Visualizing NFL Draft Data

A Data Visualization Project

Project Team:

Dante Amicarella: deamicarella@wpi.edu

Jack Lafond: jwlafond@wpi.edu

Mason Perham: mfperham@wpi.edu

Project Advisors

Prof. Lane Harrison, *Computer Science Department*

Table of Contents

Overview and Motivation.....2

Related Work.....3

Questions.....7

Data.....8

Exploratory Data Analysis.....9

Design Evolution.....10

Implementation.....14

Evaluation.....19

Works Cited.....21

Overview and Motivation

The NFL is a multi-billion dollar industry whose fans, media, and investors continue to grow each year both in the US and globally. Each off-season much of the excitement and anticipation continues as fans and analysts await the next NFL draft. For many teams and fanbases the draft can either mean years of glory to come, something that the Chiefs and Patrick Mahomes know all too well. However, for others this could mean wasted picks and years of sub-par seasons.

As of now there is no formal way of viewing NFL draft data for analysts or NFL fanatics around the world. The data from these drafts is greatly interconnected with much of the overall American Football data landscape, and as a result can be complicated to properly view. Each player picked represents an athlete with physical scores and measurements, a college that helped develop that athlete, a position that represents that athlete's particular set of skills, a round and pick number that represents the assessed value of that player by the team that picked them, and a possible future of success and accolades for that team.

In this project we will explore various visualizations and attempt to encode and display draft data in a way that makes extracting information from this dataset easier for fans and draft analysts. We have gathered a large sample of NFL draft data that spans from 2010-2020. In this data set we include all relevant variables for the draft like player position, college, draft pick, and the team that drafted them. We have also gathered career accolades for each of these players like number of MVP awards, Super Bowl wins, and other NFL honors. Due to the complex nature of the dataset we will utilize different types of visualizations like network and flow visualizations, such as Sankey and Hive plots, as well as dimension comparison plots like spider charts and parallel coordinate plots.

Through this project we hope to establish efficient ways of visualizing the NFL draft in an expository manner so that we can improve the current state of NFL data visualization for fans, analysts, and media.

Related Work

This project was an add-on to previous work from our very own Jack Lafond in collaboration with Aaron Brady during MA 4804. This project utilized the same data except they were trying to answer the question: Do college pipelines exist to the NFL? To answer this question, Brady and Lafond used a Sankey diagram to demonstrate the flow of players from the draft to their associated college conference and finally to the division they were drafted to. In addition, they created a linked view with a stacked bar chart to further analyze the teams in a division selected.

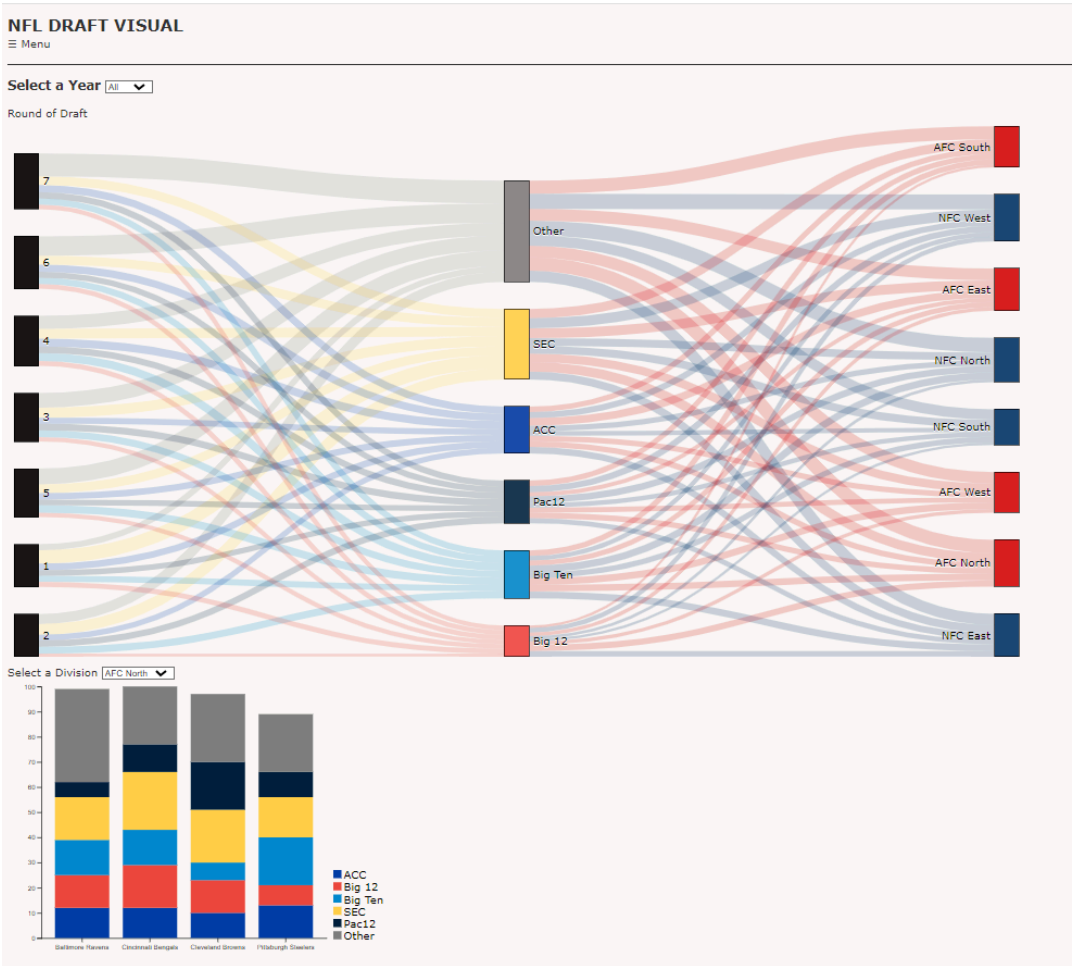

Figure 1: Brady and Lafond linked views diagram

Inspired by this draft data and project, the team went looking for ways to improve the way the data was used. In 2016, the *Washington Post* published an article diving into the

visualization of NFL draft data and decomposing the draft performance from best to worst teams (*Washington Post*, 2016). This article broke down several visualizations and composed a story as to how a team has performed, in terms of drafting, over the past 20 years. In the vein of storytelling, we sought out not only to create attractive and useful visualizations, but to tell our thought process through words on a screen. As seen first in *The New York Times "Snow Fall: The Avalanche at Tunnel Creek",* scrollytelling has taken the visualization community by storm due to its ability to tell a compelling story by breaking elements apart (Morth et. al, 2022).

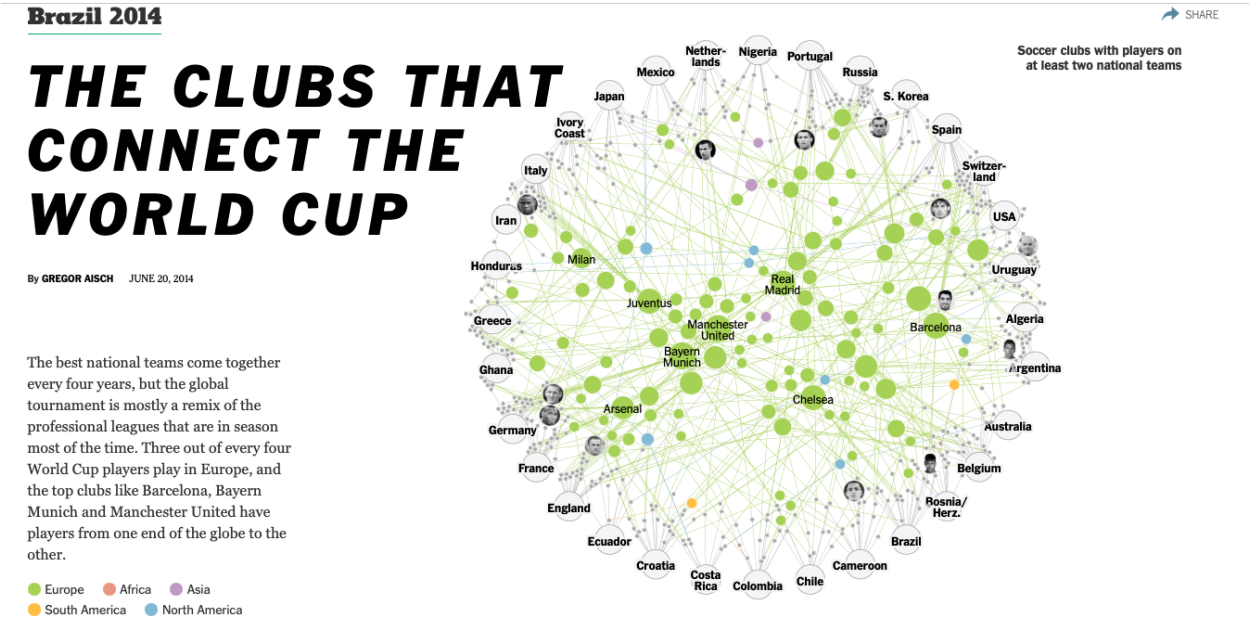

Figure 2: A sports scrollytelling example in The New York Time's (Aisch, 2014)

As seen in Figure 2, captivating visualizations are accompanied by descriptive text that not only explains the visualization but progresses the story.

The ability to interact with the visualizations within the story created a new appeal for the team and led to exploring the power of interaction within visualizations. This was explored through making interaction visible as discussed by Fu and Statsko (2022). By hiding the ability to interact with visualizations, the user may skip over or not enable the interaction available within a diagram. Through the usage of visible queues, a user is able to teach themselves how the information is supposed to be digested.

Additionally, we explored different graphs and how they were able to be adapted to our data. In bioinformatics, sparse data is very common, so a visual network is common within the field. Krzywinski et. al recommended the usage of a hive plot–a three axis model that breaks

apart the network model into a digestible visualization (2011). Hive plots as in at the bottom of Figure 3, are made up of axes, nodes, and connections between nodes to model their relationship. Nevertheless, these plots allowed for a quantitative approach compared to their network graph ancestors. Other visualizations explored included radar charts, also known as spider charts, and parallel coordinate charts. As seen, utilizing parallel coordinates would be a straightforward approach to visualizing multidimensional data (Janetzko et. al, 2016). On the other hand, radar charts had to be deployed carefully as they are frequently used improperly and are also misinterpreted by the user (Casals et. al, 2022).

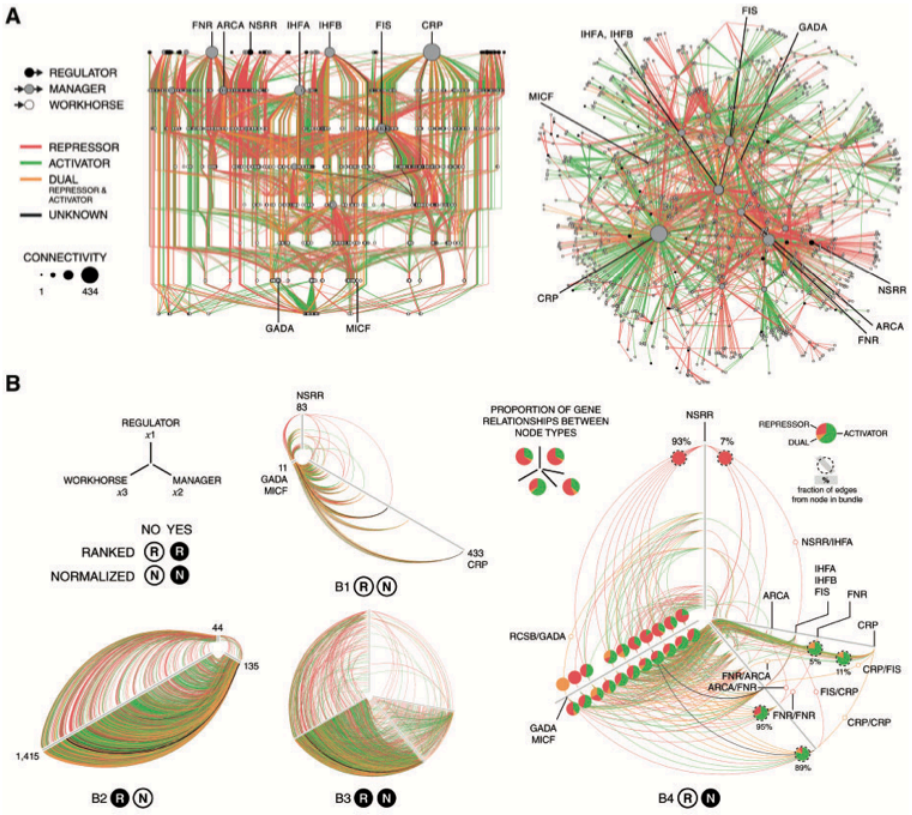

Figure 3: Hive plot breakdown by Krzywinski et. al.

Questions

When creating our project, our visualizations would change along with the questions they were trying to answer. During this process, our group questioned and re-questioned the thesis of our project and the purpose behind the visualizations. At the beginning of out project our thesis questions started as:

1. How do we want to present our visualizations?
2. Do certain NFL teams or colleges produce more successful NFL players?
3. Do players picked in earlier rounds always succeed more than those selected later?

After revision of our work and many design changes our questions changed to:

1. How can the content of one visualization transition to the content of the following visualization?
2. What elements in the graphs distract the user from what we are trying to show?
3. Which college conferences create the most NFL-ready players?
4. Are there ways to display the flow of college players to the NFL more neatly?
5. How can we incorporate player accolades into a network graph?

Data

The data used for this project was collected using a web scraper. The tool accessed Wikipedia data for the NFL draft classes between 2010 and 2020. It collected information on each player, including round drafted, pick number, NFL team drafted to, player position, college drafted from, college conference, NFL division, and individual awards won.

The scrapping method we used made HTTP requests to Wikipedia for each of the Draft years. The script uses the BeautifulSoup library to parse HTML content and extract data from relevant tables that list draft picks. It navigates through the web pages, identifies the correct table, and then iterates over the rows of these tables to extract the data. For player-specific achievements, the script further scrapes individual player pages linked from the draft tables, focusing on sections that list career highlights and awards.

The data cleanup consisted of stripping unnecessary characters and standardizing the data. A function called ‘makeID’ was used to create an unique individual player ID by combining the draft year and pick number for each player. The data was saved in a csv file for each draft year and then compiled into a complete dataset.

Exploratory Data Analysis

Our project was a continuation of a past project, so we already had a few visualizations to look at when we began. Initially, we had the Sankey diagram shown above in *Figure 1*. This diagram connects the players drafted in each round to their college conference, which is connected to the NFL division that the players ended up being drafted into. We used this design as a starting point for our hive plot. We saw that each of the three axes could be connected to the other two, so we researched ways to implement our findings, which led us to our hive plot design. Additionally, we created bar charts to display the data we collected. We used the bar charts to display the counts of player awards per conference and position. They helped us notice the prominence of the SEC in creating NFL ready players. These counts were used to create the links in the arc plots that were created later in the project. We then decided to separate the counts by year. From these yearly bar charts we could see that the SEC was consistently dominant across the years from 2010 to 2020. These findings gave us our initial story that the SEC was the best college conference for producing NFL ready players. From here we focused on finding plots that would best portray that story.

Design Evolution

Throughout our project journey, we leveraged the Design Activity Framework (McKenna et al., 2014) to guide our process from understanding to ideation, creation, and presentation of our final product. This framework enabled us to break down each component of our project, ensuring its alignment with our objectives. In the continuous cycle of design, our group met regularly to deliberate on our visualizations, their conceptual underpinnings, adaptability, and the intended message they conveyed.

As discussed, our project was a branch-off from an already existing project. Our initial project idea was to adapt the Sankey diagram and improve the overall quality of the visualization. We envisioned rotating the diagram so that it would flow within the constraints of the screen size in addition to replacing the stacked bar charts with spider charts to add a visual flare to the page. As seen in Figure 4, these ideas were explored using hand-drawings to better understand the task that we were trying to accomplish.

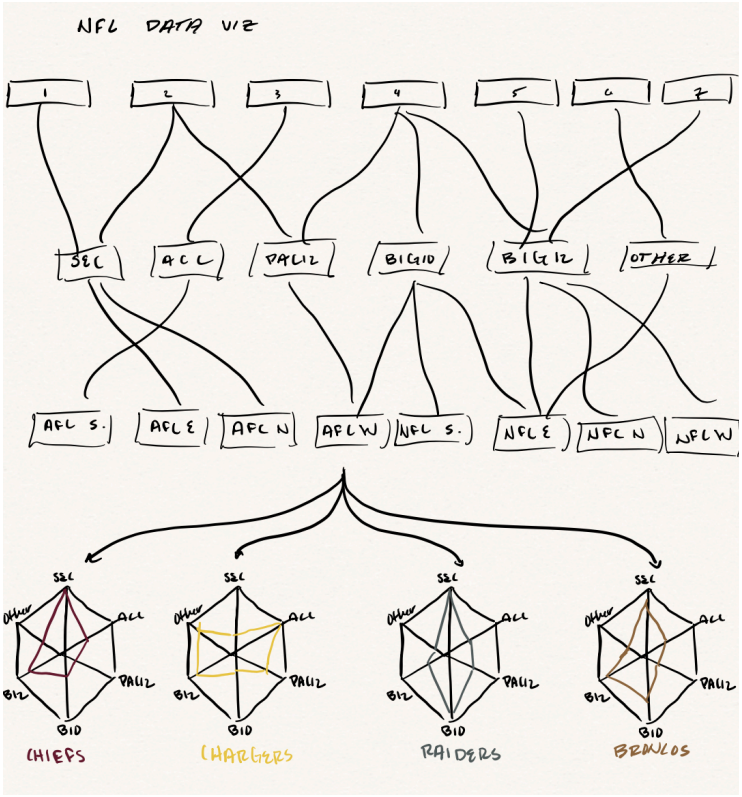

Figure 4: Hand drawing of design visualization idea

While making the adjustments to the Sankey diagram, we also started creating other visual elements. It was at this time that we decided as a group that we wanted to tell a story as a

part of the project. During the exploration of other visualization techniques we created a Hive plot, parallel coordinate plots, arc charts, small multiple bar charts, line plots and more to better guide our story.

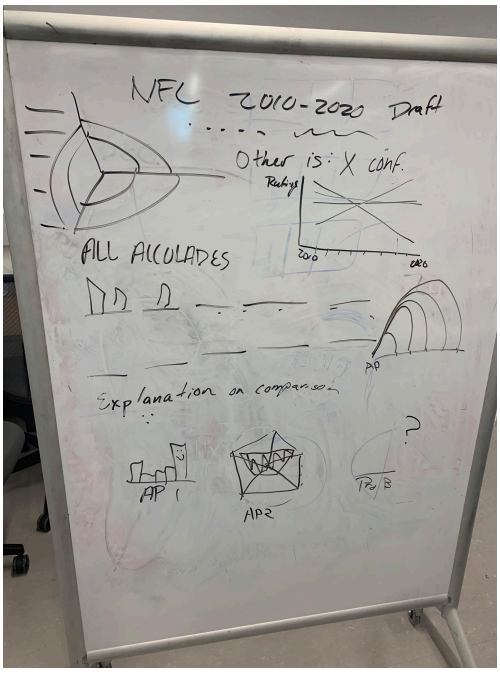

Figure 5: White board result of design evolution

During this process, among all the graphs, the Hive plot, as seen in Figure 6, emerged as the most promising due to its uniqueness and captivating visual appeal. The mapping of links to nodes added an attractive staple visualization that was unique to our project, setting it apart from conventional approaches. This subset of a network graph proved to be the anchor of our project. Additionally, the incorporation of context pop-ups enriched user understanding by providing clear insights into the nature of each node on an axis, whether it represented a player, college conference, or team. Moreover, empowering users to manipulate the chart by valuing accolades according to their preferences enhanced interactivity. This feature allowed users to experiment with the data, exploring which players excelled in specific accolade categories. As a result, the Hive plot not only stood out for its aesthetic appeal but also for its dynamic and user-centric functionality, contributing significantly to the overall success of our project.

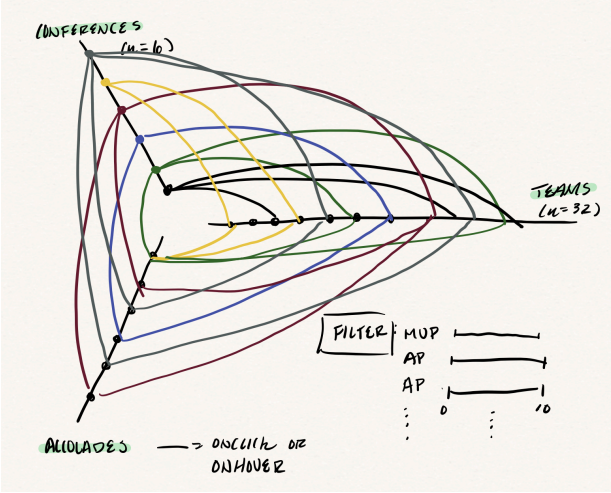

Figure 6: Hand drawing of Hive plot

After finalizing the design and functionality of all graphs using d3.js, we delved into the discussion on color schemes for each. Consistency in coloring, aligning with categorical categories, was deemed essential. For instance, employing dark blue consistently across graphs signified the Pac-12, ensuring viewer comprehension. Furthermore, we prioritized user perception of the visualizations. Implementing equal axes across small multiples and integrating on-hover events allowed users to focus on specific elements within the graphs. These considerations were also extended to the Hierarchical Bubble Chart, enhancing its user experience.

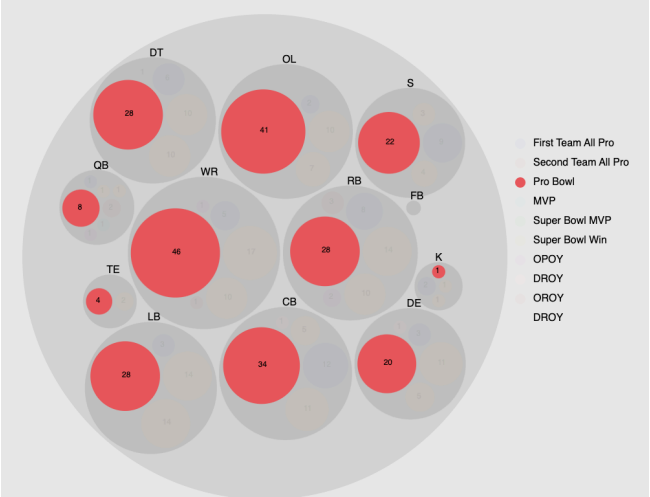

Figure 7: Hierarchical bubble chart with active on-hover event

Completing our project, we incorporated a final element aimed at enhancing user understanding of the Hive plot. We developed a dedicated page that served as an interactive

tutorial, guiding users through the intricacies of the Hive plot. Employing a scrollytelling format, users embarked on a journey of discovery, gradually unraveling the nuances of each component of the graph as they scrolled through the page. Using group meetings, we were able to break down what the user should be seeing as they scroll in Figure 8. This immersive approach not only engaged users but also provided them with a comprehensive understanding of how the Hive plot functions and how to interpret its visual cues effectively. By seamlessly integrating education with exploration, our scrollytelling page transformed learning into an interactive and enriching experience, further enhancing the overall utility and appeal of our project.

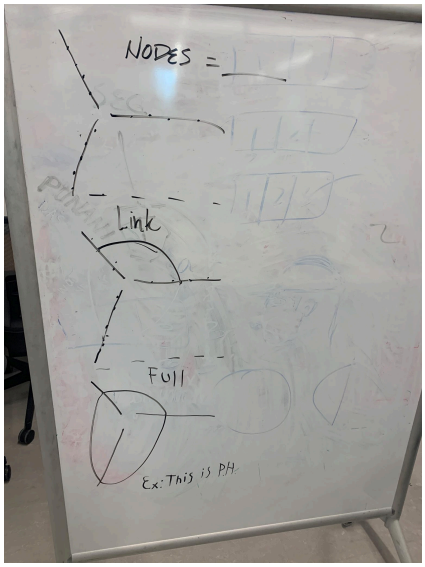

Figure 8: Hand drawing of scrollytelling breakdown

After finalizing our primary visualizations for the main pages, we curated a separate section titled "Other Vis" to house the remaining visualizations that didn't make it to the main interface. This thoughtful inclusion ensured that users could explore and appreciate the breadth of our analysis, even beyond the focal point of the project.

By providing access to these additional visualizations, we offered users a comprehensive view of our data exploration process, showcasing the range of insights and perspectives considered throughout our project journey. This dedicated space not only served as a repository for unused visualizations but also as a testament to our commitment to transparency and thoroughness in presenting our findings.

Implementation

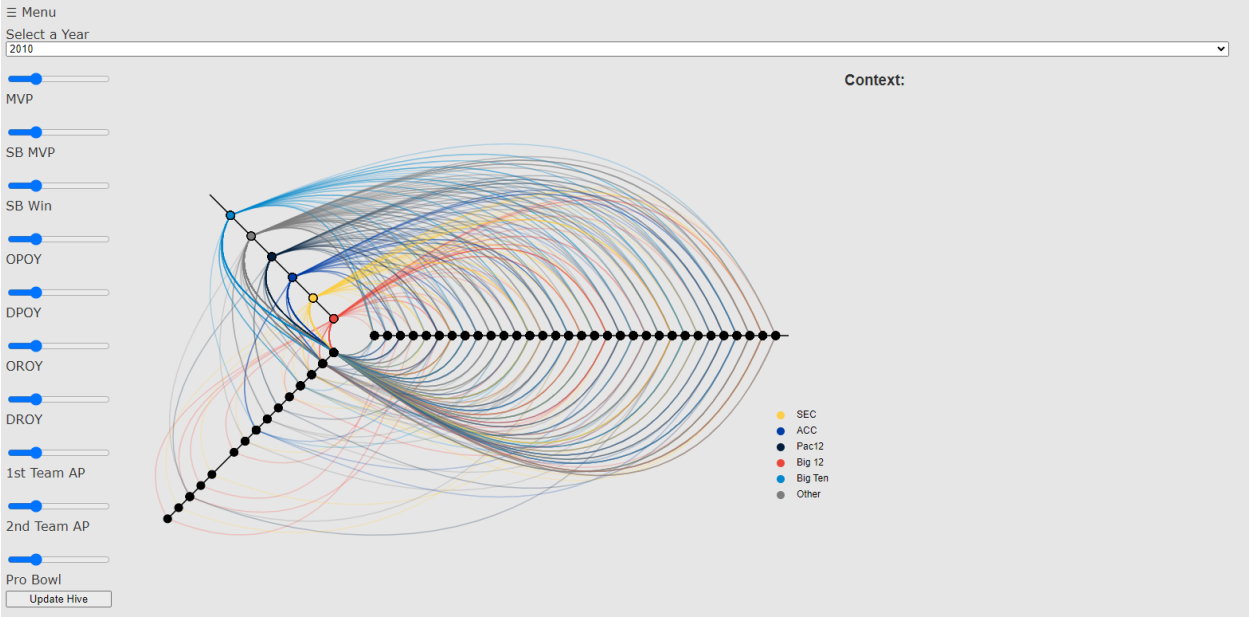

Figure 9: Hive plot

The hive plot we created is shown above in *Figure 9*. The plot displays College Conference, NFL team, and sum of accolades received. When the user hovers over the nodes along the axis, the connected links are highlighted, shown below in Figure 10.

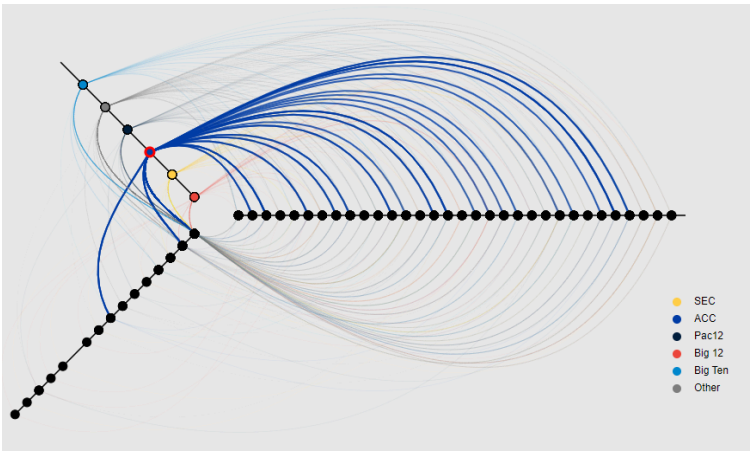

Figure 10: Hive Plot with Highlighted Node

Users have the ability to select from any of the draft years or to display all of the data at once. However, when the total data frame is selected, some of the interactive features are disabled inorder to allow the page to run smoothly. On the left side of the page there is a series of

sliders. These sliders control the weights of the awards, on a scale from 0 to 4. By adjusting the sliders, users can increase or decrease the importance of certain awards. For example, by setting all of the sliders to 0, except for the MVP slider, the hive plot changes to look like Figure 11 below.

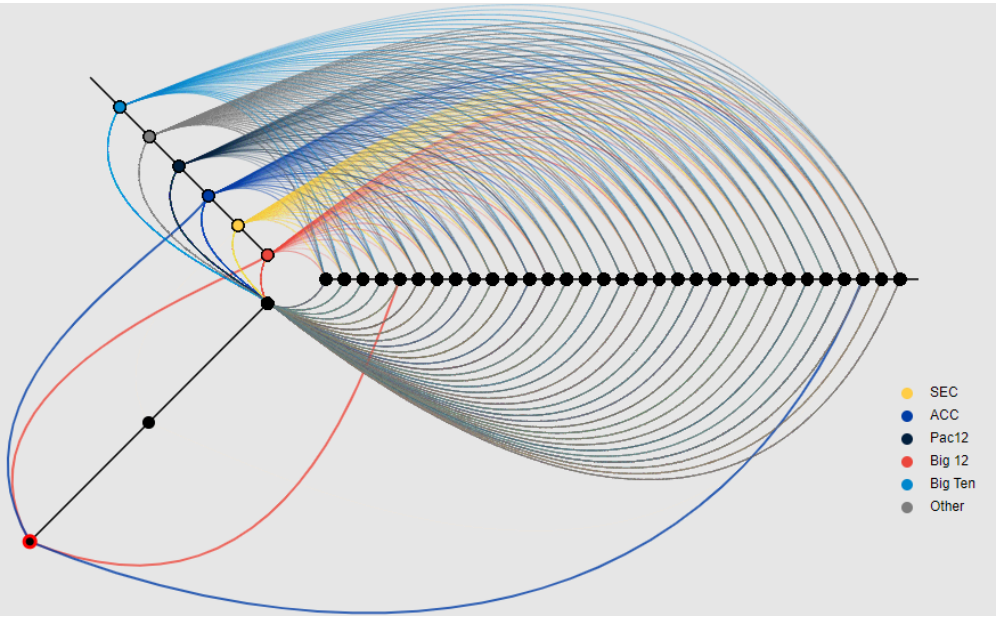

Figure 11: Hive Plot with Adjusted Slider Weights

When one of the award aggregation nodes is selected, the top five players who have award scores equal to the node are displayed. Figure 12 shows the players who have an award score of 15. Player name, position, NFL team, and college are displayed for each player.

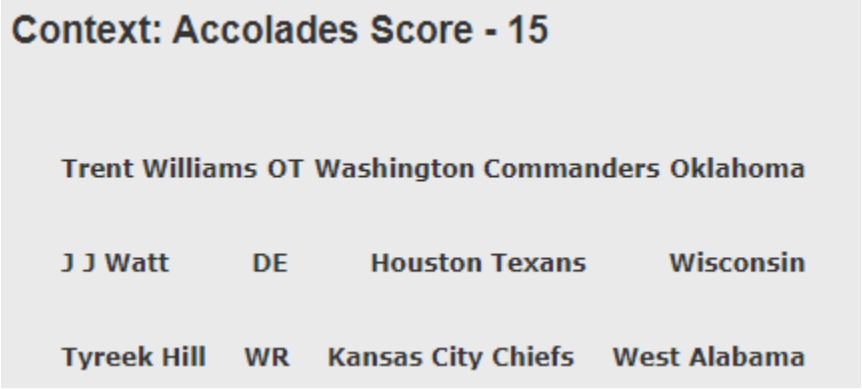

Figure 12: Hive Plot Accolade Score Output

We also created a scrolly-telly page for our visualization. It contains the hive plot, and as users scroll they can see the process of the hive plot being created. It starts with just the axes, then adds the nodes, and finally the lines are drawn.

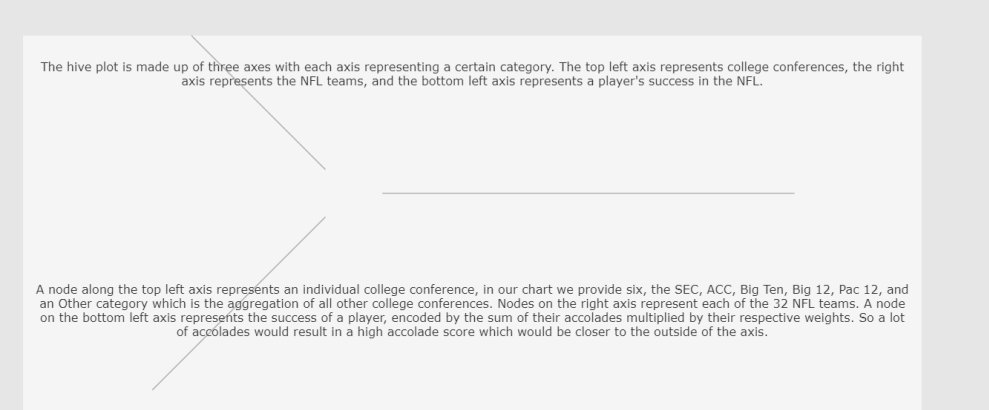

Figure 13: Hive plot Axes

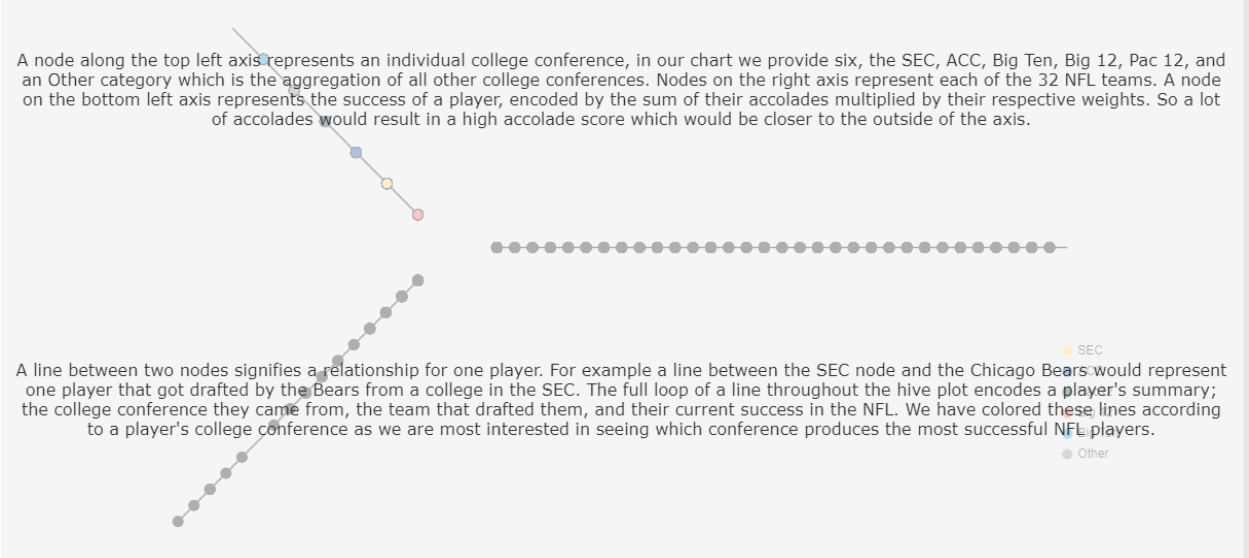

Figure 14: Hive plot Nodes

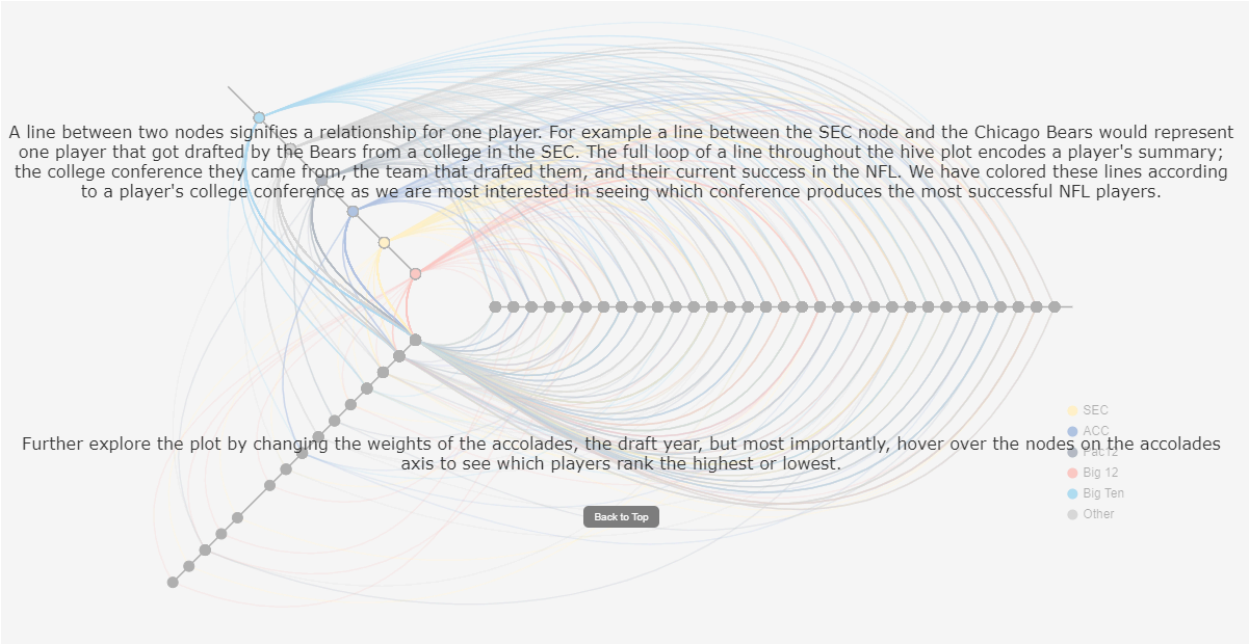

Figure 15: Hive plot Lines

We created small bar plots to show which conferences produce which awards. When hovering over a conference, all other bars fade away.

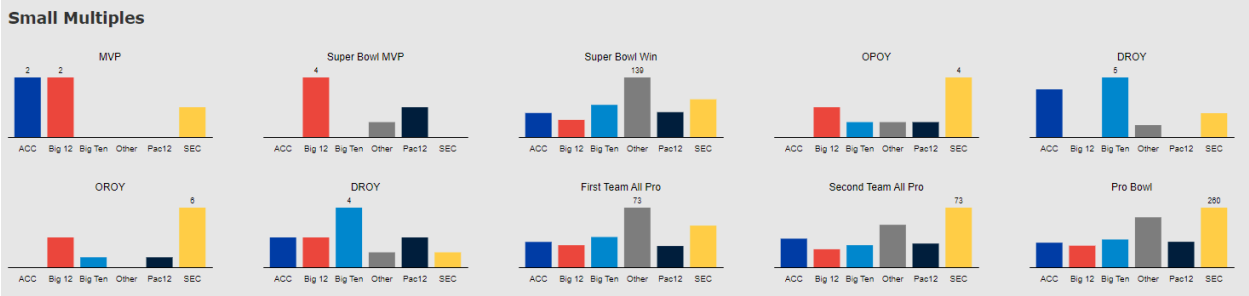

Figure 16: Small Multiples

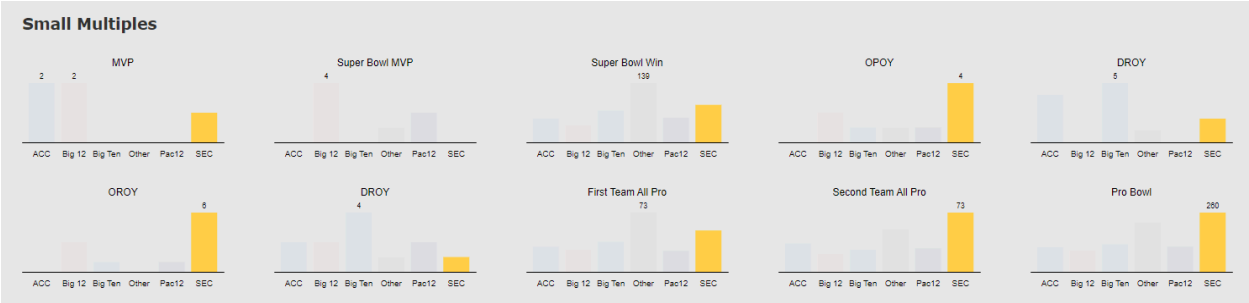

Figure 17: Small Multiples - SEC Highlighted

We created line plots to show which conferences produced the most First Team All-Pro, Second Team All-Pro, and Pro Bowl awards over time. When hovering, the other lines fade away. Users also have the ability to change the weights of the sliders.

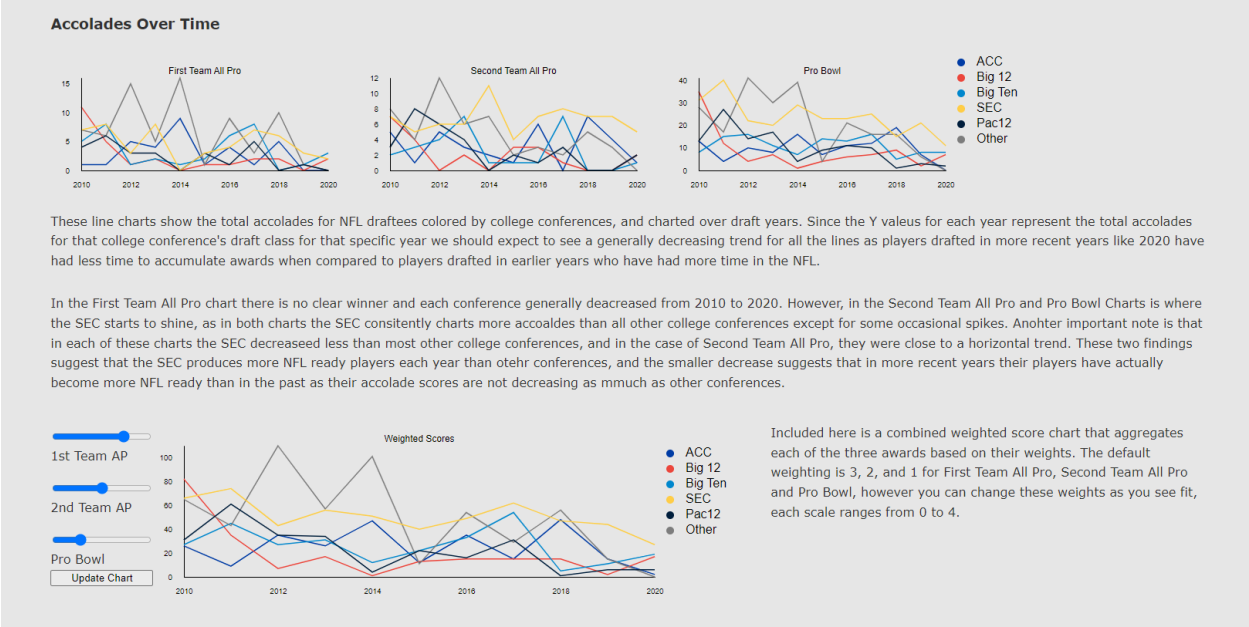

Figure 18: Accolades Over Time

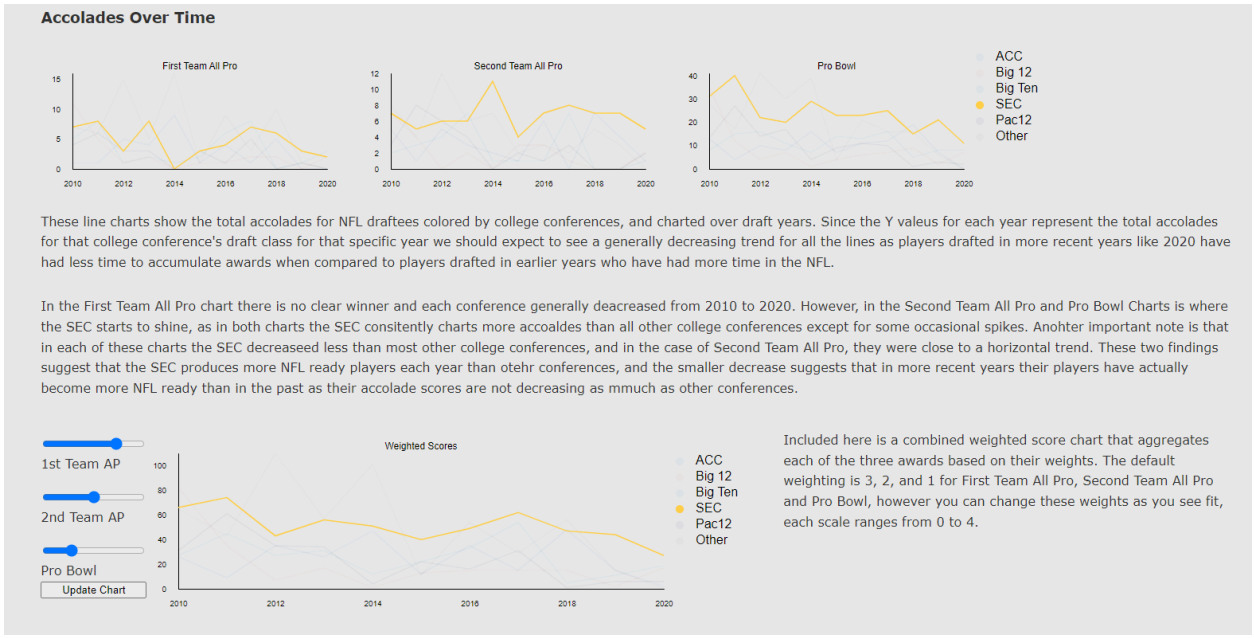

Figure 19: Accolades Over Time - SEC Highlighted

We created a bubble chart to break the SEC down by position based on awards. When highlighting an award, all other awards fade away as mentioned in Figure 7.

Evaluation

We see the Hive plot as a more neatly organized implementation of the flow from college conference to NFL team that was previously attempted by the Sankey diagram. We also found that using the hive plot allowed us to easily add another dimension to the chart, player success, that allowed us to evaluate the connections between a player's team and college conference, and their overall achievement in the NFL. The Hive plot also acted as an attention grabber for our audience. While this chart is well organized, there are many lines and good use of color that can draw a viewer in. We were also able to implement useful interactions like highlighting lines that passed through a node, and displaying the player context relevant to a hovered node in order to maximize the value of our chart and engage users. Lastly, the adjustable weights allow a user to dive in and find their own insights outside of our analysis. For future work on our hive plot we would recommend developing a more efficient interaction technique. While we are able to mostly avoid it by filtering our hive plot by year, when the total is shown there are many lines plotted, one for every player that was drafted from 2010 to 2020. As a result the hover interactions that highlight these lines can be delayed or seem laggy as there are many svg elements to update.

With our visualization anchored by the Hive plot we were able to introduce our overall analysis which we continued with more approachable visualizations. Through the use of small multiples and simple techniques like bar charts and line plots we were able to portray the story of how the SEC has dominated all other college conferences when producing NFL-ready players. We feel that the curiosity sparked by the Hive plot is complemented well by these simpler techniques that follow which help to answer and guide a user's questions. To finish our story we implemented a hierarchy bubble chart which acted as a source of inspiration for our users to explore further. We provided some insights that a user could find in this chart but again implemented useful hover interactions that keeps a user engaged and sparks their curiosity.

Lastly we feel that we created a user friendly experience for our project. Our website has a clearly defined navigation menu that is constant across all pages. We also include a home page link at the footer of every page. Our home page has a concise but complete About section that helps to inform users. These elements combine to create a web page that is easily navigable. Along with this we also recognized that Hive plots are not as common to some of our audience

as other data visualization techniques. As a result we dedicated a page in our project to quickly introducing the hive plot to a user. This page explains the elements of it, including the axes, nodes, and links, and how a user should comprehend each piece. By providing this guide it adds to the overall approachability of our project and helps add value for a user.

Works Cited

Aisch, G. (2014, June 20). *The clubs that connect the World Cup*. The New York Times. https://www.nytimes.com/interactive/2014/06/20/sports/worldcup/how-world-cup-players-are-connected.html

Casals, M., & Daunis-i-Estadella, P. (2022). Violinboxplot and enhanced radar plot as components of effective graphical dashboards: An educational example of sports analytics. *International Journal of Sports Science & Coaching*, *18*(2), 572–583. https://doi.org/10.1177/17479541221099638

Fu, Y., & Stasko, J. (2022). Supporting data-driven basketball journalism through interactive visualization. *CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3502078

Janetzko, H., Stein, M., Sacha, D., & Schreck, T. (2016). Enhancing parallel coordinates: Statistical visualizations for analyzing soccer data. *Electronic Imaging*, *28*(1), 1–8. https://doi.org/10.2352/issn.2470-1173.2016.1.vda-486

Krzywinski, M., Birol, I., Jones, S. J., & Marra, M. A. (2011). Hive plots--rational approach to Visualizing Networks. *Briefings in Bioinformatics*, *13*(5), 627–644. https://doi.org/10.1093/bib/bbr069

McKenna, S., Mazur, D., Agutter, J., & Meyer, M. (2014). Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2191–2200. https://doi.org/10.1109/tvcg.2014.2346331

Meko, T., & Lu, D. (n.d.). *Legendary? lousy? how your team has fared in the NFL Draft - Washington Post*. Legendary? Lousy? How your team has fared in the NFL draft. https://www.washingtonpost.com/graphics/sports/nfl-draft-history/

Tjärnhage, A., Söderström, U., Norberg, O., Andersson, M., & Mejtoft, T. (2023). The impact of scrollytelling on the reading experience of long-form journalism. *Proceedings of the European Conference on Cognitive Ergonomics 2023*. https://doi.org/10.1145/3605655.3605683