Project 2 Final Paper | NFL Offensive Play Analysis 2009 - 2017

Project Team | Abhishek Shetty, Bikram Khaira, Tyler Gustafson
Github Repository | https://github.com/UC-Berkeley-I-School/Project2_Khaira_Shetty_Gustafson

Data Structure

Primary Dataset | NFL Play by Play Data 2009 - 2017
This dataset covers all regular season plays for every NFL game between 2009 and 2017. The dataset has over 100 columns describing the circumstances around each play.

Introduction

As we get ready to deep dive into the evolving world of the NFL we want to take a second to set the stage for what we're about to explore. We've got our hands on a ton of data! We have all of the NFL offensive play data from 2009 to 2017. This data has 8 years of NFL games with every play listed out. We have over 100 columns that capture everything that is going on during these plays.

Abhishek Shetty, Bikram Khaira, and Tyler Gustafson will be sifting through all of this data to uncover different trends and strategies that went on during this time in the NFL. Before we jump into the analysis we needed to do some sanity checking on the data. For example, the Rams and the Chargers both moved cities and we need to make sure we get these details correct before diving into the data.

We are going to be looking into not only who wins these games, but more complex questions such as:

- How do play types change with field position?
- Does the pressure of the clock winding down affect decision-making?
- Is there really such a thing as home-field advantage?
- Are teams run heavy or pass heavy?
- Are they correctly picking the right plays?

With the stage set let's dive into what makes the NFL so exciting and why some fans are obsessed with this thrilling sport!

Sanity Checking

In order to sanity check the data, we checked the number of games within the data – there were a total of 2304 unique games and 9 seasons within the dataset, with 256 games for each of the nine seasons. Next, we checked the number of times each team appears within the HomeTeam and AwayTeam columns. We found that in each season, there were a total of 32 teams listed in the HomeTeam and AwayTeam columns, matching our expectations. Furthermore, each team was listed alongside 8 games in the HomeTeam column and alongside 8 games in the AwayTeam column, also matching our expectations from the general NFL schedule.

A minor complication within the data was that two franchises had moved cities during the time period between 2009 and 2017 seasons. The St Louis Rams moved to Los Angeles and became the Los Angeles Rams prior to the 2016 season. The San Diego Chargers also moved to Los Angeles and became the Los Angeles Chargers prior to the 2017 season. Looking at the data we see St Louis appearing till the 2015

season, and the Los Angeles Rams appearing after that. We also see the San Diego Chargers appearing till the 2016 season and the Los Angeles Chargers appearing in the 2017 data.

Another wrinkle in the data is inconsistent naming for two teams. Jacksonville Jaguars are referred to as JAX and JAC and the Los Angeles Rams are referred to as LAR and LA. To fix this during the home and away team analysis, we replaced mentions of LA and JAX in the dataset with LAR and JAC respectively.

Data Cleaning

For our analysis we made several adjustments to the columns in order to make the data base more workable in our analysis. We focused on adding new classification columns such as Yardline Groupings and Quarter Time Categories. We also included some Calculated columns such as Total Points and Total Turnovers (Fumbles + Interceptions).

For the home and away data analysis, we only wanted high-level data, so the dataset was trimmed to include only key such as variables Date, GameID, Possession Team, Defensive Team, Play Type, Penalized Team, Interception, Fumble, Sack, Possession Team Score, Defensive Team Score, Home / Away Team and Season.

Added Key Columns

New Classification / Grouping Columns
- Quarter Time Category (Start, Middle, End)
- Yardline Groupings (Ten yard bins)
- Various Binary columns for charts (Run Location, Pass Location, etc.)

Calculated Columns
- Total Points
- Total Turnovers

Removed Columns

Precalculated Probability Columns
- Scoring Probabilities (Touchdown, Two Points, Field Goal, etc.)

Unused Statistic Columns
- Player Information (Receiver, Rusher, Blocking Player, etc.)
- Timeout stats

Initital Exploration

Before jumping into our three key analytical groups on specific offense trends, we wanted to understand in the data how the NFL has evolved on various offensive metrics during the seasons in the dataset. Some of the key metrics we were curious about were Avg Total Points per game, Total Turnovers, Penalties and Touchdowns per game. Here is a table that dives into this data a little bit more.

Season	Avg Total Points	Avg Total Turnovers	Avg Penalties	Avg Touchdowns	Avg Field Goals	Avg Yards Gained
2009	42.20	4.88	11.83	4.97	2.97	861.11
2010	43.17	4.85	12.09	5.04	3.18	865.26
2011	43.40	4.55	12.78	5.07	3.30	839.17
2012	44.48	4.51	12.56	5.21	3.35	843.21
2013	45.65	4.48	12.18	5.37	3.42	833.70
2014	44.16	4.42	13.22	5.15	3.29	824.98
2015	44.46	4.28	13.81	5.28	3.30	816.75
2016	44.69	4.11	13.43	5.19	3.38	801.25
2017	42.25	4.14	13.34	5.00	3.43	767.33

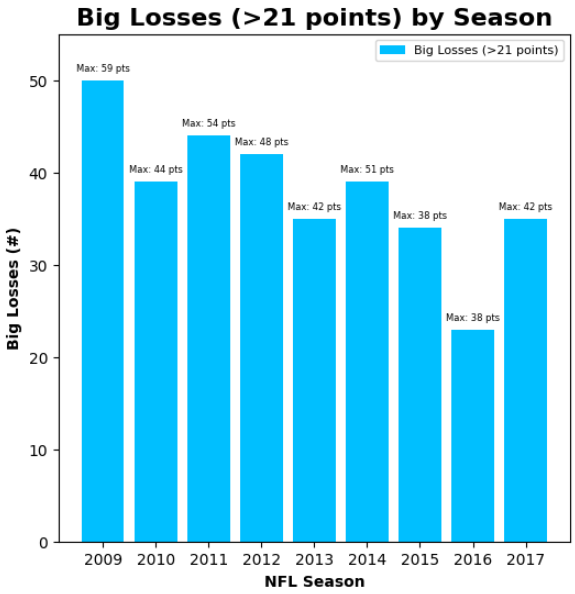
At first glance, this snapshot of the season averages almost appears that the NFL hasn't changed much:

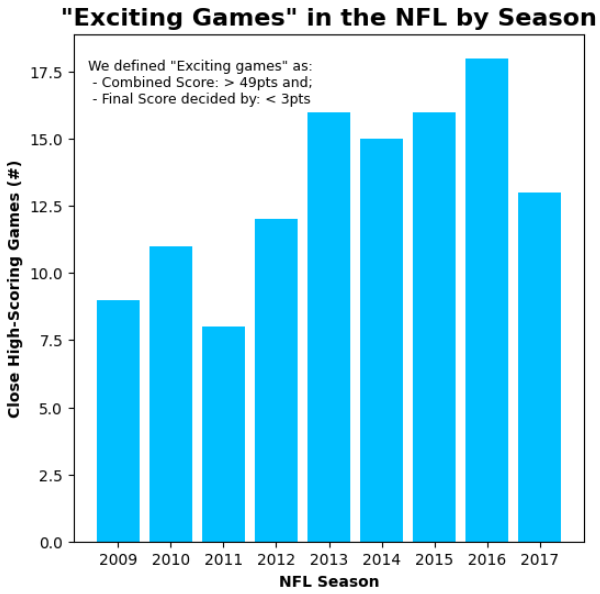
- Average Total Points are about the same at ~42 points
- Total yards actually go down
- And penalties are up!

Does this mean the NFL has gotten more boring or stale? No. First, the lower average total yards can be accounted for the rule change that occurred in 2016 where touchbacks start 5 yards closer (5 yards X 5 average number of TDS per game = an already 25 yards per game boost) and then if we account for the increased penalty yards we are at approximately the same number of yards gained a game.

Now if we look closer at those total points that have remained stagnant, we can see a couple of trends emerge. The following two graphs look at two things 1) the number of big losses (defined as losses by

more than 3 scores and 2) the number of Exciting Games (defined as combined score of > 49 pts (7 TDs) and a final score decided <3) there are some interesting patterns:



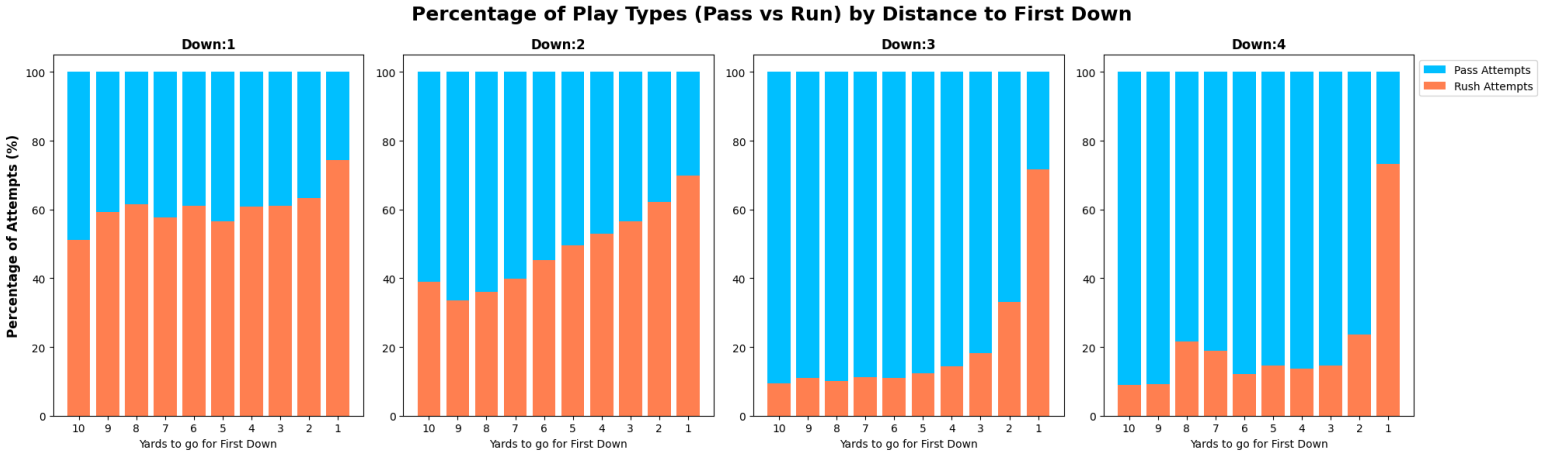


First we can see a clear pattern trending down of these less interesting blow out games (~15-20 games less a season) and then an uptick in high scoring close decided games. The takeaway? A more competitive interesting league to watch.

Now that we have an understanding of these dynamics we want to spend some more time digging into what these offenses look like and how different teams are executing their strategies.

Impact of General Factors on Offense

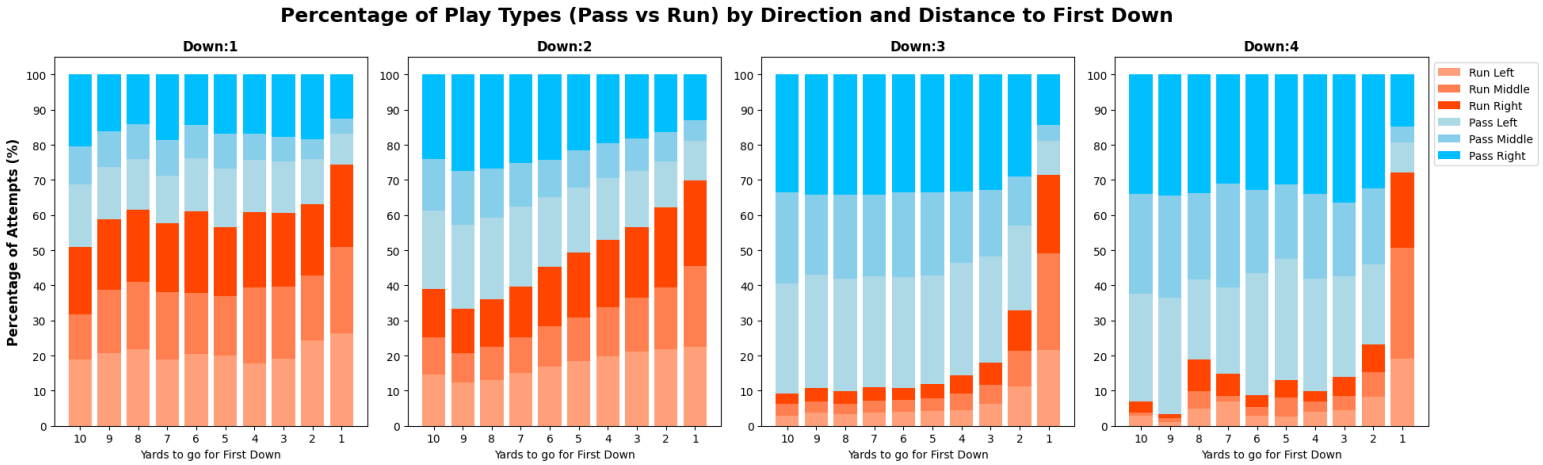
Understanding the distribution of what plays are typically executed and where can provide valuable insights to any coach on both the offensive and defensive sides of the ball. When exploring the influence of field position on offensive strategies, we created a detailed chart that categorizes play types based on the distance remaining for a 1st down separated by the specific down. This chart offers insightful data on how teams adapt their play-calling in varying field positions:



We can quickly see a hand full of insights by Down:

- **On 1st Down:** the data shows a balanced distribution between different play types, indicating a balanced strategy and that defenses should play conservatively
- **During 2nd Down:** a notable shift in play selection occurs if more than five yards were gained on 1st down, with a higher inclination towards running plays
- **On 3rd and 4th Downs:** the trend strongly favors passing plays, except in scenarios where only one yard is needed for a first down

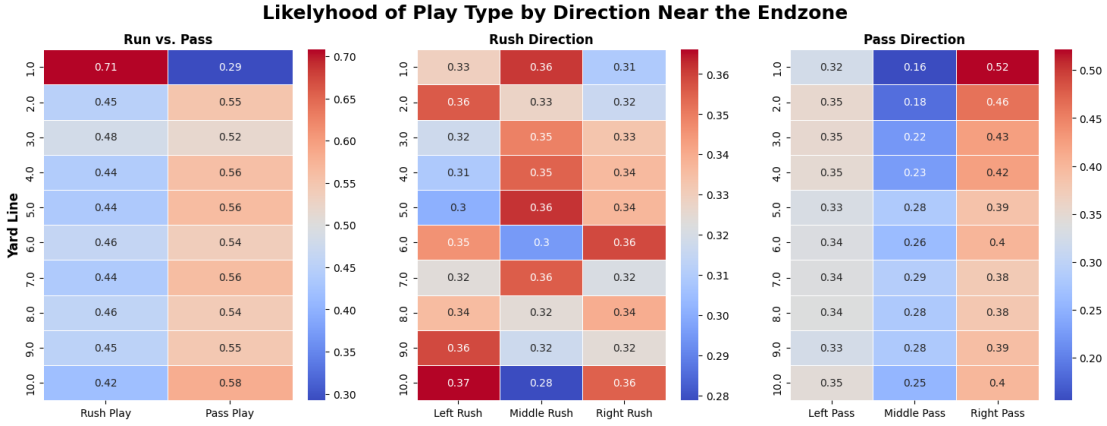
All of these insights can inform an offensive and defensive coach's strategy. We can also break this down even further by what direction the play is actually executed:



In addition to the insights we already highlighted, several more patterns emerge by play type:

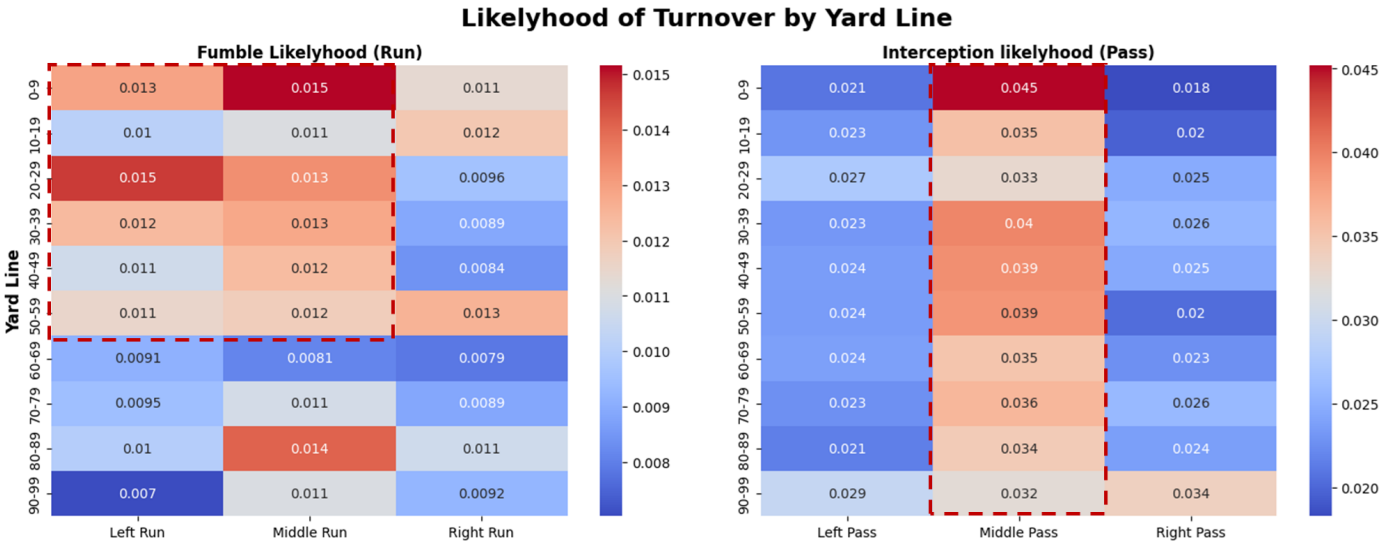
- **Pass Plays:** During 1st and 2nd downs, there's a tendency to steer clear of passes to the middle, with teams generally opting for passes to the left or right side of the field. On 3rd and 4th downs, though there's a slight increase in passes directed towards the middle. Finally, it appears the closer the quarterback is to the 1st down the more likely they are to throw a right sided pass.
- **Run Plays:** On 1st down, the direction of running plays is balanced, showing no strong preference for any side. However, in later downs, the direction for running plays varies more, influenced by the distance remaining to a first down. Left or right sided runs are preferred until only a yard remains.

These are all helpful insights as play moves down the field, but lets do a double click into the redzone and see what happens to play distribution when the team is only ten yards away from the endzone.



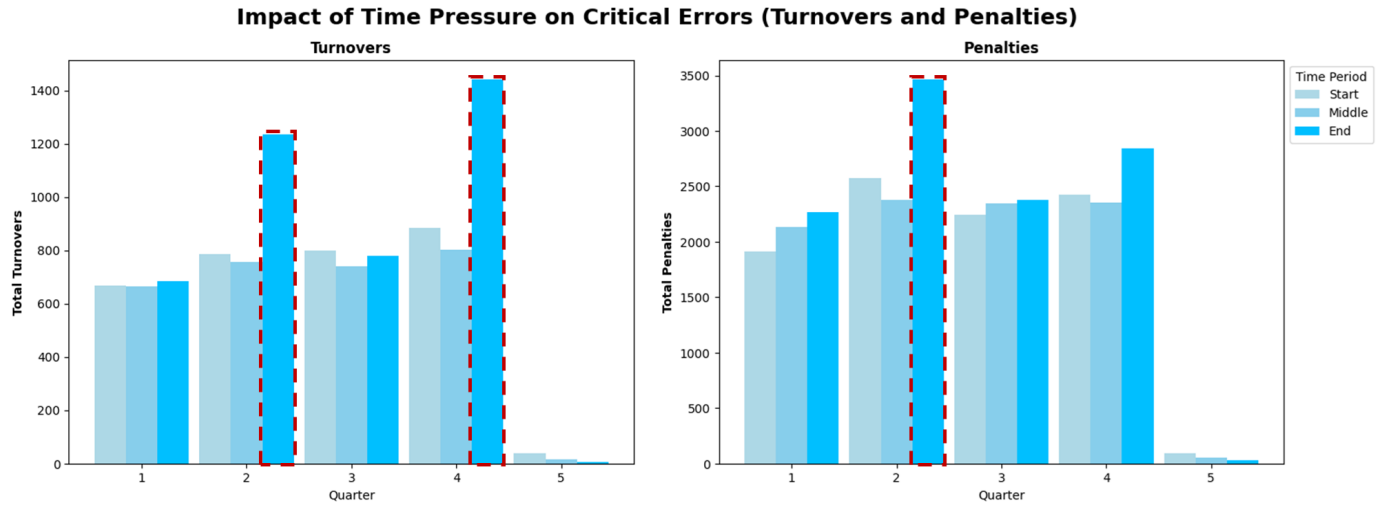
When we look at the chart above we can see that passing plays dominate near the end zone until the 1-yard line, where preferences shift. Right-sided passes are favored, while middle passes are rare.

Now that we have this understanding lets take a closer look at what areas of the field might be high risk areas for turnovers (fumbles and interceptions).



According to the data, running backs need to be extra cautious protecting the ball across mid field and running on the left side or middle. Quarterbacks, however, should be extra cautious about throwing the ball across the middle. One interesting link from the previous chart is we see the lowest interception rate is the right pass within the goal line which was also the highest play type from our heatmap on the previous slide so coaches are coaching to minimize mistakes near the endzone.

The final lens we want to understand is what is the impact of of time pressure on critical errors:

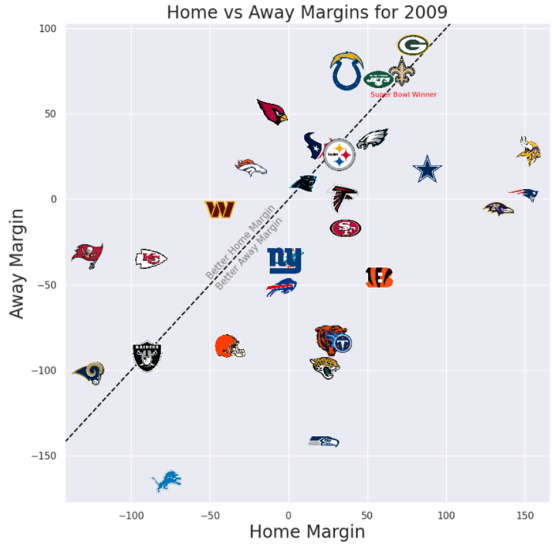


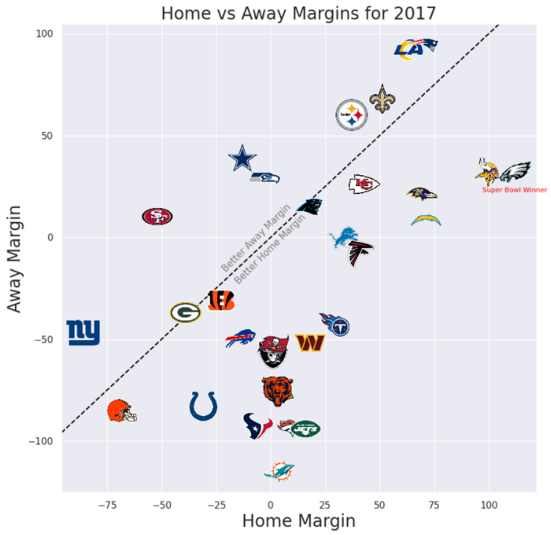
We see that time pressure increases turnovers in the late 2nd and 4th quarters and drives up 2nd-quarter penalties. This is a good indication that coaches should ensure their teams are extra

discipline going into the end of the first half. Now that we have looked at the general factors of time and field position we will take a look at another pressure that impacts teams… home field advantage.

Impact of Home Field Advantage

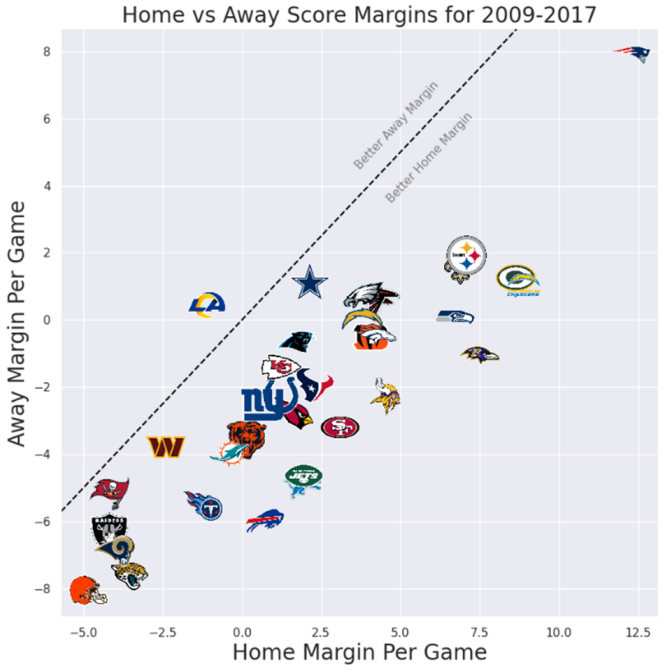
Is there a home field advantage in the NFL?





We wanted to check if teams have a home field advantage, and if so, what is the median advantage? In order to quantify this, we set out to measure the difference between the number of points a team scored and the number of points scored by their opponent. If a team won, this would be a positive number. If the team lost, this would be a negative number. This margin was added up for each team for each season at home and away, giving the net margin or difference in points scored for each team both home and away. Above are two charts showing what that looked like for 2009 and 2017. Part A of the Appendix contains a legend for team logos.

Looking at individual seasons, there was some home field advantage for most teams, but not all of them. In order to get a better sample size, we looked at all nine seasons of data that we had. Now, the Chargers and Rams changed cities between 2009 and 2017, so the total points scored for teams like the Los Angeles Rams and Los Angeles Chargers would be lower than for other teams. Therefore, in order to normalize the data, the home and away advantage was measured in terms of the net margin per game at home and away. The chart on the next page and the table of Part B of the Appendix show the result.



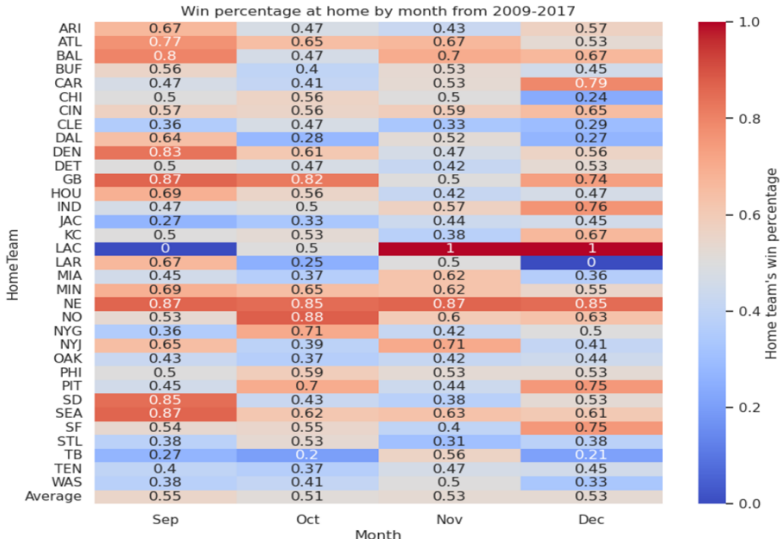
Takeaways:

1) Every team with a large enough sample

size had a home field advantage: The only team that was better away was the LA Rams, and this can be attributed to the fact that the Rams' data encompasses only one season. The median home advantage, per Part B of the Appendix, was 4.17 points per game.
2) The Patriots were very good: The New England Patriots were far and away the best team over this time period. It didn't matter if they were away or at home, they beat their opponents by huge margins.
3) No relation between team's quality and their home advantage: The Patriots only had the 14th best home advantage. One of the worst teams during this period, the Jaguars, had the 18th best advantage.

What was the impact of month/time of year on the home advantage?

When winter rolls around, you can hear NFL analysts talk about how teams like the Green Bay Packers and the New England Patriots benefit from cold weather at their home fields. But is this a real effect? We didn't have weather data for each game, so we had to go for a generalization: Does home field advantage vary by month? For this, we calculated the percentage of games each team won during each month. January was excluded due to the very small number of games played in that month. Here's a heatmap of our analysis:

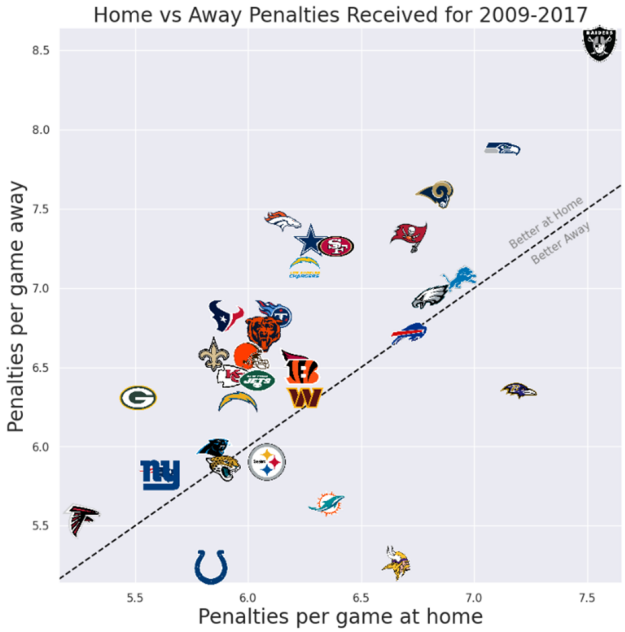


HomeTeam	Sep	Oct	Nov	Dec
ARI	0.67	0.47	0.43	0.57
ATL	0.77	0.65	0.67	0.53
BAL	0.8	0.47	0.7	0.67
BUF	0.56	0.4	0.53	0.45
CAR	0.47	0.41	0.53	0.79
CHI	0.5	0.56	0.5	0.24
CIN	0.57	0.56	0.59	0.65
CLE	0.36	0.47	0.33	0.29
DAL	0.64	0.28	0.52	0.27
DEN	0.83	0.61	0.47	0.56
DET	0.5	0.47	0.42	0.53
GB	0.87	0.82	0.5	0.74
HOU	0.69	0.56	0.42	0.47
IND	0.47	0.5	0.57	0.76
JAC	0.27	0.33	0.44	0.45
KC	0.5	0.53	0.38	0.67
LAC	0	0.5	1	1
LAR	0.67	0.25	0.5	0
MIA	0.45	0.37	0.62	0.36
MIN	0.69	0.65	0.62	0.55
NE	0.87	0.85	0.87	0.85
NO	0.53	0.88	0.6	0.63
NYG	0.36	0.71	0.42	0.5
NYJ	0.65	0.39	0.71	0.41
OAK	0.43	0.37	0.42	0.44
PHI	0.5	0.59	0.53	0.53
PIT	0.45	0.7	0.44	0.75
SD	0.85	0.43	0.38	0.53
SEA	0.87	0.62	0.63	0.61
SF	0.54	0.55	0.4	0.75
STL	0.38	0.53	0.31	0.38
TB	0.27	0.2	0.56	0.21
TEN	0.4	0.37	0.47	0.45
WAS	0.38	0.41	0.5	0.33
Average	0.55	0.51	0.53	0.53

Takeaways:
1) For the entire league, there is little variation in the home advantage by month.
2) The teams in colder climates also did not have a benefit from winter: Teams such as the Patriots, Packers, Minnesota Vikings and the Buffalo Bills did not see a higher win percentage at home in December compared to other months.

Are the referees biased in favor of home teams?

We looked at several different metrics such as interceptions thrown, sacks allowed and fumbles committed per game for teams home and away. The charts for these can be found in Part C of the Appendix. For these stats, there was some minor home field advantage, with around 60% of teams doing better at home. However, one statistic showed a clearer difference: penalties received. We calculated the average number of penalties each team received per game at both home and away games.



Takeaways:
1) There was a clear home field advantage for penalties received: The median difference between the number of penalties received per game on the road versus at home was 0.31.
2) Only six teams out of our sample of 34 received more penalties at home.
3) The results are notably different from other minute-level stats: One could attribute the difference between home and away games to teams simply playing better at home. However, this kind of clear advantage is not observed for stats like interceptions, fumbles and sacks (as seen in Appendix Part C).

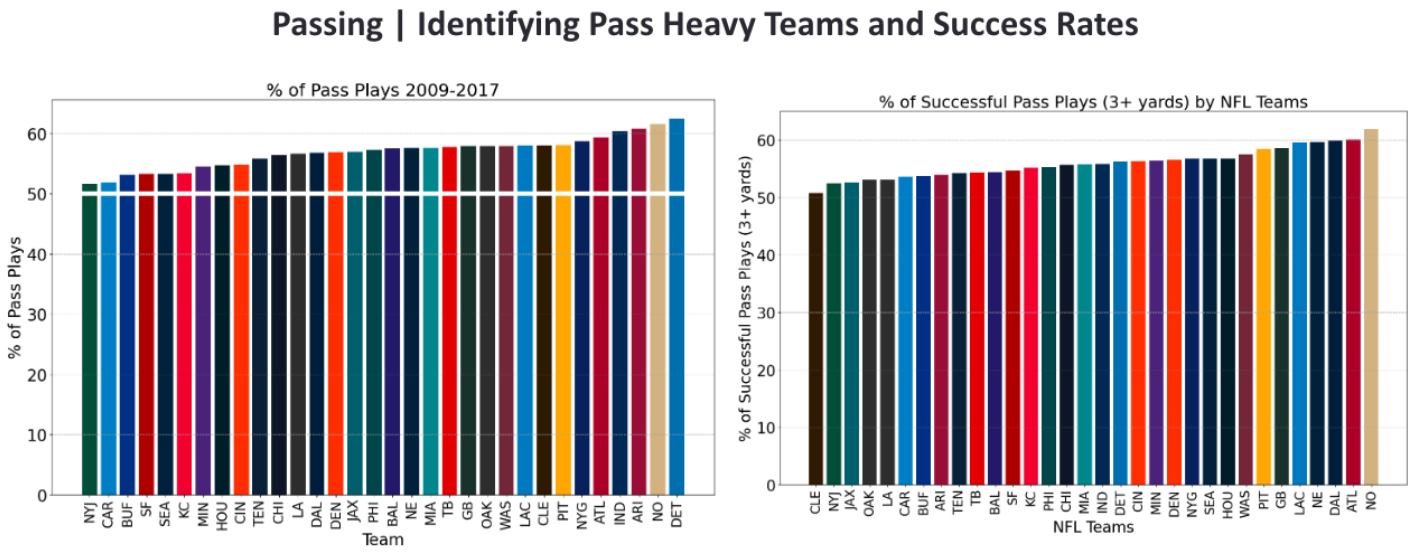
NFL Team Specific Strategies

Do teams favor the run or pass more?

We wanted to start out by looking at if teams pass or run the ball more often. This will help us find insights into offensive strategies for each team. After only selecting plays where teams ran or passed the ball we saw that every team passed the ball more than 50% from 2009-2017! This is surprising because running the ball is such a fundamental part of football.

We can also see that the New Orleans Saints passed the ball the most while also having the most successful pass plays. We attributed a successful pass play as a gain of 3+ yards since that would get you a 1st down if you executed it for three downs.

Looking into the team specifics for the Saints they were led by Drew Brees during this time period. He has the second most passing yards in NFL history! When we look at the other end of % of successful plays we can see the Cleveland Browns at a little over 50%. This can be attributed to them having 18 different starting quarterbacks from 2009-2017!

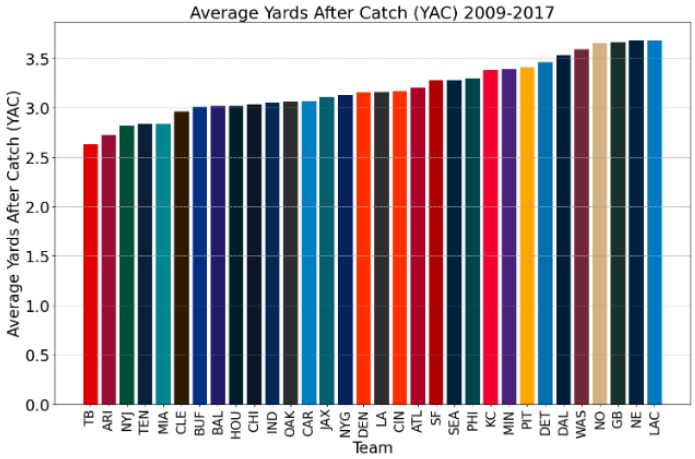


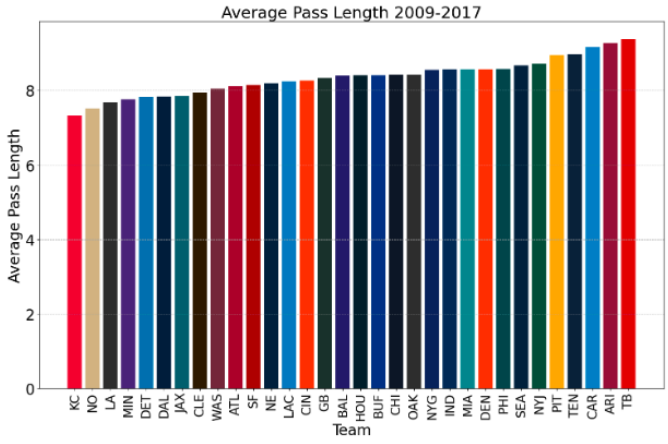
How do teams receivers rank in yards after catch (YAC)?

Since the league is so pass heavy, let's look at the average yards after catch (YAC) and the average pass lengths for each of these teams. In the YAC chart we can see the Tampa Bay Buccaneers and Arizona Cardinals have the lowest YAC and one of the highest teams with YAC are again the New Orleans Saints!

When looking at the average pass length chart we can see the Arizona Cardinals and Tampa Bay Buccaneers have the longest average pass length and the New Orleans Saints have the second lowest! This is indicating that the Cardinals and Buccaneers receivers would catch the ball further down the field rather than catching it shorter and running. The Saints tended to throw the ball short and have their receivers run to gain more yardage.

Passing | Identifying Teams with Dynamic Receivers



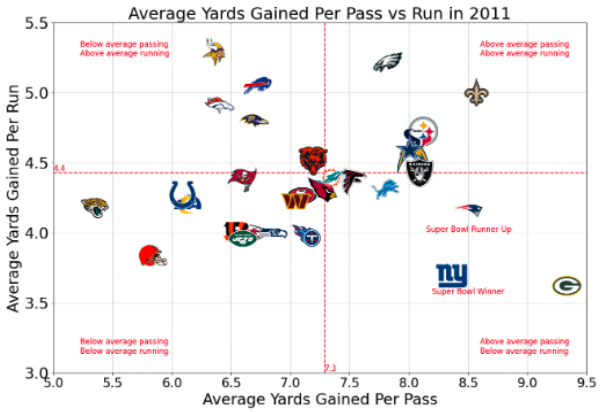


Do winners need a better passing or running attack?

Since running is still around 40% of a team's offense, let's see if the best teams have a strong running and passing attack! In the charts below we are comparing average yards gained per pass vs run for the years of 2009 and 2011.

We can see in 2009 the Super Bowl Finalists were better than league average in passing but the Colts were much below average in running the ball! In 2011 we can see that both teams were below average again at running but had elite passing attacks. The overall trend here is that if you want to make it far in the playoffs, having a great passing attack is necessary.

Pass & Run | Identifying Winning Teams Avg Yards Gained



Conclusion

By grasping the key aspects and risks in these areas that we have highlighted; a coach can create an effective offensive and defensive coaching strategy for the NFL.

A strong comprehensive NFL coaching strategy is..

1) Coaching for the general factor effects of:

- Play distribution by field position and down
- High interception rate in the center of field
- High fumble rate after mid field
- Dangers of time pressure on critical errors (Turnover and Penalties)

2) Cautious of homefield advantage risks when playing away:

- It's evident from the sample size of all seasons that the net margin for every team is higher at home than away. Over a long time frame of several years, there is a clear home advantage for teams in the NFL, but there is significant variation in the advantage from season
- The cold-weather teams don't have a stronger home field advantage in December compared to other months.
- There is a clear difference between how likely referees are to penalize the home team versus the away team. Overall, away teams are likely to be penalized more.

3) Planning for specific team strategies:

- The NFL is a pass heavy league with the Saints leading the way during 2009-2017
- There is an inverse relationship between passing length and YAC
- Having an elite running attack is not necessary to have success in the NFL

APPENDIX

Part A: Table showing teams, their abbreviations and their logos

Team	Abbr.	Logo	Team	Abbr.	Logo
Arizona Cardinals	ARI		Los Angeles Chargers	LAC	
Atlanta Falcons	ATL		Los Angeles Rams	LAR	
Baltimore Ravens	BAL		Miami Dolphins	MIA	
Buffalo Bills	BUF		Minnesota Vikings	MIN	
Carolina Panthers	CAR		New England Patriots	NE	
Chicago Bears	CHI		New Orleans Saints	NO	
Cincinnati Bengals	CIN		New York Giants	NYG	
Cleveland Browns	CLE		New York Jets	NYJ	
Dallas Cowboys	DAL		Philadelphia Eagles	PHI	
Denver Broncos	DEN		Pittsburgh Steelers	PIT	
Detroit Lions	DET		San Diego Chargers	SD	
Green Bay Packers	GB		San Francisco 49ers	SF	
Houston Texans	HOU		Seattle Seahawks	SEA	
Indianapolis Colts	IND		St Louis Rams	STL	
Jacksonville Jaguars	JAC		Tampa Bay Buccaneers	TB	
Kansas City Chiefs	KC		Tennessee Titans	TEN	
Las Vegas Raiders	LV		Washington Redskins/Football	WSH	

Part B: Table of Home vs Away Margins for each Team

	Team	Home Margin Per Game	Away Margin Per Game	Net Home Advantage Per Game
1	BAL	7.486111	-1.069444	8.555556
2	LAC	8.875000	1.000000	7.875000
3	GB	8.625000	1.263889	7.361111
4	DET	1.902778	-4.916667	6.819444
5	MIN	4.500000	-2.263889	6.763889
6	BUF	0.694444	-6.000000	6.694444
7	SEA	6.722222	0.069444	6.652778
8	NYJ	1.958333	-4.638889	6.597222
9	SF	3.097222	-3.208333	6.305556
10	NO	6.875000	1.625000	5.250000
11	PIT	7.083333	1.916667	5.166667
12	DEN	4.222222	-0.444444	4.666667
13	CIN	4.041667	-0.527778	4.569444
14	NE	12.291667	7.875000	4.416667
15	ARI	1.597222	-2.763889	4.361111
16	TEN	-1.236111	-5.541667	4.305556
17	HOU	2.277778	-1.958333	4.236111
18	JAC	-3.555556	-7.652778	4.097222
19	ATL	4.347222	0.486111	3.861111
20	SD	3.781250	0.015625	3.765625
21	CHI	0.138889	-3.583333	3.722222
22	MIA	-0.402778	-3.972222	3.569444
23	IND	1.208333	-2.250000	3.458333
24	CLE	-4.847222	-8.055556	3.208333
25	NYG	0.638889	-2.513889	3.152778
26	PHI	3.819444	0.722222	3.097222
27	STL	-4.035714	-6.892857	2.857143
28	KC	1.236111	-1.402778	2.638889
29	CAR	1.736111	-0.666667	2.402778
30	OAK	-4.194444	-6.319444	2.125000
31	WAS	-2.416667	-3.791667	1.375000
32	DAL	2.138889	1.138889	1.000000
33	TB	-4.194444	-5.166667	0.972222
34	LAR	-1.125000	0.437500	-1.562500

Part C: Charts of Home vs Away Sacks, Fumbles and Interceptions