Michael Nwauche
Ethan Stavisky

Process Book

Overview and Motivation:

Our primary objective is to create an interesting and interactive visualization. We hope to create something that people would be eager to play around with and investigate the features and displayed information. We hope to learn more about the data we gather and insights about the NFL. In general, we expect to learn more about how to turn raw data into an effective visualization by utilizing user friendly features and visual designs.

The primary motivation for this project is that we both love to watch and analyze football. Football certainly gets less attention when it comes to data analytics when compared to the NBA and MLB. This is likely due to the nature of football being such a team sport and that the offense does not play at the same time as the defense. However, there is an increasing amount of data out there pertaining to NFL performance. We are both interested in looking at this data and how it is being used to alter how NFL teams make decisions today. So, when we were discussing a possible topic for this group project, we both naturally thought about football. We think it would be really cool to create some visualizations based on this topic and this data that we are interested in. This topic is not related to any research interest we have but, rather, simply a shared hobby.

Related Work:

The main inspiration for this project idea was our class project with NBA data. We both enjoyed creating that visualization because it related to sports and was able to effectively display lots of data in a small space in a highly comparable way. We thought that we could do something similar for this project while incorporating some interactive features. Another inspiration was simply the website pro-football-reference.com. This is a website that we often find ourselves looking at when researching football statistics. We both noticed the robust amount of information on the site and thought we could utilize it in order to create our visualization

Questions:

The primary question we are trying to answer is how has the NFL changed over time. Specifically, what trends can we find in NFL statistics over the course of its entire lifetime. We look at passing, rushing, receiving, overall offense, kicking, punting, returning and scoring. Through these main categories of NFL performance, we hope to

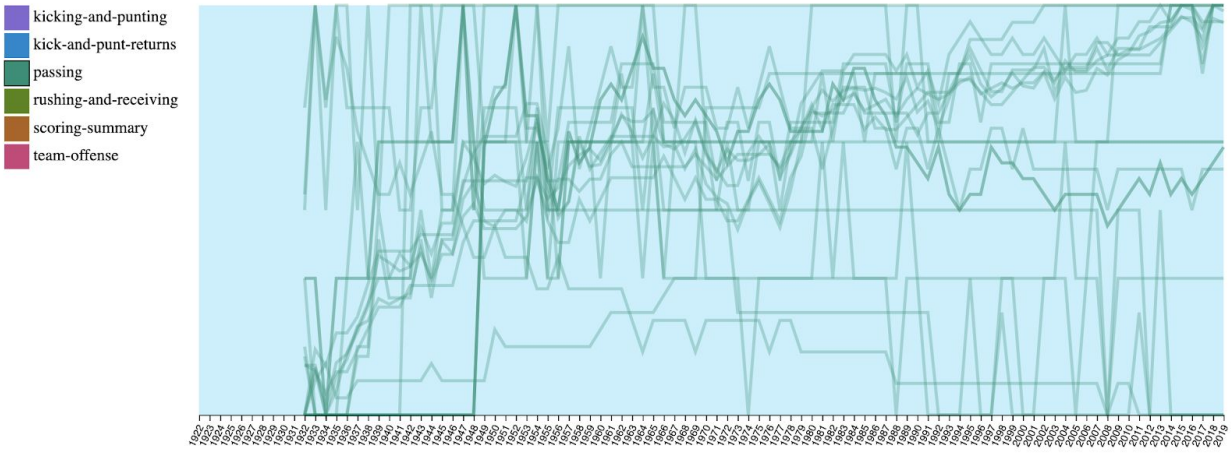
uncover how the game has changed over the years. Originally we also planned to incorporate defensive statistics but decided against it. In the middle of the process we considered including individual player statistics. However, we determined player statistics would not be insightful in determining overall trends in the game over time. Also, the data for individual statistics way back when the league was first created was challenging to find and unreliable. We eventually decided not to include defensive statistics because they are essentially included in offensive statistics. For example the yards and touchdowns scored would be the exact same and the yards and touchdowns allowed. So we decided to only investigate the categories stated above.

Data:

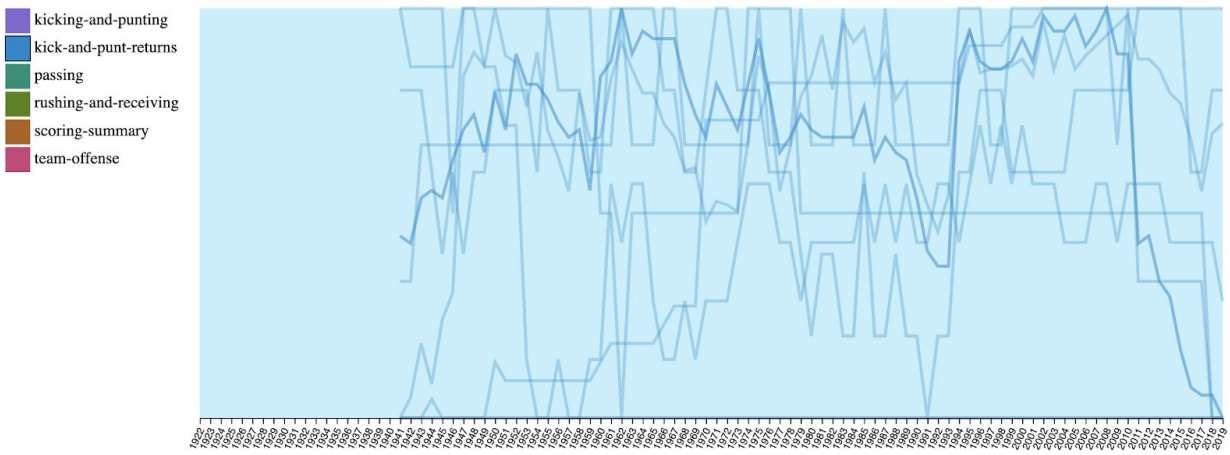
We gathered all of our data from pro-football-reference.com. This website has a tremendous amount of NFL data and had everything we needed. The data can be broken down by numerous features such as year, team, offense or defense, rushing or passing etc. Pro football reference allows you to download the data into a CSV file which is a format we are very comfortable with. Originally we thought the data scraping and cleanup would be very easy. However, we actually struggled to gather the desired statistics together onto one page and download them into CSV. Although the website had everything we needed, its process to download the data was actually quite challenging. But, we were able to clean the data from within the website through its search features. In the end, this portion of the project was a larger challenge than expected.

Implementation:

Our visualization shows various NFL statistics over time. The main functionality of the visualization is that you can choose the category of statistics you would like to look at by clicking one or multiple boxes on the left side of the visualization to show that category. This allows the user to decide which statistics they would like to look at.



For example, this shows passing statistics by clicking the “passing” box on the left.



This shows kick and punt return data by clicking the corresponding box on the left.

Exploratory Data Analysis:

The first thing we did once we found, scraped and cleaned our data was to make some initial visualizations. We made some basic bar graphs that allowed us to see some general trends in each specific statistic we looked at. The main takeaway from these visualizations was that there were noticeable and interesting trends worth putting into our final visualization in nearly all of these statistics over time. These visualizations also gave us a sneak peak at what our final visualization would show us. These insights let us know which statistics to keep and which to get rid of in the data. It also showed us that these trends are related to one another and that should be apparent to some degree within the visualization.

Design Evolution:

Originally we considered making aligned bar charts in order to show all of the data. We thought that we could make separate graphs for each type of statistics opr for individual statistics. However we decided against that choice and to instead do what you can see in the final project. We believed that this type of visualization would be more pre-attentive for the viewer. We thought they would be better able to see trends in the data and that the space would be used more effectively. However, in order to do this type of visualization we would need to incorporate more interactivity in order to make sense of such a large amount of data in a smaller space. So, we deviated from our proposal because we decided to condense our amount of visualizations from multiple into one.

Analysis:

The main thing that we learned from our visualization is this story of how the NFL game has evolved over time. It all begins with massive increases in passing attempts,

yards, touchdowns and accuracy and huge decreases in interceptions. Due to this, receiving yardage and touchdowns have increased massively. Because of the rise of the passing game, rushing attempts decreased largely. However, due to the effectiveness of the passing game, rushing touchdowns only decreased as teams found themselves in opponents territory more often. Similarly, rushing yardage decreased only slightly as increases in rushing effectiveness due to defenses focusing on the passing game more offset the decreased attempts. The prominent passing game put offenses in field goal range more often so field goal frequency and scoring increased. Because offenses were more effective, punt frequency decreased. We also saw that field goal accuracy, especially at further distances, and punt distance increased likely due to advances in athlete training over the years. Due to the large increases in offensive success we also saw that kick and punt returns per game, yardage and touchdowns decreased. Offensive statistics improved across the board due to increases in total yards, touchdowns, plays, yards per play and first downs and decreases in turnovers. In the end we saw that scoring has increased greatly and is now heading towards 25 points per game per team. Clearly, we were able to learn a lot about the trends that have taken place in the NFL game since its inception.

Our visualization works okay. It could be further improved by a few things. First, it would greatly benefit from a y axis which changes based on the type of data being displayed. It would also benefit from the lines being labelled as to what statistic is actually being shown. Also, our statistics are grouped together based on category however, they are not evaluated by the same metrics so the visualization would benefit from further splitting up the data. Passing yards and touchdowns are two very different stats measured in very different ways for example. Perhaps being able to choose the category first and then the individual statistic through some sort of drop down menu. This would make the visualization much easier to understand for the user. Or maybe splitting the visualization into a few separate visualizations based on the category and then selecting the individual statistics in a similar way to how you select the category currently. The visualization could also benefit from the incorporation of more interactivity. In sum, our visualization could use a lot of improvements in order for the user to better understand the data that is being shown.