UG-02 Process Book: Modeling NFL Play Calling Decisions

Collaborators:
Simon Draper, ssdrape@clemson.edu, C10395791
Brigid Murphy, bnmurph@clemson.edu, C66342697
Will Sutton, wfsutto@clemson.edu, C54960997

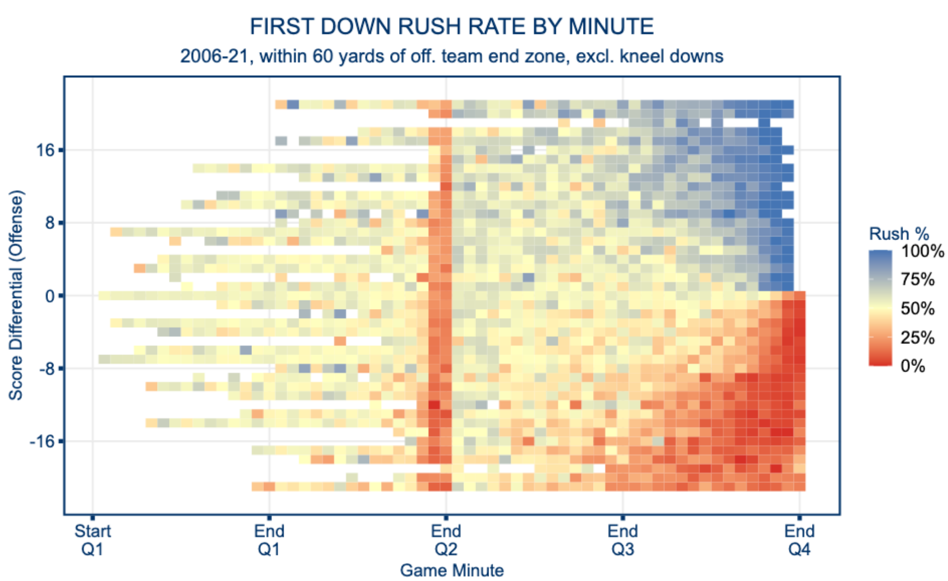
Link to project repository: https://github.com/bnmurph/bnmurph.github.io
Link to website: https://bnmurph.github.io/

Overview and Motivation

We wanted to do some sort of sports analytics for our project as we are all big sports fans. Additionally, Will is an intern in the sports science department with Clemson Athletics and has experience in sports analytics. We chose to narrow in our topic from sports in general to football specifically, as we know there is lots of data available and it's a sport we all love to watch! Thinking of what happens during a football game and what could be measured, we started coming up with questions. We were most interested in how play calling decisions are made, and what factors affect them. We also want to look at the success of these decisions, because that's what creates an exciting football game. We then found a dataset that would work well to answer these questions.

Related Work

Will previously worked with the NFLfastR dataset over the summer. When watching football on TV, there is often a flood of stats thrown at the viewer. We thought visualizations for these statistics would be super helpful and help the average football fan understand them better.



Above is one of our inspiration visualizations, it can be found at:
https://operations.nfl.com/gameday/analytics/stats-articles/
This visualization uses score differential and minutes of the whole game. We made a similar heatmap, but we decided to look at minutes remaining in the half due to the space available on our dashboard. Overall the patterns we found are fairly similar. We also decided to make a similar heatmap with down and distance on the axes.

Questions

Overall question: What factors most affect play calling decisions in NFL games, and how successful are the calls?

Questions we came up with while exploring our data:

1. How are run/pass rates affected by down and yards-to-go?
2. How successful are different types of playcalls?
3. Does clock time and score differential affect the play calls?

Data

We pulled our dataset using nflfastR, which is an open source tool to scrape and collect data from the National Football League (NFL). It can be used to compile a record of every play, penalty, timeout, and coach's challenge from every game of every season since statistics have been tracked. We decided that using this to create a workable dataset of play-by-play information would be a viable option to explore the factors that impact play decisions and the results driven by those decisions. Details on nflfastR: https://www.nflfastr.com/

We pulled data from the 2018-2022 seasons into Python using nflfastR. We had to do some data cleanup. This included removing all penalties, kickoffs, field goal attempts, punts, timeouts, and extra point attempts, as well as trimming off many extraneous attributes. We want to be left only with information about run/pass plays, so that we can analyze these types of play calling decisions most directly. Using the attributes we have in our dataset, we will be able to answer our key questions.

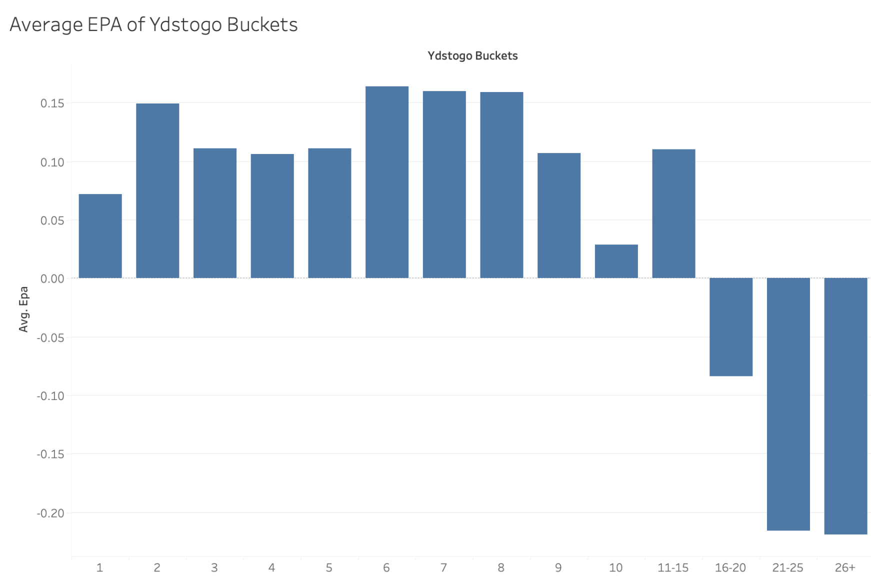
Then, we used some data from the original set to create aggregated attributes that could be better used in our visualizations.

This left us with 153,357 rows and 11 columns. The rows are the items and the columns are the attributes. They are listed below.

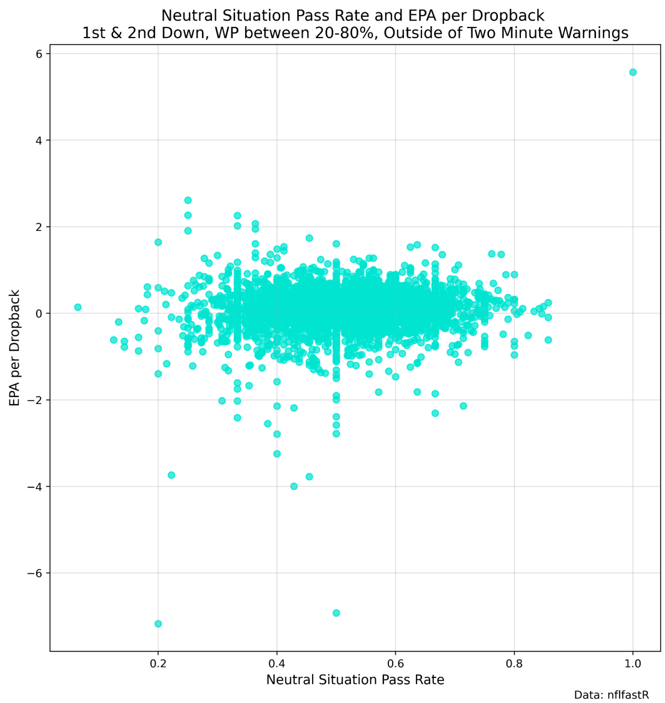
- play_id: A numeric play id that, when used with game_id and drive, provides the unique identifier for a single play

- game_id: A ten digit identifier for a NFL game. The unique play of a game is what we are analyzing, so we have 180,050 items.
- posteam: String abbreviation for the team with possession.
- Epa: Expected points added (EPA) by the team in possession for the given play.
- pass: Binary indicator if the play was a pass play (sacks and scrambles included).
- Half_minutes_remaining: We created this attribute based on the original data's half_seconds_remaining. We agreed minutes is a much better indicator of time than seconds.
- down: The down for the given play.
- xpass: Probability of dropback scaled from 0 to 1.
- new_play_type: We created this attribute by combining the pass length column and the play type column to make categories of deep pass, short pass, and run.
- score_differential_buckets: We created this attribute based on the original data's score_differential. We agreed to group them based on values of 7, as that is the value of a touchdown + extra point.
- ydstogo_buckets: We created this attribute based on the original data's ydstogo. Discussed later in our design evolution section, we agreed categories of yards-to-go would be a better way to show distance.

Exploratory Data Analysis



Here we explored how EPA changes with down. As you can see, there seems to be a pattern when the distance of yards-to-go is above 15, the EPA (ie our success rate measure) is negative. We expect this to continue to be a pattern in our dashboard.



Above we explored the relationship between EPA and a neutral situation pass rate. We used this design idea of a large scatterplot with a high density of points.

posteam	epa
KC	0.150450
GB	0.069273
LAC	0.054515
NO	0.048978
BAL	0.041380
TB	0.039891
NE	0.039159
DAL	0.029954
LA	0.029702
SEA	0.029374
PHI	0.028503
SF	0.021640
BUF	0.015400
TEN	0.015122
ATL	0.013276
MIN	0.013207
PIT	0.002015
LV	-0.007462
IND	-0.011741
CIN	-0.017176

We created a table to explore how each team ranked in regards to their EPA. Interesting to note that Kansas City has an EPA of almost 2.5 times greater than the next best team. We'd expect the teams with a positive EPA to win more game and be successful in more plays compared to teams with a negative EPA.

Design Evolution

Down and distance visual:

First in Tableau, we created a few different visualizations to show the relationship between down and distance, specifically how it relates to a team passing the ball or rushing the ball.

Pass Rate by Down & Distance

Ydstogo	Down 1	Down 2	Down 3	Down 4
1	26.4%	30.0%	26.7%	28.9%
2	32.9%	39.3%	66.1%	78.0%
3	37.5%	44.4%	81.9%	94.4%
4	38.7%	48.5%	89.4%	93.3%
5	43.4%	54.6%	91.9%	96.0%
6	38.6%	60.5%	93.4%	95.1%
7	39.4%	66.1%	94.0%	94.3%
8	38.9%	68.0%	94.5%	91.9%
9	45.2%	73.1%	94.4%	89.7%
10	51.4%	68.2%	95.2%	96.8%
11	47.6%	78.7%	94.2%	97.3%
12	54.8%	78.7%	92.9%	87.9%
13	56.5%	80.8%	92.8%	96.8%
14	58.5%	83.2%	90.7%	93.1%
15	67.9%	84.8%	94.0%	100.0%
16	72.5%	85.1%	91.2%	100.0%
17	73.8%	81.1%	89.5%	90.0%
18	78.9%	80.3%	83.7%	95.0%
19	72.1%	78.2%	83.7%	90.9%
20	76.5%	82.7%	87.9%	100.0%
21	60.0%	81.8%	86.0%	100.0%
22	50.0%	75.2%	83.2%	88.9%
23	77.8%	81.8%	71.6%	100.0%
24	28.6%	75.0%	73.7%	100.0%
25	82.0%	78.4%	90.2%	50.0%
26	50.0%	83.8%	66.7%	100.0%

Rush Rate by Down & Distance

Ydstogo	Down 1	Down 2	Down 3	Down 4
1	73.6%	70.0%	73.3%	71.1%
2	67.1%	60.7%	33.9%	22.0%
3	62.5%	55.6%	18.1%	5.6%
4	61.3%	51.5%	10.6%	6.7%
5	56.6%	45.4%	8.1%	4.0%
6	61.4%	39.5%	6.6%	4.9%
7	60.6%	33.9%	6.0%	5.7%
8	61.1%	32.0%	5.5%	8.1%
9	54.8%	26.9%	5.6%	10.3%
10	48.6%	31.8%	4.8%	3.2%
11	52.4%	21.3%	5.8%	2.7%
12	45.2%	21.3%	7.1%	12.1%
13	43.5%	19.2%	7.2%	3.2%
14	41.5%	16.8%	9.3%	6.9%
15	32.1%	15.2%	6.0%	0.0%
16	27.5%	14.9%	8.8%	0.0%
17	26.2%	18.9%	10.5%	10.0%
18	21.1%	19.7%	16.3%	5.0%
19	27.9%	21.8%	16.3%	9.1%
20	23.5%	17.3%	12.1%	0.0%
21	40.0%	18.2%	14.0%	0.0%
22	50.0%	24.8%	16.8%	11.1%
23	22.2%	18.2%	28.4%	0.0%
24	71.4%	25.0%	26.3%	0.0%
25	18.0%	21.6%	9.8%	50.0%
26	50.0%	16.2%	33.3%	0.0%

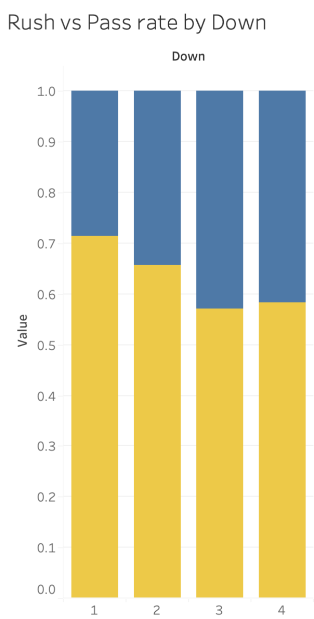
These two heat maps use color to show the percentage of pass rates (left) and rush rates (right) by down and yards to go (distance). The vertical position channel is used to represent the distance to first down (aka yards-to-go) and the horizontal position channel is used to represent down. We realized these two visualizations are just the opposite of each other because we filtered our data to only be looking at plays that are pass or rush. Including both graphs in our final dashboard would be redundant, so we decided we would just show the pass rates graph.

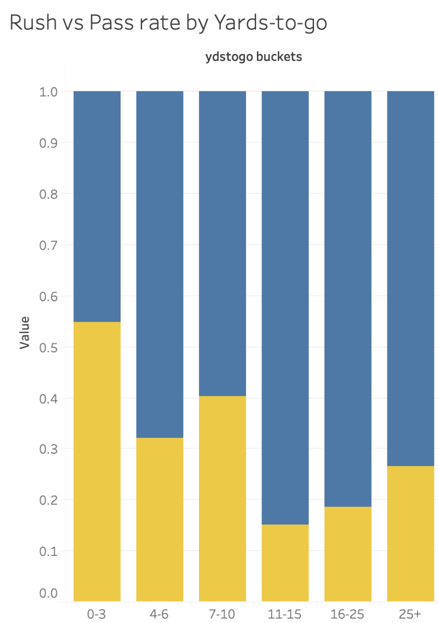
Additionally, the visualization takes up a lot of vertical space and is a lot to digest since there are so many rows. We thought we could improve this by grouping yards to go into different buckets.

Pass Rate by Yards-to-go categories

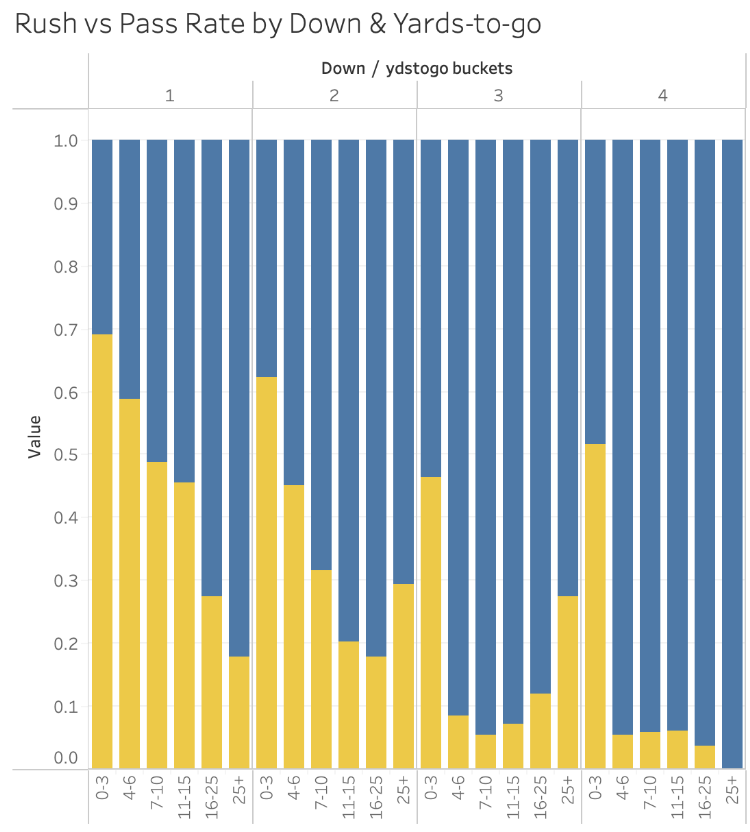
	Down			
ydstogo buckets	1	2	3	4
0-3	30.9%	37.7%	53.7%	48.5%
4-6	41.1%	55.0%	91.5%	94.6%
7-10	51.2%	68.5%	94.6%	94.1%
11-15	54.4%	79.7%	92.9%	93.8%
16-25	72.6%	82.2%	88.1%	96.3%
25+	82.1%	70.5%	72.6%	100.0%

We liked the idea of grouping yards to go, but thought this visualization might be too broad of categories. For example, 4th down and 1 yard to go has a 65% difference to 4th down and 3 yards to go, yet they are grouped into the same category in this visualization.

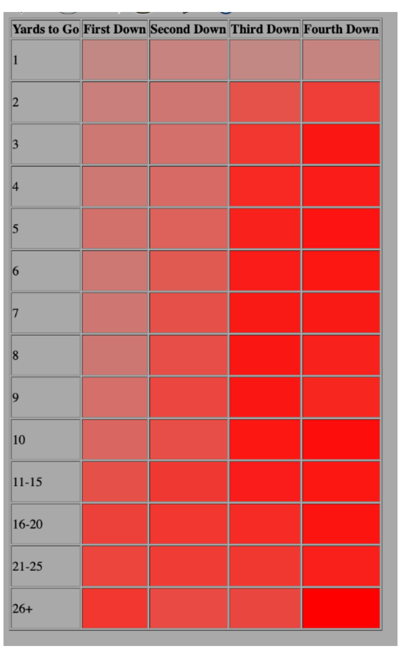




Above are more visualizations we created. The stacked bar charts show rush percentage in the color yellow and pass percentage in the color blue (this further shows how pass % and rush % are opposites of each other). This visualization works well, since we are only looking at pass and rush, their percentages add up to 100%. We liked how these visuals showed the data well, but wanted to show how BOTH down and distance affect the pass/rush rates. We had the idea of merging these two charts into one chart and below is what that looked like.

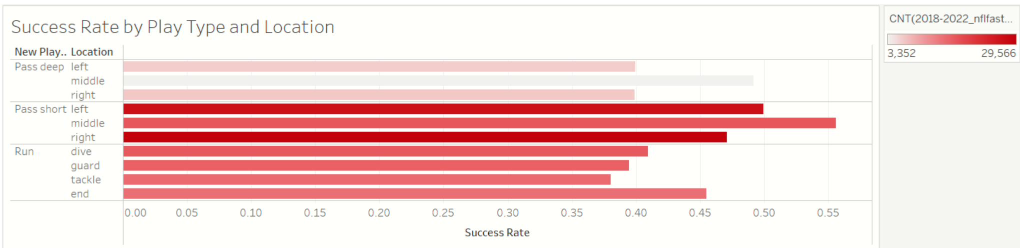


This stacked bar chart has each down separated into yards-to-go categories. However, if we wanted to create more yards-to-go buckets since we felt they were too broad, it would create even more bars in this visual, making it harder to digest

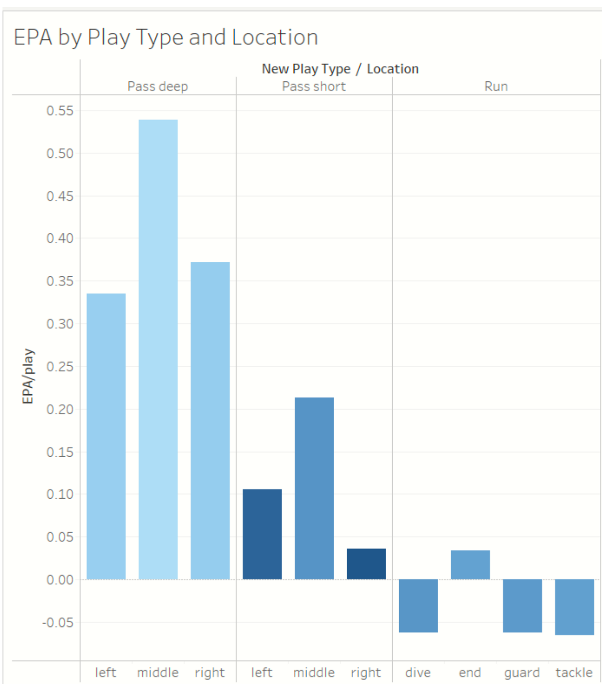


For our prototype, we decided to go with the heat map with yards-to-go grouped into broader categories. We felt the heat map did the best job of visualizing the questions posed. We came up with better yards-to-go categories of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-15, 16-20, 21-25, 26+. Since the yards 1-10 are the most common, we felt they should get their own category. More than 10 yards is less common, so we grouped 11-25 into three groups of 5 yards. The highest yardage was 43, but many points between 26 and 43 had no data available or were very rare. So we decided to group any yardage of 26 and above into the final group.

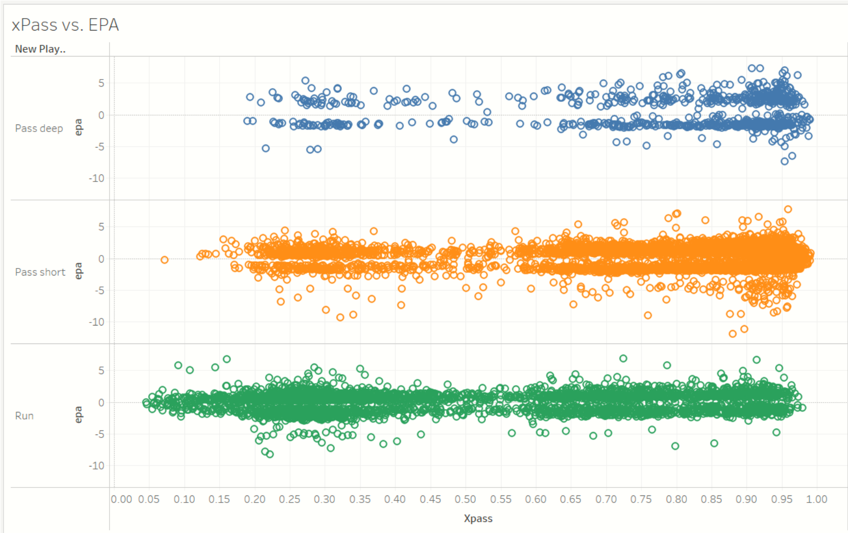
Success rates visual:
First in Tableau, we created a few different visualizations to find an interesting relationship regarding success rate (and how we wanted to measure that).



This grouped bar chart uses line marks to represent the play subtypes. The horizontal position channel is used to represent the success rate attribute, and the luminance channel is used to represent the frequency of each play type.

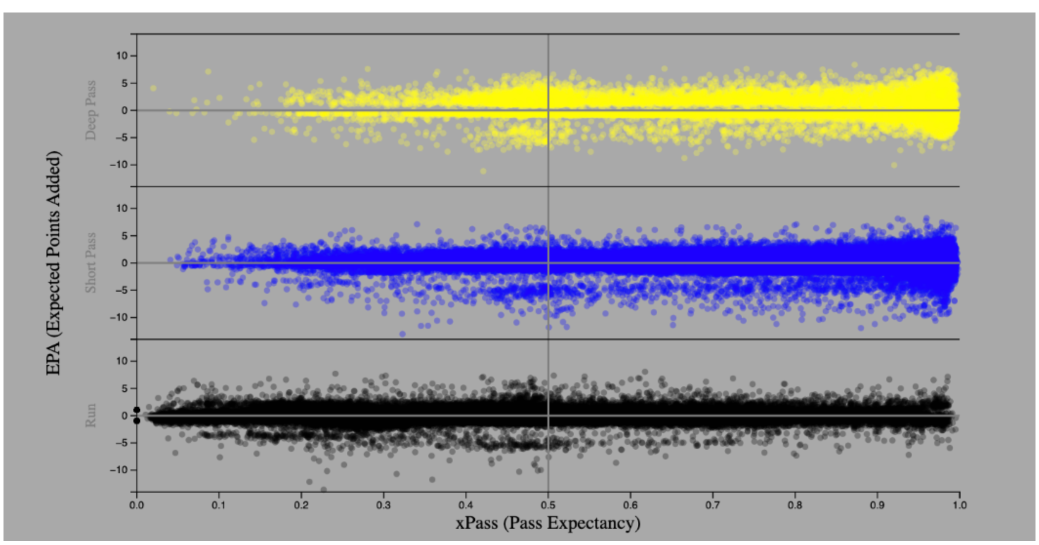


This grouped bar chart uses line marks to represent the play subtypes. The vertical position channel is used to represent the EPA/play attribute, and the luminance channel is used to represent the frequency of each play type. We decided to use the Success Rate attribute instead of EPA/play as EPA/play is an explosiveness metric, and is thus always biased towards deep passes. Success rate was more encompassing of efficiency, and revealed more interesting trends.

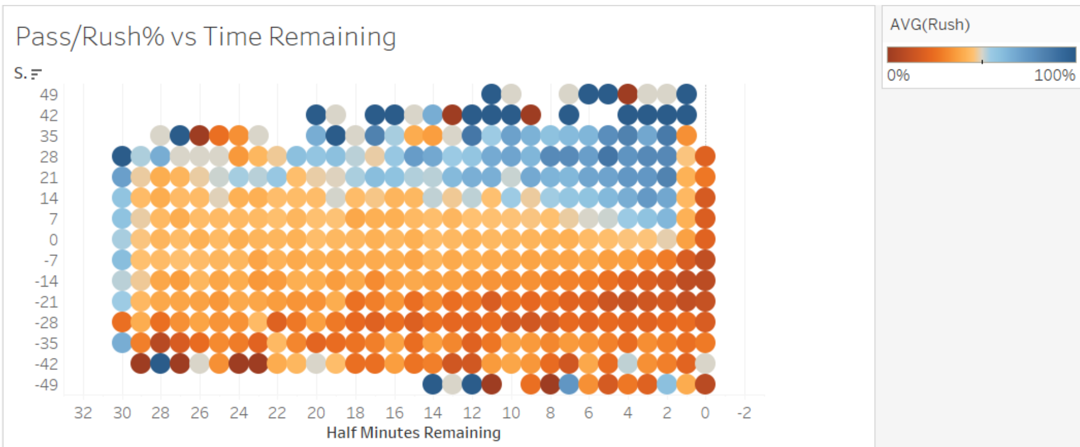


This scatterplot uses points as marks to represent individual plays. The color channel and vertical position channel are representing the play type attribute. The vertical position channel is also representing the EPA of a play. The horizontal position channel is representing the xPass attribute, which is the expectation that a team will pass given a variety of characteristics about the play.

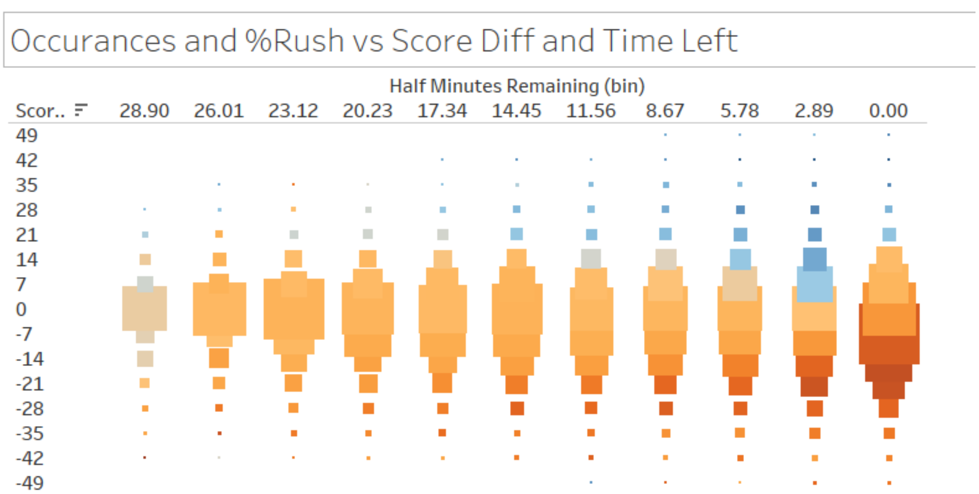
In our proposal, we opted to go with the second visualization. We thought it best described the data in a concise way. But when considering that our final dashboard would include filtering, we wanted a visualization with high granularity. We thought there were too many marks in the third scatterplot that it did not reveal many trends. However, we decided this busy scatter plot could be filtered by our other visualizations and show the trends we were hoping for. So, for our prototype, we decided to go with the scatterplot.



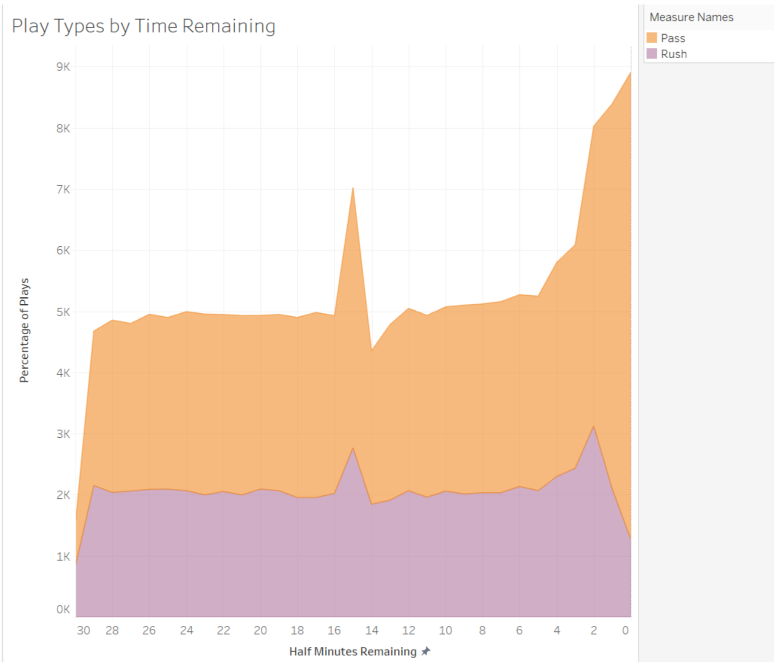
Clock Time and Score Differential visual:
First in Tableau, we created a few different visualizations to show the relationship between time remaining and score differential, specifically how it relates to a team passing the ball or rushing the ball.



In this heatmap, the odds of a rushing play or a passing play are placed on a divergent color scheme, while the time remaining in the half and the score differential of the game are shown as horizontal and vertical position attributes, respectively. One potential improvement would be to make it more clear that the score differential is a divergent axis – note that the y-axis has both positive and negative values.

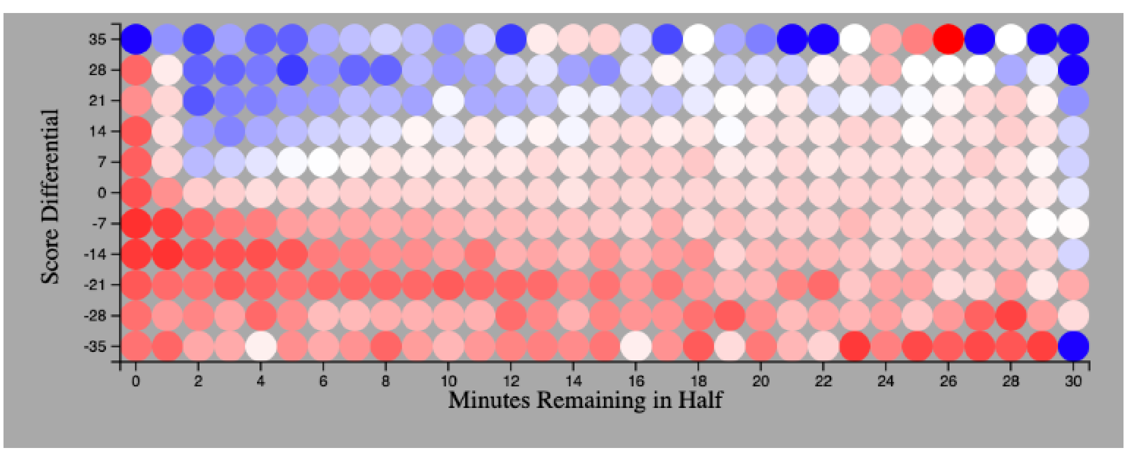


This graph is another version of the previous heat map, but each block has a size attribute corresponding to the frequency of each play. This does give some interesting new information, but the overlap between the squares could mislead the audience about the message. Furthermore, the very infrequent scenarios are almost invisible. The additional information displayed on this graph was not worth those tradeoffs, especially since the frequency of each scenario isn't directly tied to our problem statement.

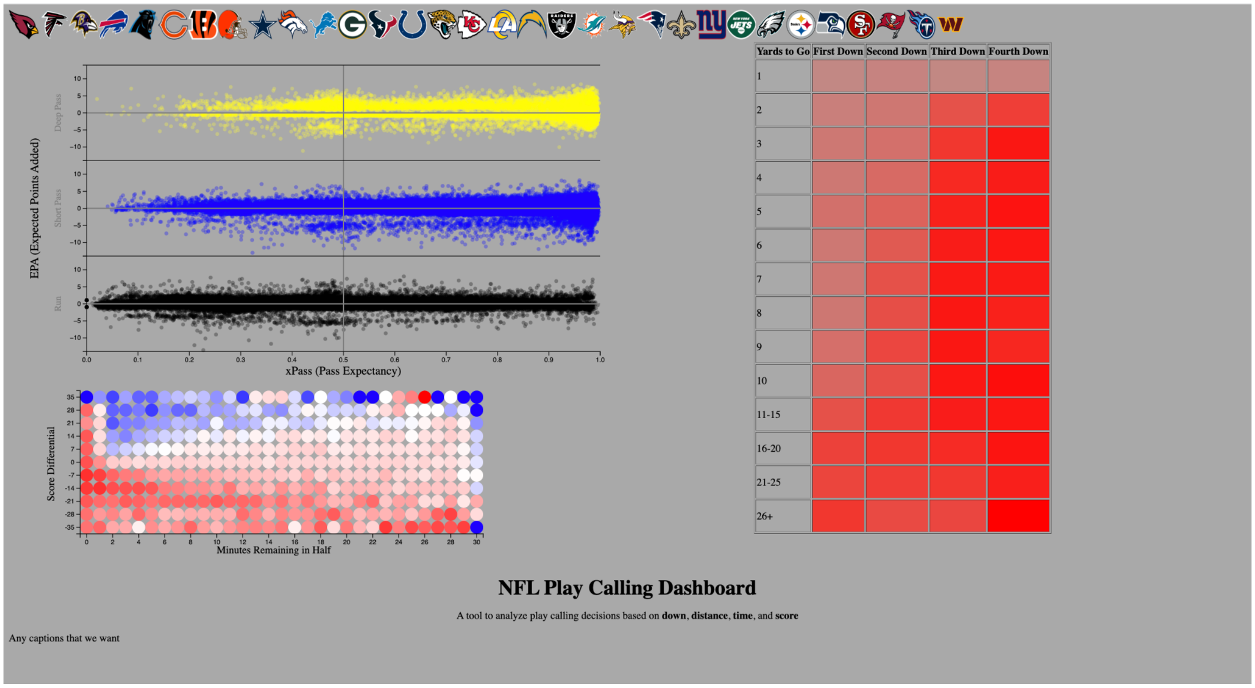


This simple area plot shows the total number of pass and rush plays called at each point in a half across our dataset. We also graphed it as a percentage, which was likely clearer for our stated objectives, but the variation in total number of plays over the course of the half was interesting enough that it seemed worth including – note how halfway through the half (at the end of the first or third quarters) and at the end of the half (at the end of the second or fourth quarters), the total number of plays increases.

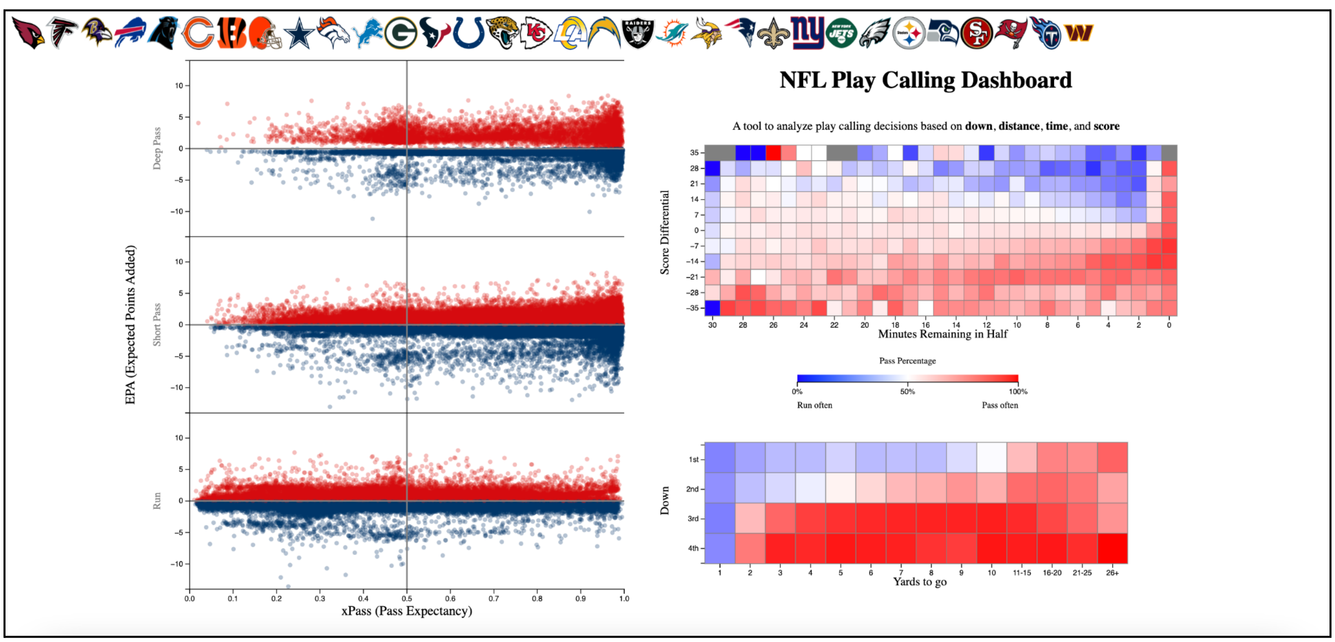
For our prototype, we decided to go with the dotted heat map which allows us to see the relationship between run/pass rates, score differential, and time remaining in the half.



Prototype to final design:



Above was our initial prototype. We included our 3 chosen visualizations discussed above, as well as NFL team logos across the top of the screen. We had to make improvements to the overall design based on feedback we received. Below is our final dashboard.

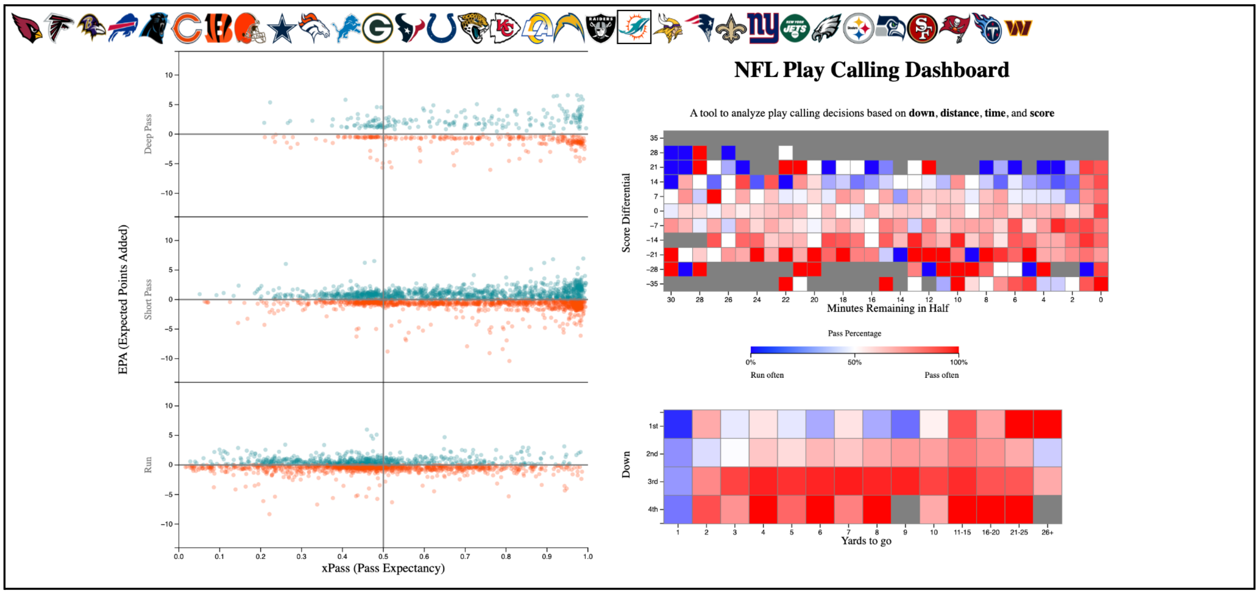


Improvements we made and why:

- We wanted to stay away from background colors and simply use either black or white. We didn't want the background color distracting from our visualizations.
- To better use the color channel, we thought it would be helpful to make the scatter plot change color to use the team colors of the NFL team selected. (We chose red and dark blue as the default when no teams are selected because those are the NFL's colors). We realized it is not necessary to have three different colors for the three sections of the scatterplot but rather use color to differentiate between positive and negative values in each subsection.
- To better use the color channel, we changed the down/yards-to-go heatmap to a diverging color scale of blue, white, red to match the score differential/minutes remaining heat map. This better shows how the pass rate and rush rate are both intrinsically shown in the visualizations, ie a low pass rate means a high rush rate and a high pass rate means a low rush rate. We also needed to add a legend to explain to the user what these colors mean – blue indicates a low pass rate (run often), white indicates a 50% pass rate (50% rush rate), and red indicates a high pass rate (pass often).
- For the scatterplot, we wanted to add the channel of opacity so that the user can see where points are more dense.
- From a layout perspective, we wanted to make the scatter plot bigger since it has so many points. We wanted to move the title to the top of the screen rather than at the bottom. We also wanted to put the two heatmaps in the same orientation and same style. We decided to put both of them horizontally to better use the space available. We decided to use squares instead of circles to avoid the background color peeking through the spaces between the circles. Then, we added a gray background behind the heatmaps, as some sections do not exist within our dataset (especially when filtering is enabled). We did not want the user to be confused between white squares from a 50% pass rate and non-existant squares that seemed white because the dashboard background color is white.

Implementation

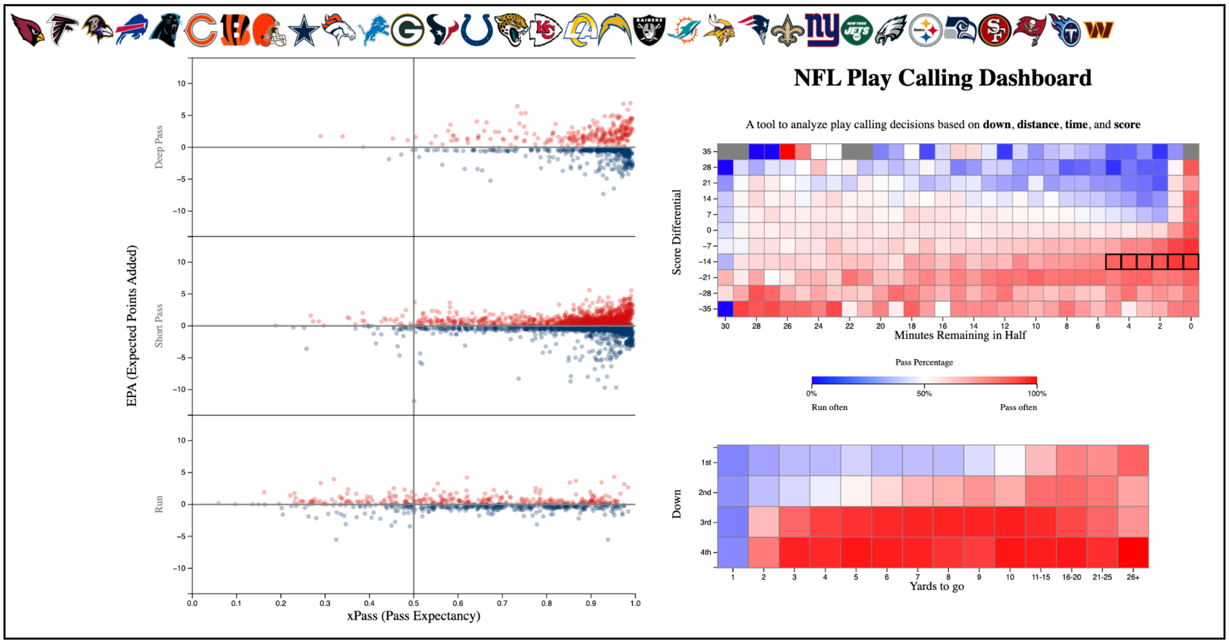
<u>Filtering by team:</u>



If you mouse over a logo, a black border will appear around it and the cursor will change to a hand pointer. This is to show that it is clickable. When you click on a logo, it filters all three visualizations to only show data that is applicable to that selected team. You will notice that there will be less points in the scatterplot. You will notice that many squares will gray out in the heatmaps (showing those scenarios do not exist) and the squares remaining might change hue. Above, you can see that the Miami Dolphins logo is selected.

To deselect a team, you can click the team's logo again. This will bring back all visualizations to include data from all teams. Alternatively, when one team is selected, you can simply click on a different team logo to switch directly to that team's data.

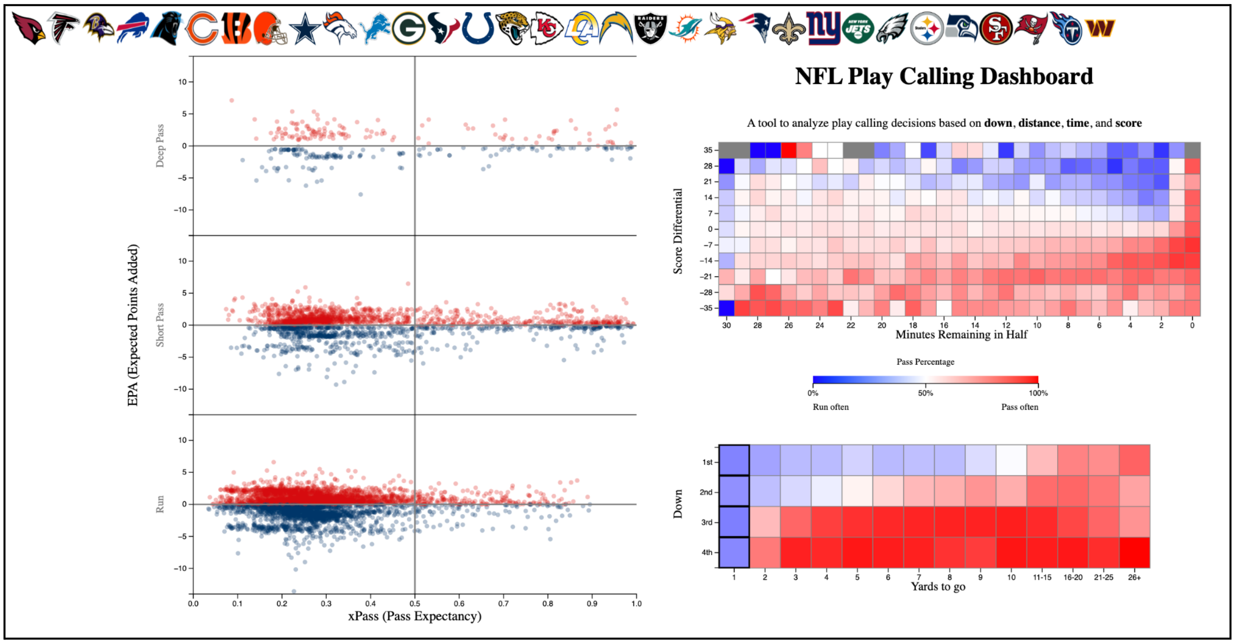
Filtering scatterplot by score differential/minutes remaining:



If you mouse over a square within the score differential/minutes remaining heatmap, a black border will appear around it and the cursor will change to a hand pointer. This is to show that it is clickable. When you click on a square, it filters the scatterplot to only show data that is applicable to the square clicked in the heatmap. You can click multiple squares. You will notice that there will be less points in the scatterplot than the unfiltered dataset, and more points will appear the more squares you click. Above, you can see that no NFL team is selected (so it is showing data from all teams) and 6 squares are selected in the top heatmap. A trend you can see here is that when a team is down by 14 points with 5 minutes or less remaining of the half, they pass the ball more often (squares selected are red, more points in the short pass & deep pass scatterplots) and in the cases that they run, they were more expected to pass (majority of points in the run scatterplot fall in the right section of the scatterplot showing high pass expectancy).

To deselect a square, you can simply click the square again or refresh the page.

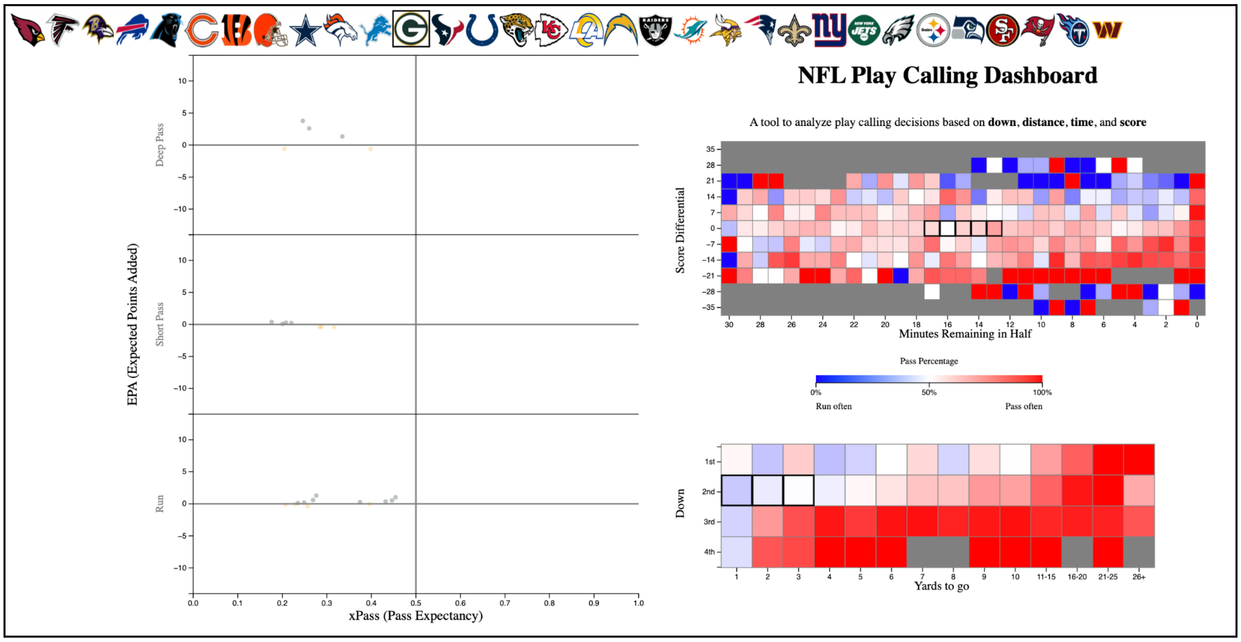
Filtering scatterplot by down/yards to go:



If you mouse over a square within the down/yards-to-go heatmap, a black border will appear around it and the cursor will change to a hand pointer. This is to show that it is clickable. When you click on a square, it filters the scatterplot to only show data that is applicable to the square clicked in the heatmap. You can click multiple squares. You will notice that there will be less points in the scatterplot than the unfiltered dataset, and more points will appear the more squares you click. Above, you can see that no NFL team is selected (so it is showing data from all teams) and 4 squares are selected in the bottom heatmap. A trend you can see here is that when a team has only 1 yard to go, they run the ball more often (squares selected are blue, more points in the run scatterplot) and in the cases that they pass, they were more expected to run (majority of points in the pass scatterplots fall in the left section of the scatterplot showing low pass expectancy).

To deselect a square, you can simply click the square again or refresh the page.

Filtering scatterplot by everything:



You can also filter by both heatmaps and by team. Above, you can see Green Bay is selected, 0 score differential with 13-17 minutes remaining in the half is selected in the top heatmap and 2nd down and 1-3 yards-to-go is selected in the bottom heatmap. Under these conditions, you can see in the scatterplot that teams are always expected to run, although there are few instances of these scenarios shown by the few number of dots.

Evaluation

What we learned:

Through this project, we were able to identify multiple trends within overall playcalling tendencies in certain scenarios. For example, we saw that in short yardage situations, teams are of course more likely to run, however, this becomes even more true on second downs, where teams run much more frequently, and are more successful in running.

On top of this, we were able to learn more about the tendencies of individual teams. The distributions of overall playcalls for most teams were relatively similar, however, when looking at specific scenarios, we saw many differences, both in tendencies and successfulness. One example of this that we identified was between the Patriots and Chiefs. In 3rd down and 10+ yards to go scenarios, the Patriots often threw short passes and ran the ball. This very infrequently resulted in a positive play for them. On the other hand, the Chiefs did not run the ball in those scenarios unless they had a high xPass value, likely meaning that they were leading very heavily. Excluding those plays, the Chiefs only threw passes, and threw deep often. This led to them having more successful plays and even some explosive plays with very high EPA values.

Through analysis and examples such as these, we are able to explore the answers to our questions. With thorough analysis, we believe that an NFL playcaller could use this dashboard to identify trends of their own team and of opposing teams. They could identify both what types of plays are most often called by particular teams in different scenarios, as well as what plays most often result in both successful and explosive plays, both within those scenarios and as a whole.

What we could improve:

We could further improve our dashboard by being able to more easily compare teams in similar situations. In our final dashboard, when you click on a team logo to filter by that team, it resets all the data in the visualizations. As we cannot rely heavily on our short term memory, it would be extremely difficult to compare the Cowboys on first down vs the Eagles on first down, for example. It would be an improvement to click on the team and other filters you want, then choose a different team but keep all the other same filters. You could then go back and forth to see this same situation for many different teams. However, since comparison between teams is not the overall goal of our project, we felt this was not a necessary feature for now.

Another way we could improve our dashboard is by enabling the user to select an entire row or column in the heatmaps instead of having to click each square individually. We feel it would be insightful to filter by first down, for example, to see how a team performs in all scenarios they have achieved first down. Similarly, it would be insightful to filter by 5 minutes remaining, for example, to see how a team performs in all those cases. This is possible to do within our current dashboard but you would have to individually select each square in the row or column you are interested in, which is inconvenient to the user.

We could further improve our dashboard by filtering the heatmaps by each other. Right now, when you click on say 1st down and 10 yards-to-go square, it does not update the colors/squares in the score differential/minutes remaining heatmap. Under certain selections, no points may appear since no plays with those characteristics exist. Ideally, if these scenarios didn't exist, there shouldn't be the option to click on them.