	11	2
Iowa State	435	321	+0.6	11	3

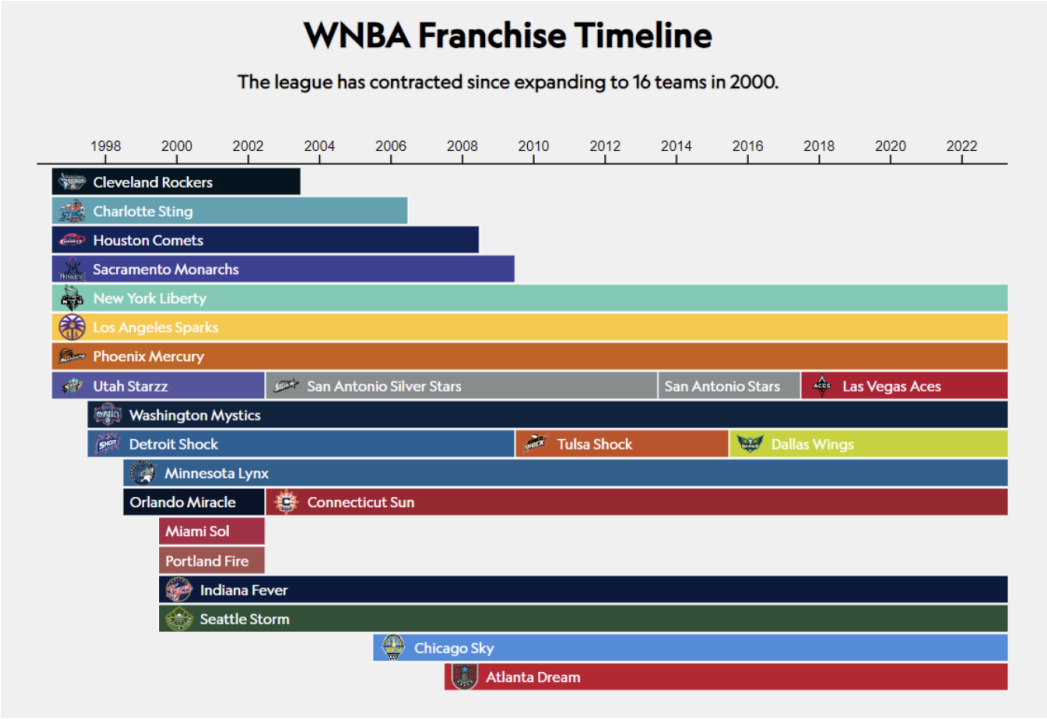
These are the raw statistics for each team in the CFBPI metric, which includes Points For, Points Against, Turnover Margin, Wins, and Losses. I believe the correlation between the CFBPI metric and these performance outcome measures are strong as just by doing an eye test looking at this data, one can see how the difference in Points For and Points Against can elevate a team in these rankings. For example, the top 5 teams in the rankings scored a large number of points, but didn't give up even close to as many points. Comparing that to the bottom 5 teams in the rankings, it's close, but they didn't score as many points and also allowed teams to score on them quite a bit more. Since these two performance outcome measures are weighted, points for and points against hold a lot of value and the metric seems to properly display this. As for turnover margin, it isn't weighted as much as points for and against, but it still is significant in ranking teams. A higher turnover margin indicates that the team gained possession of the ball through interceptions or fumbles more often than given the ball away, whereas a lower turnover margin begins to indicate that the team gave the ball away more often than they gained possession. What this tells me about the metric is that the performance metrics work extremely close together in determining the ranking of teams, and that weights absolutely do matter. This also tells me that the metric isn't too far off of the actual data being used, and that the metric is useful and more accurate than not. One more thing this tells me about the metric that could be a flaw is that the total number of games played varies a bit. Rather than use data through the end of the season as a whole, the data could be tailored to only go through the end of the regular season so each team has played the same number of games, providing a more accurate visualization.

What does it mean to tell a story with sport data? Choose two sport data visuals and discuss how each either meets or does not meet the guidelines set forth by Cole Nussbaumer Knafflic outlined in class.

What it means to tell a story with sport data is turning raw data or statistics on a page into understandable information that is meaningful to help the sport audience understand and buy-in to what the data is revealing. Telling a story with sport data helps get across to the sport audience what the purpose of it is, why it matters in the first place, and what the data provides them with. In the "Storytelling with Data" video featuring Cole Nussbaumer Knafflic, she uses an analogy of components of an actual story (plot, twists, ending) and applies it to telling a story with data. For the plot, we want to determine what context is essential for the sports audience and what their characteristics are. For the twists, we want to figure out what is interesting about the data and what it shows. Lastly, for the ending, we need to find out what we want the audience to do afterward with this new information.



I believe this sport data visual meets Cole Nussbaumer Knafflic's first step in the guidelines to effective storytelling, which is understanding the context, and we can tell by the bold title that it's the NFL practice facility locations and how far it is from the stadium. Then, I would say this is an appropriate visual that is easy to read and understand and tells you what you need to know quickly. As for any clutter, the visual is already very stripped down and simplified to where there almost isn't any clutter and the data is labeled and leverages consistent color through the logos of each NFL team. The visual first draws my attention to the large bold title, which is good because it informs the audience on what this visual is all about. Lastly, I think it absolutely tells a story as it provides the viewer with context in the title of what it's about, the viewer can see what's interesting about the data, and the viewer can much more easily retain the information from this visual afterward compared to having read each team name and the distance away in a standard text form where most of the information wouldn't be retained.

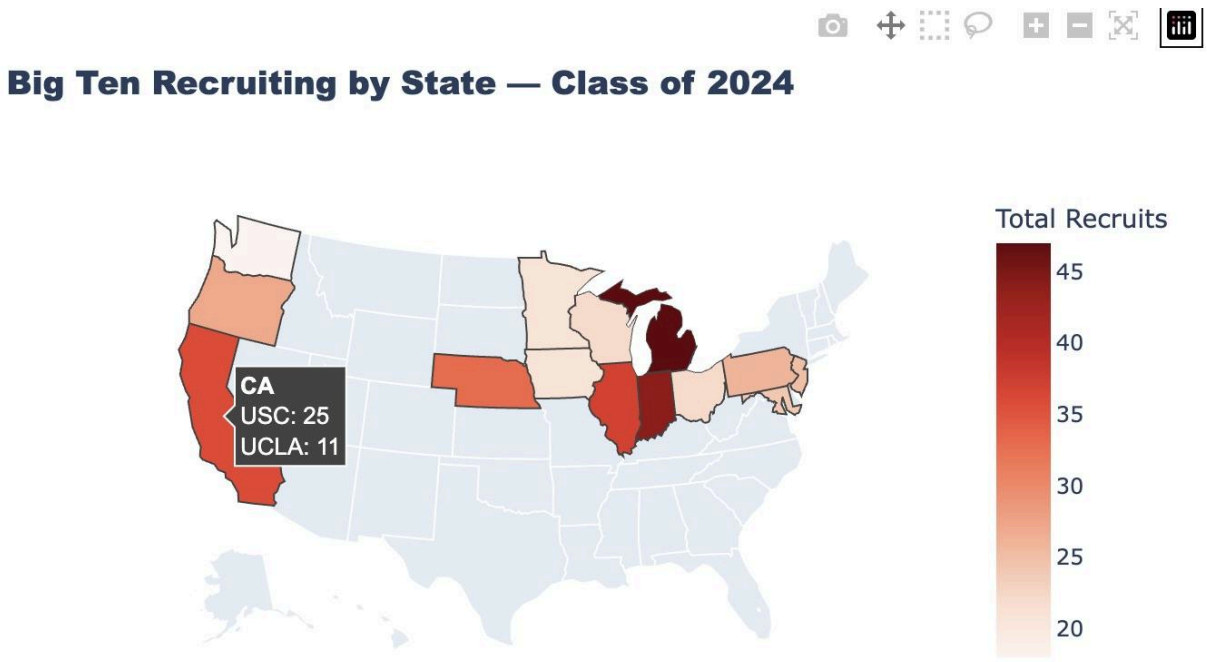


I believe this sport data visual meets Cole Nussbaumer Knafflic's first step in the guidelines to effective storytelling, understanding the context, because once again, the large title at the top is straight to the point and very noticeable when first looking at the visual. Next, I'd say this is an appropriate visual for the most

part. The structure and layout is great, but it's a bit difficult to accurately tell when a franchise began/ended. I would maybe suggest adding the year the franchise was founded to the left of each bar if possible, but I understand why they didn't as it would be redundant to put the year timeline at the top to put another year right next to each bar. I would say there's hardly any clutter at all and the visual is very clean and simple. The colors are distinguishable, each one is labeled directly, there aren't any messy axes, no gridlines or borders, and no data markers. It focuses my attention first on the title then I'm drawn to all of the colors of each bar below. Lastly, it very simply tells the story and is straightforward with the timeline at the top and each bar spanning within the width of the timeline to show the duration of each franchise.

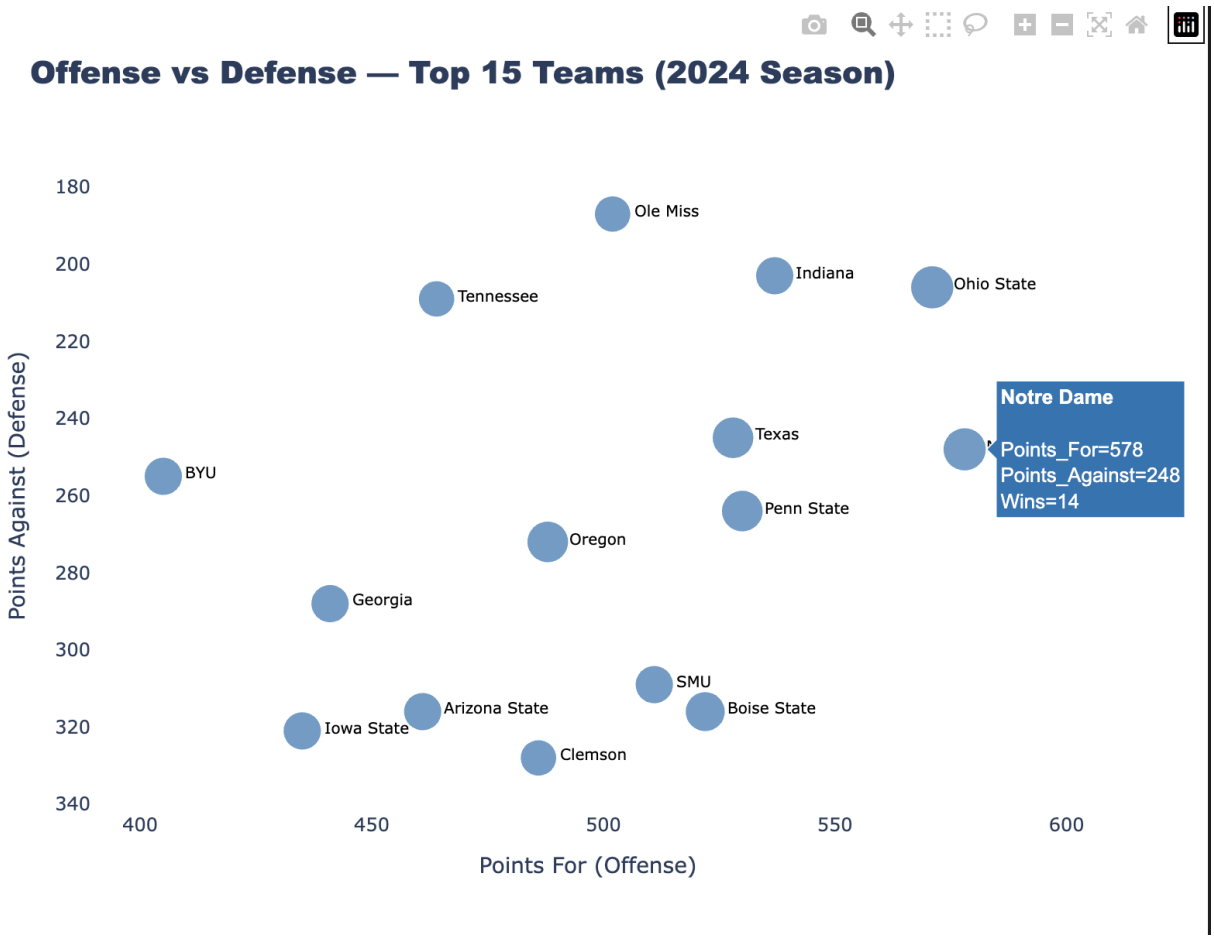
Utilize two different visuals crafted in Tableau (OR R/Python) to tell a story (or stories) embedded within the data you have available via the Big Data Bowl competition, past and publicly available NFL play-by-play data, OR data from a sport of your choice. Each visual should adhere to the Cole Nussbaumer Knafflic data storytelling principles outlined in class. Be sure to address the following:

a. **Each visual must be a compilation of several variables. Each visual must be completely distinct from the other.**
b. **Include a short narrative explaining your visuals and what they mean for your final project outcome.**



This visual simply presents a choropleth map of the United States, but focuses on the states that contain a Big Ten school(s) and the total number of recruits that state has brought in at the close of the 2024 college football season. At first, it seemed fine to simply sum all recruits within a state even if there were many schools in the same state, but things could get a bit confusing there if that ended up being the final outcome. I decided it was far more necessary to group by the state overall but display the recruits for each team when a user hovers over the state, for example, California, which displays the two Big Ten teams in that state and their recruit count, USC with 25 and UCLA with 11. The same goes for Illinois which houses Illinois and Northwestern, Michigan who houses Michigan and Michigan State, and Indiana who houses Indiana and Purdue. Also, without having to interact with the choropleth map, an observer can simply look at the map and the legend to determine what the colors mean, darker red for a greater number of recruits, and a lighter red for the lesser number of recruits. As for my final project outcome, a choropleth map like this could be used for a number of metrics, whether it's for revenue generated in each state by their housed NFL teams, the majority of fanship in each and every state even if they don't house

an NFL team, or even elevating this visual to take data on the NFL Draft and where the drafted players played in college, rather than where they were recruited since not all college players make the step up into the NFL.



This is a scatterplot that shows the offensive and defensive performance of the top 15 college football teams at the close of the 2024 season, with the x-axis representing the offensive performance, points for, and the y-axis representing the defensive performance, points against, which is also structured in a reverse order, so less points against at the top since that is the positive side of the metric, and more points against at the bottom so the scatterplot displays and behaves in a way observers are used to when viewing a scatterplot. There is a slight difference in the plot point sizes based on number of wins, larger being more wins and smaller being less wins. Each individual top 15 team has also been labeled so it's easy to read and understand which point is which. This scatterplot as a whole meets what was expected, but it does

have its outliers based on the actual outcome. For example, Indiana did a great amount of scoring and the defense played great, but their playoff run ended far earlier than those grouped around it like Notre Dame and Ohio State who played in the national championship. Another outlier is Georgia as they gave up quite a few points on defense and didn't score as many on offense as the higher end of the plot, but they still made it further into the playoffs. All in all, this scatterplot easily distinguishes the teams based on their offensive and defensive performance for the season overall, and closely reflects the actual outcome of the playoffs at the end of the season. For the final project, this scatterplot helps support that the overall success of a team is not only based on offense and defense, but takes into account other factors, like strength of schedule and turnover margin, like the CFBPI metric above uses, which is likely more accurate applying weights to each performance metric as well. This applies to the final project in the sense that there are more variables than we think there are in determining a team's success, not just things like points and wins and losses. We must really evaluate the data and the sport and bring in factors that can largely affect game/season outcomes, not just focus on easily observable data like wins, losses, and points, since those don't completely define teams and accurately represent the team and what actually happened.