Touchdown Tables

Process Book

CS 5630 (Fall 2019)

Hannah Bruns, Zachary Cutler, Joshua Morrell

Table of Contents

Team Information

Project Title - Touchdown Tables

Project Repository - https://github.com/ZachCutler04/DataVisProjectGroup

Project URL - https://zachcutler04.github.io/DataVisProjectGroup/index.html

Team Members -

- Joshua Morrell, joshua.morrell@utah.edu, u0972427

- Hannah Bruns, u0629077@utah.edu, u0629077

- Zachary Cutler, zachcutler@yahoo.com, u1025642

Overview and Motivation

As an avid football fan, Zach convinced our team to come up with a project that analyzes decision making in NFL football games. In our research of current football visualizations we were surprised to find that many of the existing visualizations are limited to static representations of small datasets. Because of the large size of the dataset provided to us through an open source NFL data scraper we have a lot of options as to what would be possible to visualize as well as the scope of the story we wish to convey. Additionally, due to the size of the dataset, we believe an interactive visualization to be the best way to navigate such data. Finally, due to the lack of interactive visualizations in this space, we believe that we have the opportunity to use techniques that have not been used to visualize football analytics before.

Project Objectives

The primary question we are trying to answer is how well NFL coaches make macro level decisions in football. Questions like how frequently a team should punt/go for it on 4th down and how frequently a team should run vs pass the ball have recently been explored more in analytical circles around football, but no major data visualization for such analysis has been made that we are aware of. In addition to exploring the data, we are hoping to appeal to a wider range of audience. Historically analytics has been met with resistance from NFL coaches and fans who grew up playing a game that looks very different from the game today, and are resistant to being told that everything they have learned is wrong. By creating an unbiased visualization that shows the data accurately, we hope to try and convince the more resistant parts of the NFL community to try and look at the data objectively.

Data

We are using data from nflScrapR, an R library for scraping data from every NFL game since 2009, updated in real time. The data provides a large suite of information from every play, ranging from simple statistics like yards gained and players on the field to advanced metrics like EPA and Success Rate. We plan to use certain columns of the data more than others. These columns include many of the different EPA and success rate metrics, as well as the type of every play, such as run/pass, punt/kick, etc. A link to the github from where the data is pulled: https://github.com/maksimhorowitz/nflscrapR

Data Processing

The R library we are using to scrape data allows us to create the exact dataset we need without the need to clean an existing dataset. We are expecting to utilize a subset of all available data provided by the data scraping library, where the available data dates back from 2009 to the present day. The only data processing that will be required is converting the games to JSON format to use with Javascript easily.

Visualization Design

We are going to create three different views that will work together to illustrate some of the most important decisions that are commonly faced by NFL teams, and compare what decisions different teams make..

The first view will show the decisions that teams make on fourth down, especially how frequently they punt/kick a field goal/go for it. The visualization will be a representation of a football field that displays the field positions of successful 4th down conversions depending on the position of play. Our goal with this illustration is to visualize the success rates of 4th down plays when teams decide to go for the 1st down conversion at different yardages, and display the value of going for it compared to punting.

The second view will be a parallel line chart with different axes variables that will show how different factors correlate to one another and correlate to the outcome of the game. For example, you could see how passing yards and rushing yards correlate to one another or how they correlate to the result. This visualization will include some brushing and sorting features on the variable axes, with different selections for more focused data, such as selecting a specific team and a range of years.

Our third and final visualization will attempt to visualize how frequently teams run the ball compared to throw it. We will show a comparison of runs/throws in a split bar chart showing the number of passes thrown at a specific down/yardage compared to the number of times the ball is run at that same down/yardage. This comparison is a popular topic of debate in the NFL, deciding whether a team is more successful running or passing the ball at certain yardages. We would also like to add the option for users to change these axes with other columns of data that could make for interesting comparisons.

Project Features

- Fourth Down View

  - A way to select a down and distance for a fourth down. Show historical statistics for that fourth down.

  - Some way to show expected EPA based on the different decisions that can be made

  - Tooltips over historical data that shows the game state/team/date.

  - Filtering on team

- Correlation Analysis View

  - Parallel line chart with every game in data

  - Brushing ability on each axis

  - Filtering on team/year

  - Ability to move axes/select new axes

- Run/Pass View

  - Split Bar Chart for 2nd, 3rd and 4th down.

  - Some way to visualize first down, though a split bar chart might not be ideal.

  - Ability to compare different teams, or compare the same team over time.

Optional Features

- Fourth Down View

  - Way to compare how frequently different teams go for it/succeed.

  - Filter by situational variables like time/score

- Scrollytelling story that steps through important decisions made by one team during a season/playoff run.

- Correlation Analysis View

- Better filtering methods, by week, playoffs, etc.

Schedule

Friday, November 1

- Create the datasets we will need for each of the three visualizations

- Setup a basic class architecture for each view with update and linking functions

- Load the data into each view

- Create D3 scales for all view axes

Friday, November 8

- Visualize all of the data in all three views

- Style all of the views to match the specifications

- Begin to add must have features

Friday, November 15

- Add interactive functionality to each view

- Finish the must have features

- Begin the optional features

Friday, November 22

- Add documentation to the codebase

- Record and edit a video discussing the visualization

- Polish the styling and animated transitions

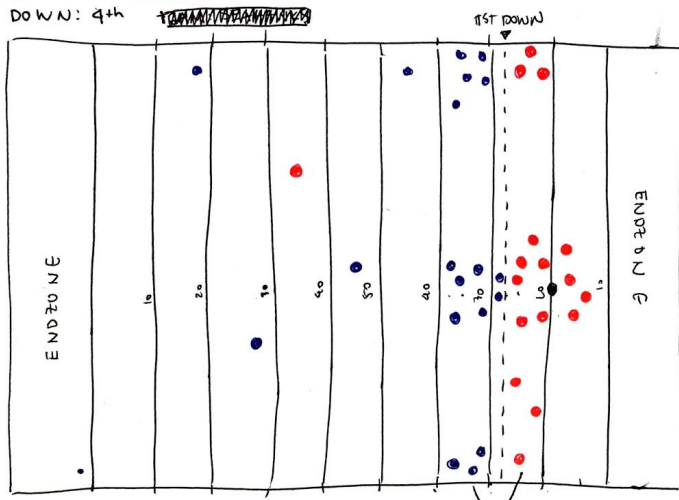- Finish coding the visualization

Wednesday, November 27

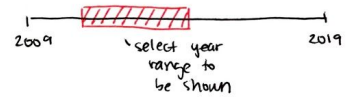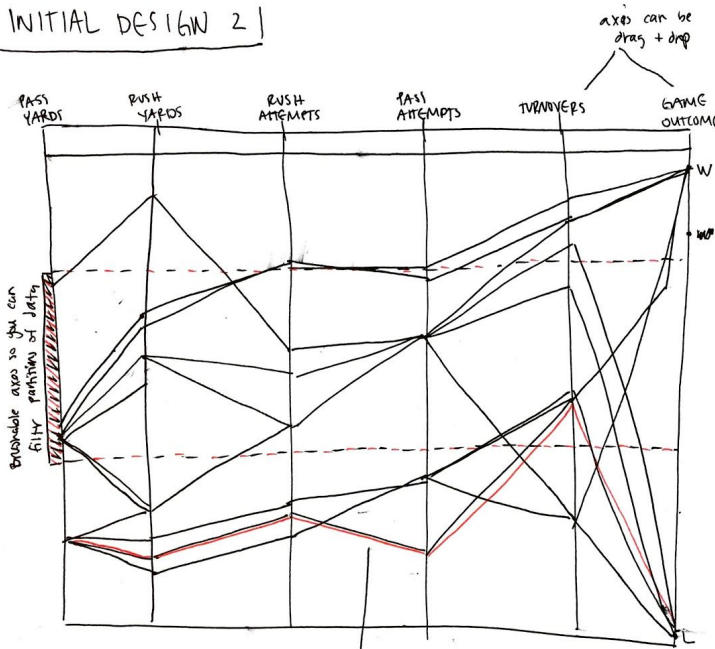- Turn in the project

Design Sketches

INITIAL DESIGN 1

DOWN: 4th

1ST DOWN

- SELECT A YARDLINE FOR 4TH DOWN PLAY ●
- SELECT WHERE THE FIRST DOWN LINE IS - - - -

SELECT A TEAM: [   ]

2009 [/////] 2019
`select year range to be shown

ENDZONE

ENDZONE

10 20 30 40 50 40 30 20 10
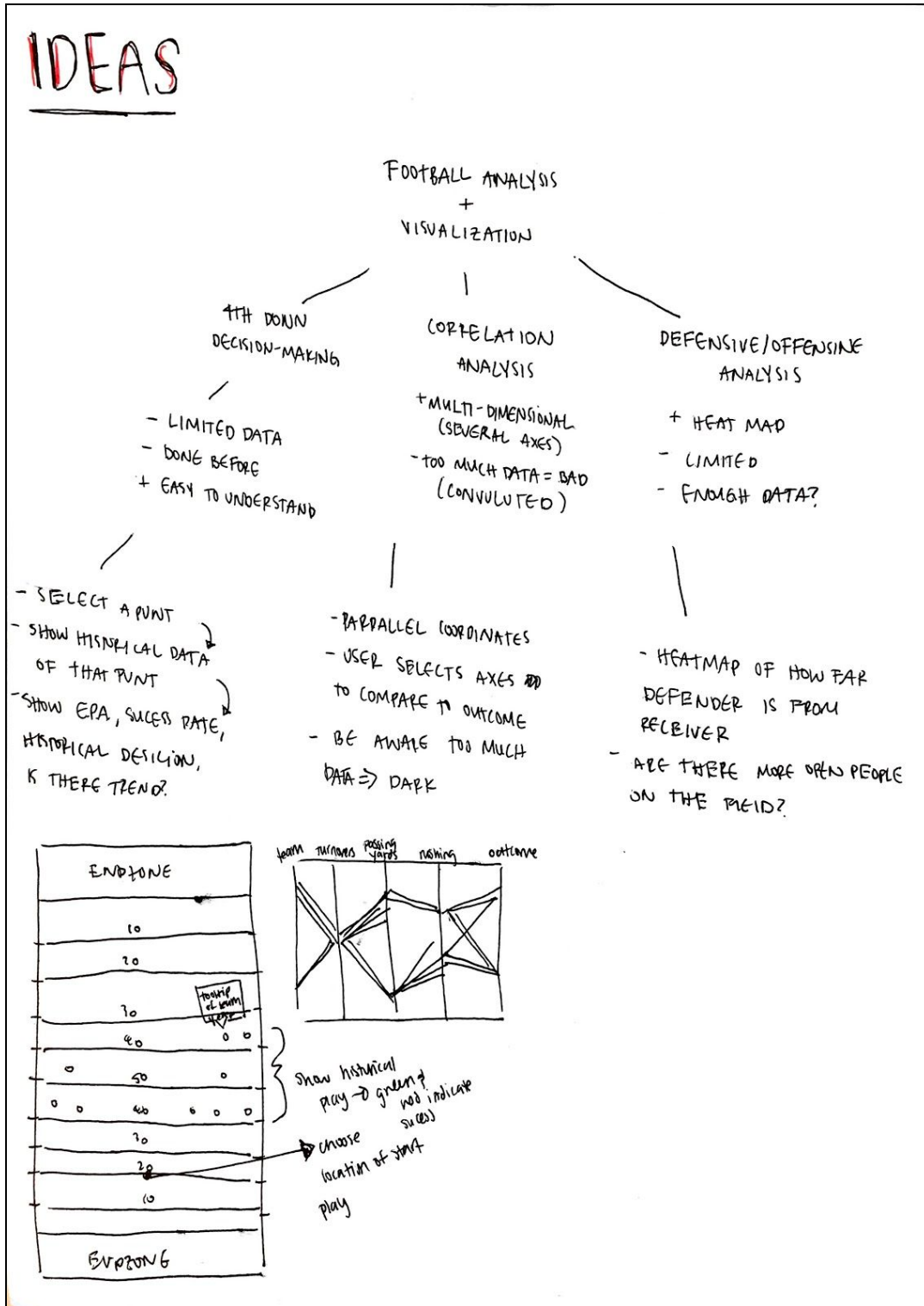
● 1ST DOWN
● TURNOVER

tooltip shows on hover
TEAMS, YEAR
SITUATION (SCORE, TIME OF GAME)

INITIAL DESIGN 2

axes can be drag + drop

PASS YARDS | RUSH YARDS | RUSH ATTEMPTS | PASS ATTEMPTS | TURNOVERS | GAME OUTCOME
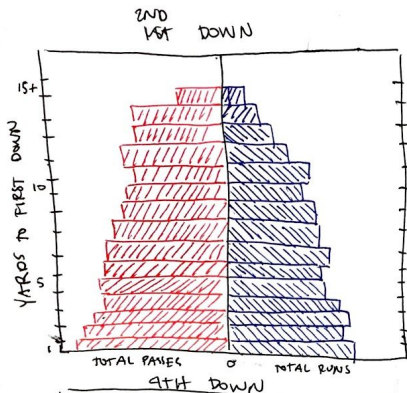
W
L

Brushable axes so you can filly partition of data

POSSIBLE AXES
- TURNOVERS
- TOUCHDOWNS
- PASS YARDS
- FIELD GOALS
- TOTAL YARDS
- NUMBER OF PUNTS

→ selectable so you can change axes.

each line represents a single game of a single team

→ on hover, more game info will appear

→ dull out other lines

9

INITIAL DESIGN 3

2ND 1ST DOWN

- WE WANT TO REPRODUCE THIS GRAPH TO SHOW HOW THE YARDS RUN VS THE YARDS PASS DIFFER FROM DOWN TO DOWN.
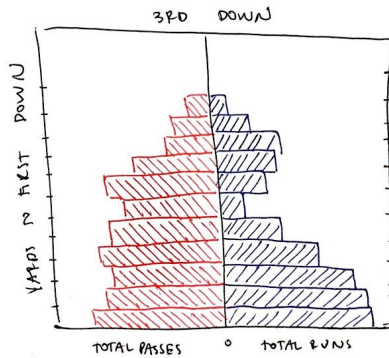- WE CHOSE NOT TO INCLUDE THE 1ST DOWN GRAPH AS IT WILL MORE THAN LIKELY BE AT THE 10 YARD LINE.
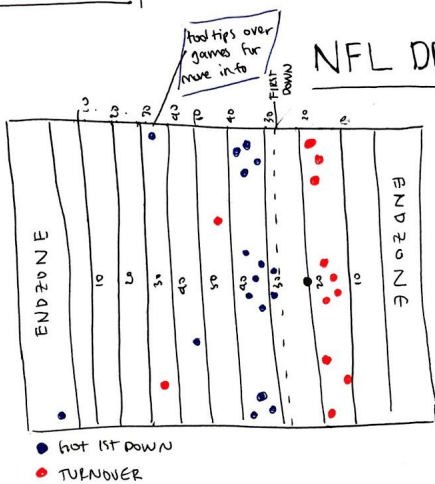
2ND DOWN

4TH DOWN

- THERE WILL BE ALWAYS BE 3 GRAPHS REPRESENTING THE DOWNS BUT WITH THE OPTION TO CHANGE THE X & Y AXIS VARIABLE.
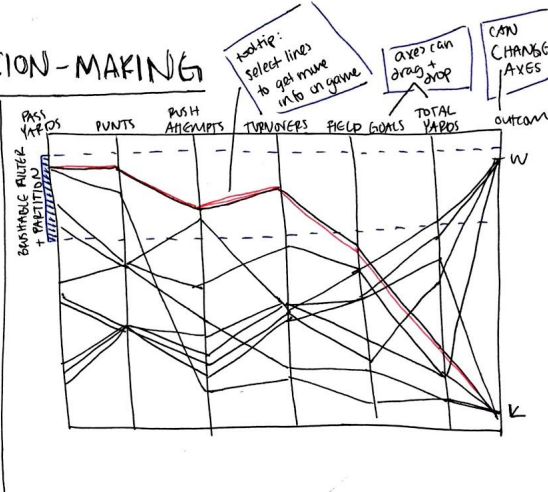- SOME EXAMPLES:
  - COMPARE 2 TEAMS
 - SUCESS VS FAILURE OF PLAY FROM THAT YARD LINE
  - INTERCEPTIONS

3RD DOWN

REALIZATION

NFL DESICION-MAKING

tool tips over games for more info

tooltip: select lines to get more info on game

axes can drag + drop

CAN CHANGE AXES

- HOT 1ST DOWN
- TURNOVER
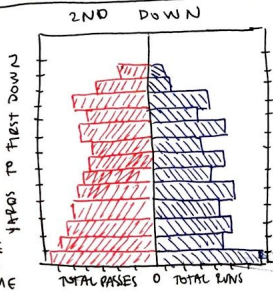
OPTIONS FOR DIFFERENT AXES ON CLICK. INSTEAD COMPARE TWO TEAMS. WE DIDN'T INCLUDE 1ST DOWN BECAUSE IT WOULD ALL LOOK THE SAME

2ND DOWN

3RD DOWN
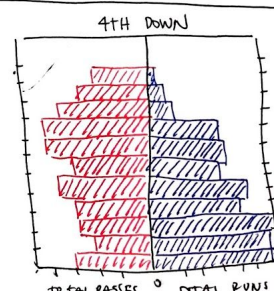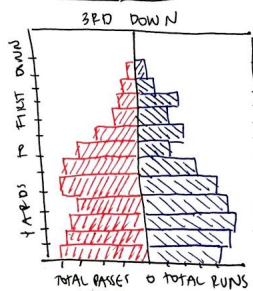
4TH DOWN

Project Proposal Feedback

We met with the Fantasy Football team which consists of the team members: Jake Maschoff, Raj Patel, and Blaze Kotsenburg.

First Chart (4th Down Play Results)

The first chart in our visualization is a visual representation of a football field with lines drawn where each 10 yard marker would be and an end zone on each end of the field. The user will have the option to drag around two lines on the field that correspond to a line of play, and the yards left to go on fourth down. On the field we will display circles that represent successful or failed plays that occurred at the selected yardages for a specific team.

Some of the critiques that the Fantasy Football team had of this view were that we should also show the current score and time left in the game in a tooltip when hovering over a circle on the field. This would help our users see a correlation to desperation of a team and the decisions they make at different times in a game as well as at different score differences. They also thought it would be interesting to know what ranking the selected team had in that specific season and how that compared to other seasons and other teams ranking. We could also include this ranking in the tooltip for each play that is represented on the field, or we could allow the user to select a specific year above the visualization and display a ranking next to the selection.

Second Chart (Football Statistic Correlation)

The second view in our visualization is a parallel line chart with multi-variable axes displaying correlations between different football statistics. The view will have the option to add, remove, and reorder the axes as well as brush an individual axis for a more fine grained selection. When hovering over an individual line on the chart, a tooltip will appear with a more detailed description of each statistic.

The Fantasy team thought that this view would be a great place for an author driven storytelling animation. At first this graphic could be difficult to know how to use without some kind of orientation. A good way to introduce the viewer to some of the features, such as changing axes variables or adding and removing axes, would be to add an animated story that highlights the features as well as something we found to be interesting.

Third Chart (Running vs Passing Plays)

Our third view is three side-by-side parallel bar charts that represent the second through fourth downs in football. The axes for this view would default to the left x axis being number of successful running plays, the right x axis being the number of successful passing plays, and the y axis being the number of yards left to reach the next first down.

The main critique we received for this view was to offer the user an option to stack each bar chart so that the user can see a comparison between not just successful first down conversions but a success versus failed passing/running conversion at each distance to the first down.

Connected View

After reviewing each view independent of one another we discussed with the Fantasy team how each view will be displayed on the webpage and the interactivity between all three views. The Fantasy Football team thought that we should find a way to link the datasets of all three visualizations in order to connect the views. This would allow the user to make a selection on any of the views and the other two views would offer up a different perspective on the same data. They also thought that because each view would most likely take up a large portion of the screen that we should find a way to show and hide each view in order to expand just one or two views at a time. This would allow the user to focus on one view at a time, and also gives our team an opportunity to add a storytelling component to each view when transitioning between each view.

Milestone 1
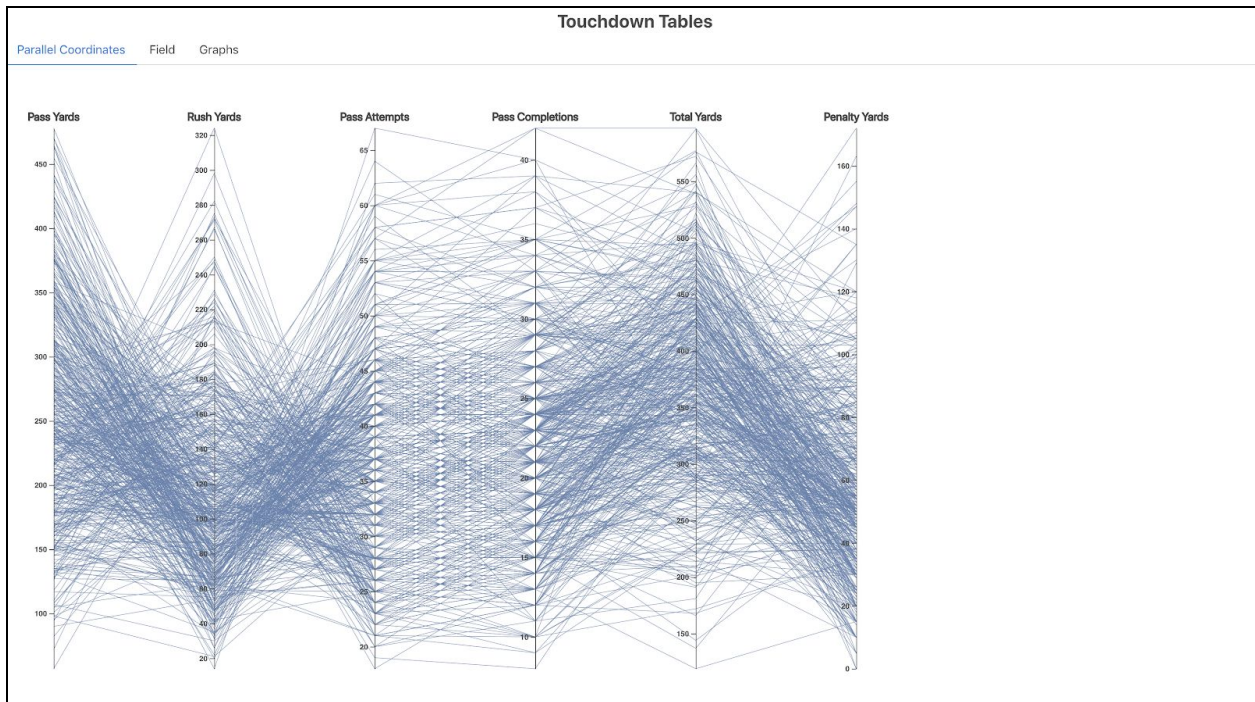
The figure above shows the first view in our visualization at the time of milestone 1. At this time we are able to show a parallel coordinate chart for a single year of NFL games with simple aggregate data for each column axes. The user is able to select and brush multiple axes on the chart and view see an updated view of the games that fall within the brushed selection. When a user makes a selection all of the lines that are not within the selection are greyed out, highlighting the games that do fall within the selection.
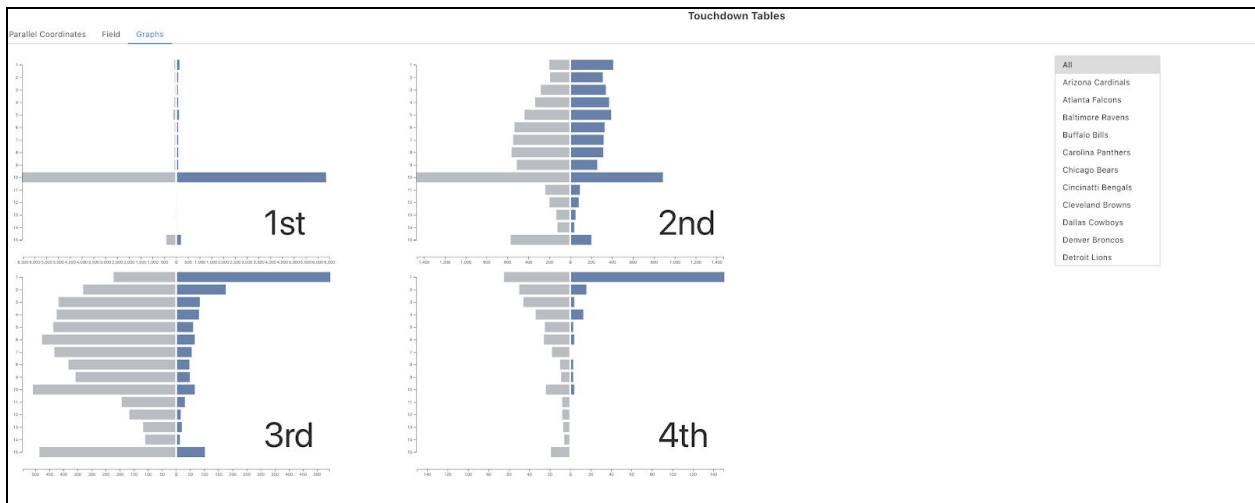
This screenshot is of the third view in our visualization, which shows the number of times teams will pass the ball versus running the ball at specific yard lines during an NFL game. There are four split bar charts in this view that represent this data at each down of a play. To the right of the bar charts we have added a list of teams that the user can choose from to filter the data down from the default full set of all NFL teams.

Data Aggregation

As of milestone 1 we were able to successfully scrape all the data we will need from the nflscrapR library into csv files and store those files in our repository. The three categories of data we have obtained are roster data for players on each team, game data which highlights the result of each game and the two teams who competed, and play by play data which gives us statistics for each play that occured in every game since 2009. The play by play data that we scraped was by far the largest file, and we spent some time writing scripts to remove column attributes of statistics we did not think we needed. We also ran into the issue of not having statistics that we originally wanted which forced us to do our own data aggregation. For example, one of the stats that was left out of the play by play data was total yards covered per team each game. This was as simple as summing the total passing and total running yards for a team during a specific game. In our repository within the totals directory is all the csv's that we created with custom aggregate data.

Hosting

Using GitHub pages we have been able to host a live look at our current visualization. Hosting the site has given us a head start on this task that was meant for the final submission. Having a live site has been a great tool to debug problems that arise when using javascript libraries outside of a local development environment. Site URL: https://zachcutler04.github.io/DataVisProjectGroup/index.html

Design Evolution

The first large change in our overall design was to get rid of our third view, which was the four charts showing run versus pass rates on each down of a play. As a team we felt that this view had little to offer in terms of user interactivity and overall had less to do with the other two views as a whole. We chose instead to focus on the connectivity and interaction between the parallel coordinate graph and the football field visualization.

The second major change to our design came after meeting with the TA around the milestone deadline. At the time we had nearly completed our parallel coordinate view and the symmetric matrix for games but had otherwise been struggling to implement the field view with little to no progress made. After discussing some methods of interaction for the field with the TA, we came up with the design that you see in the figure below. This design will allow our users the option of picking a single game to interact with or open up the entire dataset of fourth down plays to interact with. Our goal with this design was to connect the two views that had originally been separate visualizations into a single visualization.

The user would be able to filter and select a game in the parallel coordinates view and then smoothly transition into the field view where we would render all fourth down plays made in the selected game. They would have the option to click on a play and view outcomes of plays made in that same location in other games.
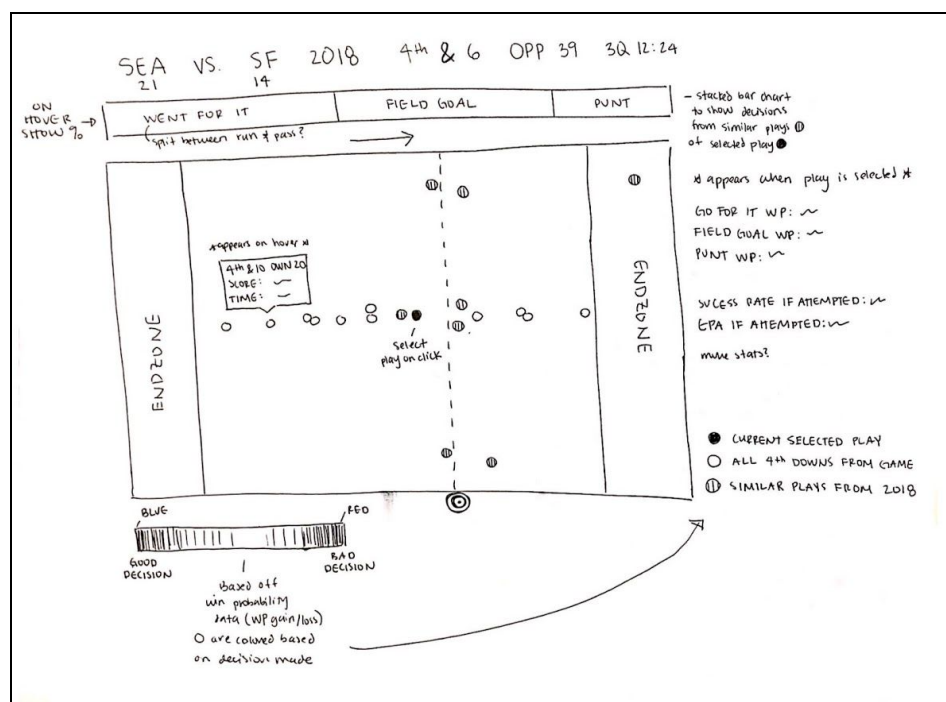
Implementation

When the user first enters the site they will immediately have the options to stay on the visualization page or navigate to the process book, or the screen cast which briefly describes the intended use of our visualization.

The first view that the user will be presented with on the visualization page is the parallel coordinate view. Here the user can see general trends in basic football game statistics across all games in a single season. There is the option to brush across the vertical axes and select a smaller subset of games. When a user selects a subset of games, the corresponding games table will update by only coloring the games that are within that subset.
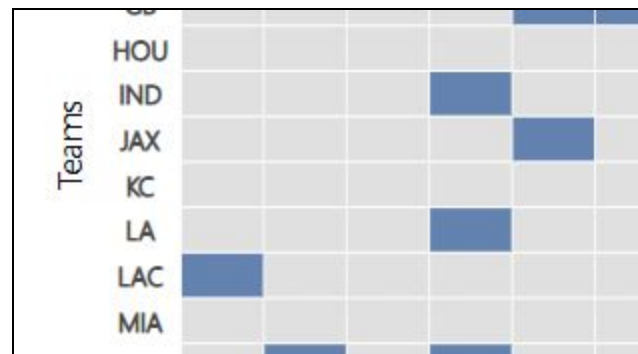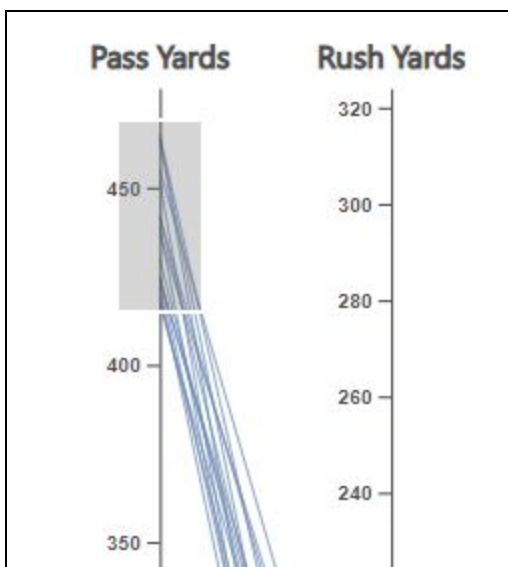

Upon clicking a game within the game table, the page will smoothly transition down to the football field view and present the user with all fourth down plays that occurred during the selected game as circles on the field. Above the visualization basic information about the selected game is displayed, as well as the option to pick a team to

be actively displayed. To change which team is active the user can click on the name of the team and the field will update with the fourth down plays of that team.

HOU vs **DAL**

4th & 1   5Q 05:40   OPP 42

After clicking a play on the field the user is presented with two main statistics, the first being a percentage breakdown of what teams are more likely to do in a similar situation, the second a breakdown of success versus turnover percentages for similar plays.

4th & 10   OPP 50

Punt: 91%   Go for it: 9%

Success rate when going for it

Turnover: 50%   Success: 50%
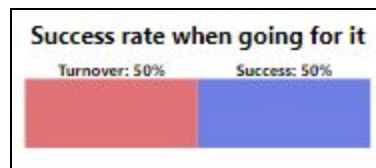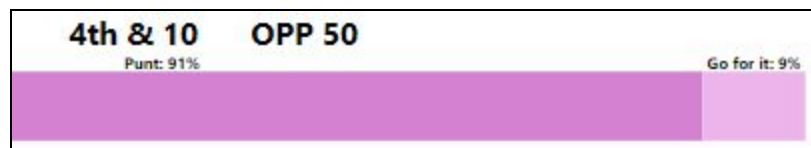
Apart from selecting one specific game, users can instead click on the "Try Your Own 4th Down" button which will open up the field to the entire 4th down dataset. After clicking on the button, the user can drag the circle, which marks the field position of a play, to any position on the field as well as drag the yellow target first down line to a position of their choosing. Once either position is changed the field will update with any 4th down play that matches that criteria.

Try Your Own 4th Down

Evaluation

An interesting observation we encountered when implementing and using our visualization was that for the most part teams will not go for the 1st down play when left with only 1 yard to go on 4th down. The main goal of our visualization was to present an argument such as this to our users to try and highlight key decisions in NFL games that could possibly lead to a change in play outcome. While our field visualization specifically focuses on 4th down plays, the parallel coordinate plot shows a direct correlation between several variables on what affects the overall game outcome. What we found to be the largest correlation with a game winning outcome was a high number of rushing yards.
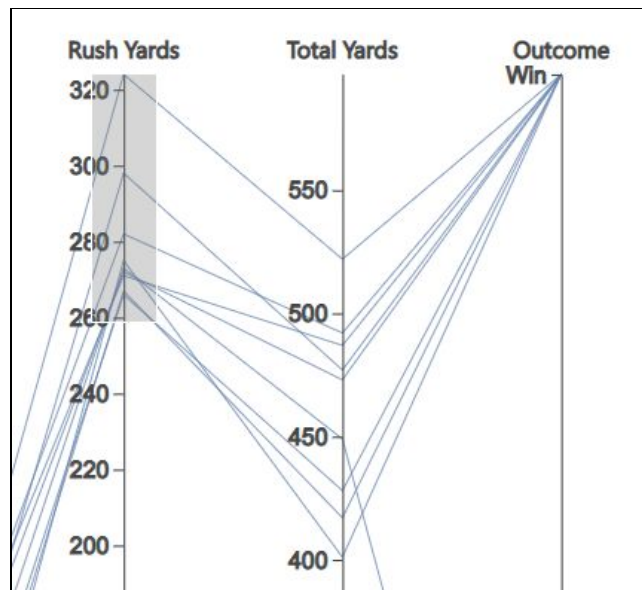
One improvement that our team would have liked to add to the field visualization is to not only display plays that match a similar situation to the one selected but to also evaluate a probability model that could point to whether the decision to go for the 1st down or not is a good choice. We began to evaluate this problem towards the end of the implementation phase of our project and would have liked to continue to pursue this feature. The result would allow a user to select a field position and yard to 1st down and be presented with a probability of success if their team went for the 1st down.