Analysis on NFL Play-By-Play Data

Nathan Krieger, Caleb Moe

Our Dataset

- https://nflsavant.com/about.php
- Compiled data from the 2024 NFL season
- Created by Daren Willman

Our Dataset

```
## # A tibble: 6 × 45
##      GameId GameDate   Quarter Minute Second OffenseTeam DefenseTeam  Down  ToGo
##       <dbl> <date>       <dbl>  <dbl>  <dbl> <chr>       <chr>       <dbl> <dbl>
## 1    2.02e9 2024-12-29       3      7      3 GB          MIN             1    10
## 2    2.02e9 2024-12-29       3      9     44 MIN         GB              0     0
## 3    2.02e9 2024-12-29       3      9     44 MIN         GB              0     0
## 4    2.02e9 2024-12-29       3      9     50 MIN         GB              1    10
## 5    2.02e9 2024-12-29       3     15      0 MIN         GB              0     0
## 6    2.02e9 2024-12-29       2      2      0 GB          MIN             0     0
## # ℹ 36 more variables: YardLine <dbl>, ...11 <lgl>, SeriesFirstDown <dbl>,
## #   ...13 <lgl>, NextScore <dbl>, Description <chr>, TeamWin <dbl>,
## #   ...17 <lgl>, ...18 <lgl>, SeasonYear <dbl>, Yards <dbl>, Formation <chr>,
## #   PlayType <chr>, IsRush <dbl>, IsPass <dbl>, IsIncomplete <dbl>,
## #   IsTouchdown <dbl>, PassType <chr>, IsSack <dbl>, IsChallenge <dbl>,
## #   IsChallengeReversed <dbl>, Challenger <lgl>, IsMeasurement <dbl>,
## #   IsInterception <dbl>, IsFumble <dbl>, IsPenalty <dbl>, …
```

Data Cleaning & Preparation

- Converted to factors when possible (Ex: PlayType, PassType)
- Parsed Description and added columns
  - FieldGoalResult
  - ExtraPointResult
  - IsSafety
  - DefensiveTwoPoint
  - PlayIndex
  - Corrected IsTouchdown
  - Several score variables

Touchdown Points Calculation

- Point calculation was derived from the strings provided in our description column.
- Thus, there were plenty of edge cases, such as touchdowns which were reversed on review or via penalty.
- We initially thought the built-in isTouchdown variable could be used to help get scores however, isTouchdown does not differentiate between recalled touchdowns and actual touchdowns.

Defensive Point Calculation

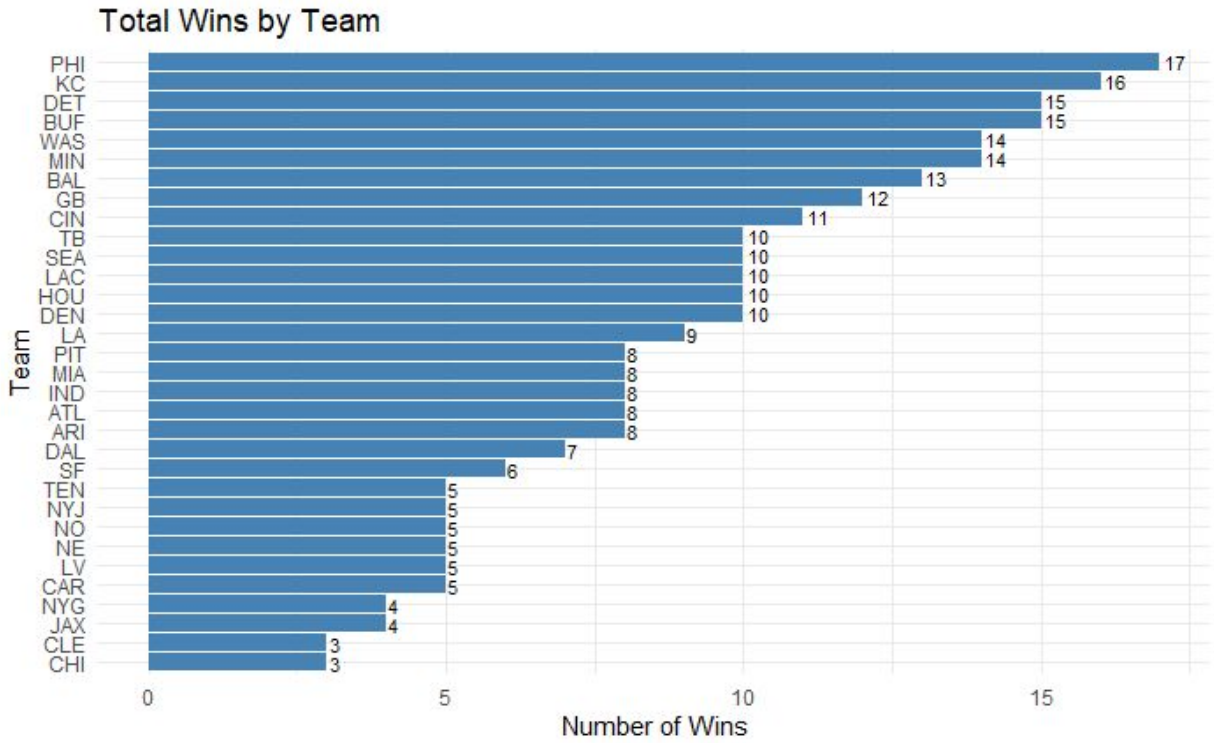
- Defense Point Calculation was also tricky.
- We parsed the description variable for key defensive plays like interceptions, defensive two point conversions and safeties
- There were plenty of edge cases for defensive points like penalties and “No Plays”

Result of Points Calculations

Columns like these were added to our dataset:

- OffPointsPlay (How many points were generated for the offense on that specific play)
- DefPointsPlay (How many points were generated for the defense on that specific play)
- OffenseScore (How many points the team which are on offense has in total)
- DefenseScore (How mant points the team which are on defense has in total)
- ScoreDiff (Offense score - defense score)

What Makes a Team Successful?

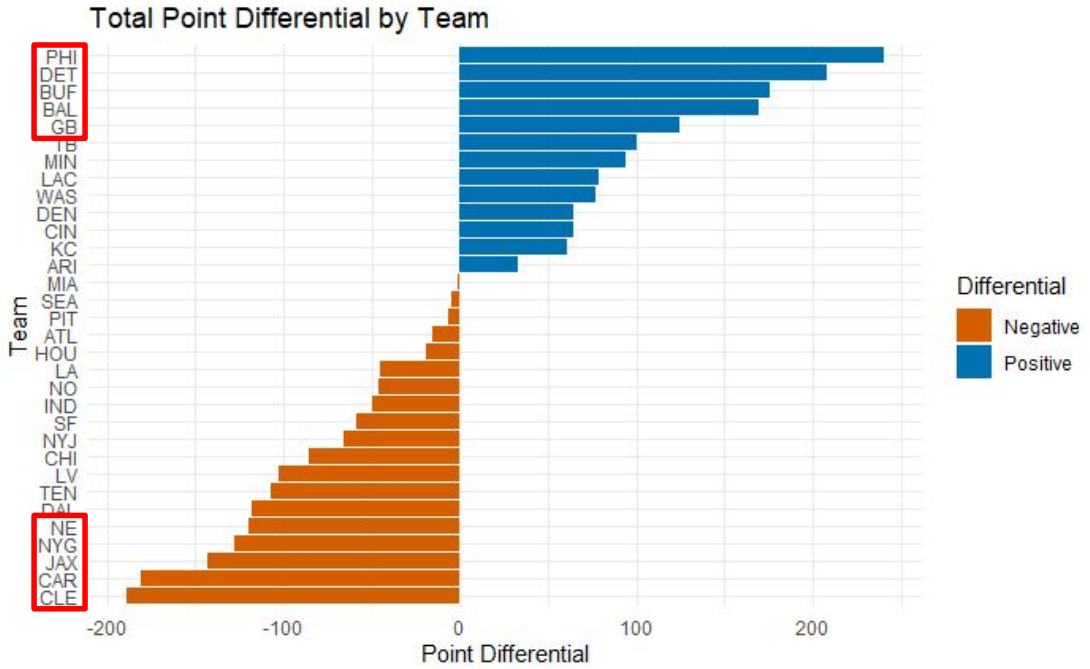


What Makes a Team Successful?

Point differential is a better indicator of a team's strength

Wins don't tell the whole story

Ex: Strength of schedule

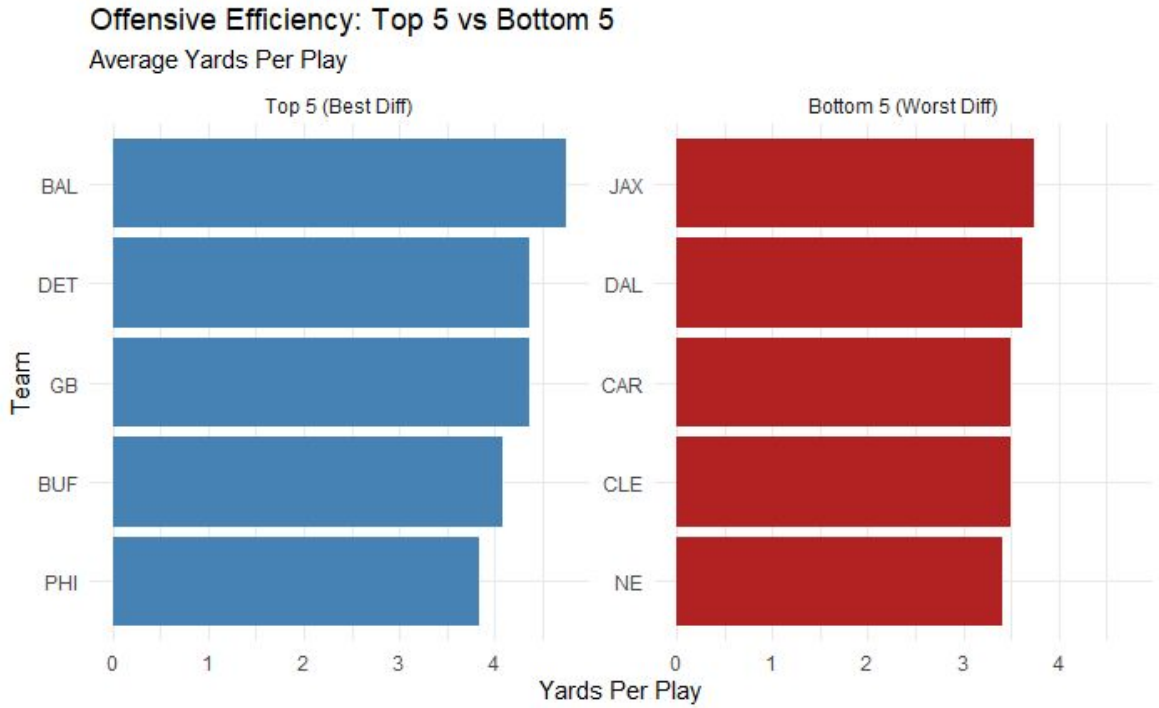


Offensive Efficiency

Better teams get more yards

Baltimore is especially efficient

BAL > 4.5 yards per play

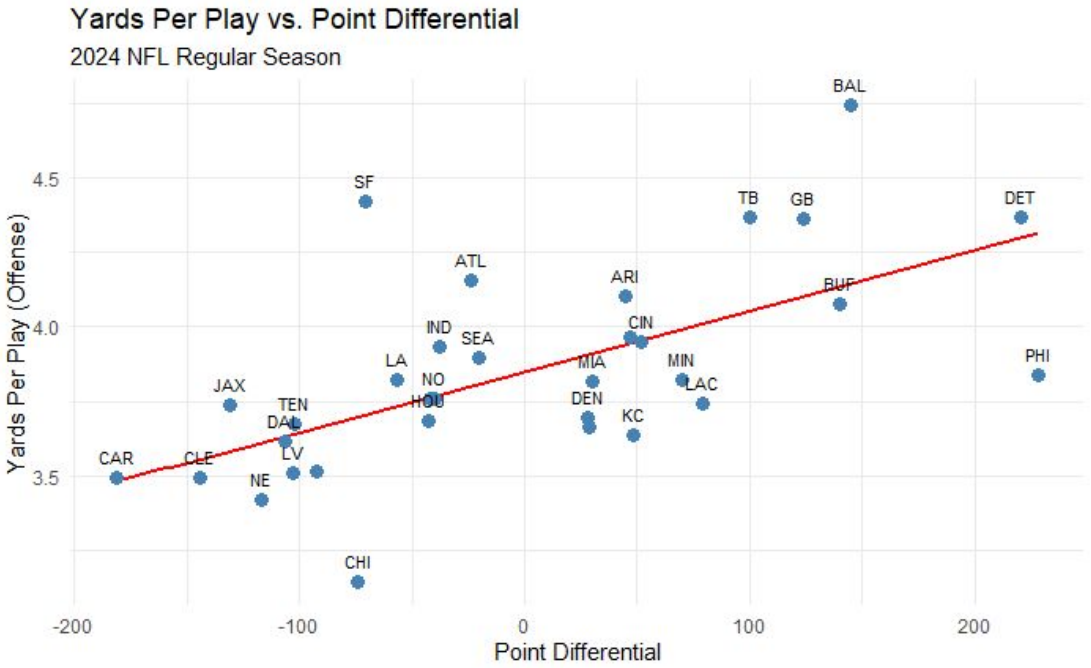


Offensive Efficiency

Clear positive correlation between yards per play and point differential

Correlation (r) = 0.6384

Teams that get more yards per play are better



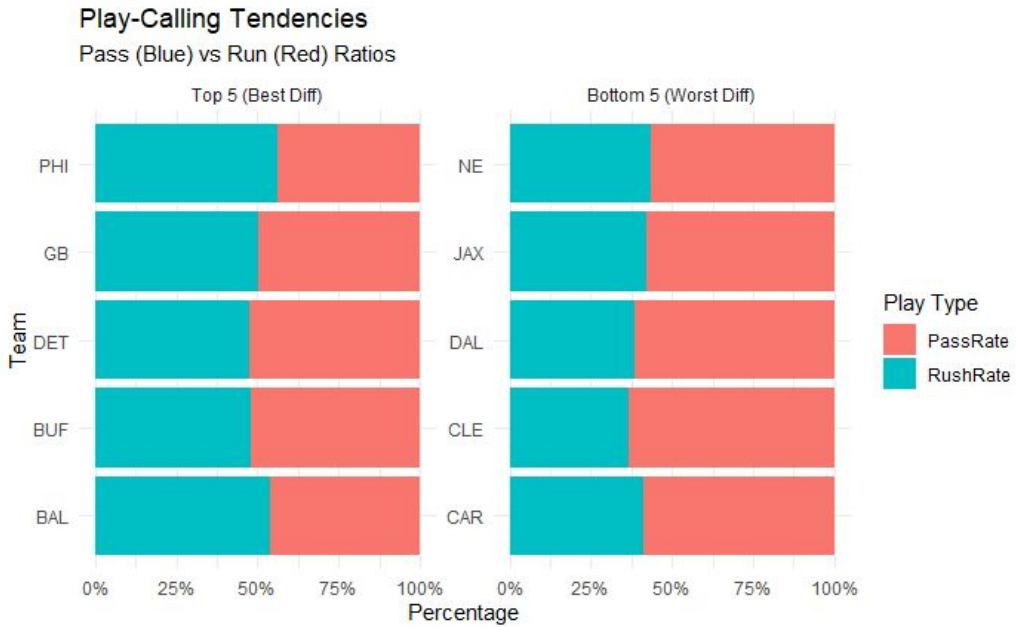
Play-Calling Tendencies

Forces the defense to “respect the run”

3 of the top 5 run ~50%

0 of the bottom 5 run >50%

Good teams run the ball more

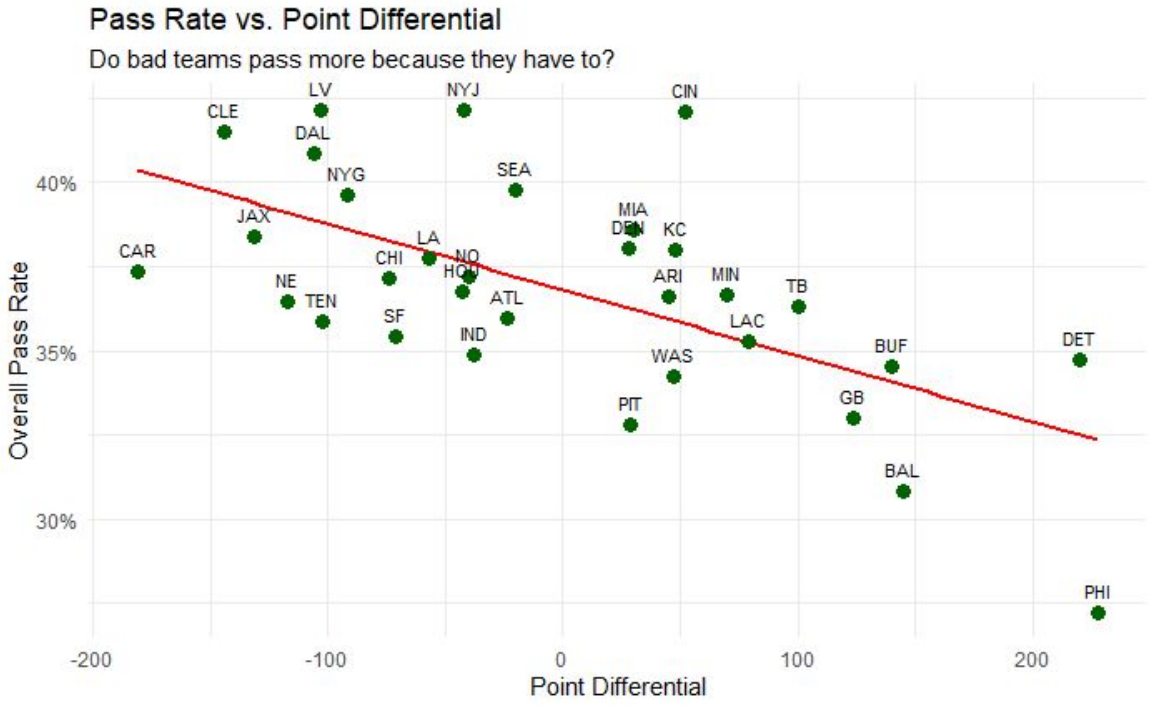


Play-Calling Tendencies

Teams that can't run the ball don't win games

Correlation (r): -0.6021

The less teams pass the ball, the better they are.

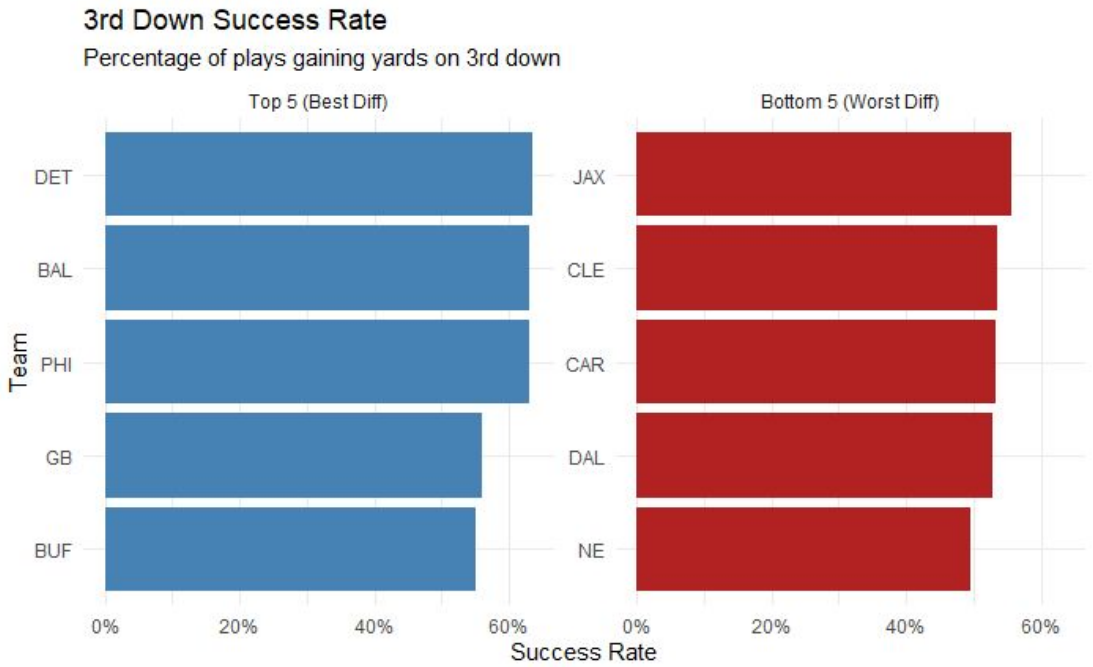


Down Efficiency

Better teams convert on more opportunities

DET, BAL, and PHI > 60%

JAX < 55%

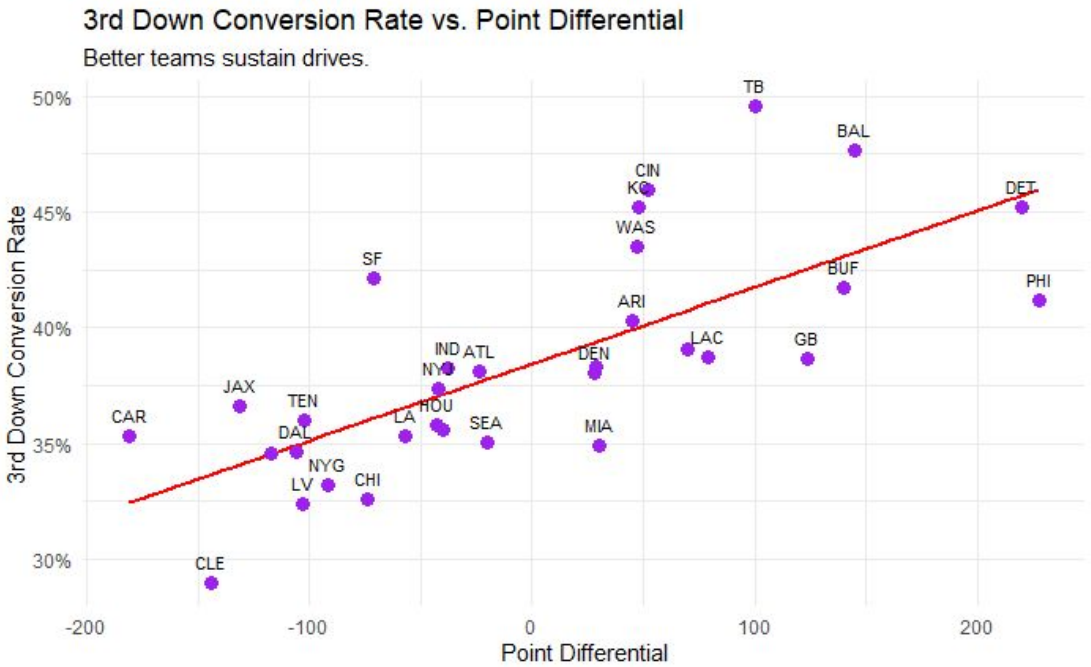


Down Efficiency

Good teams are highly effective on 3rd down

Correlation (r): 0.7406

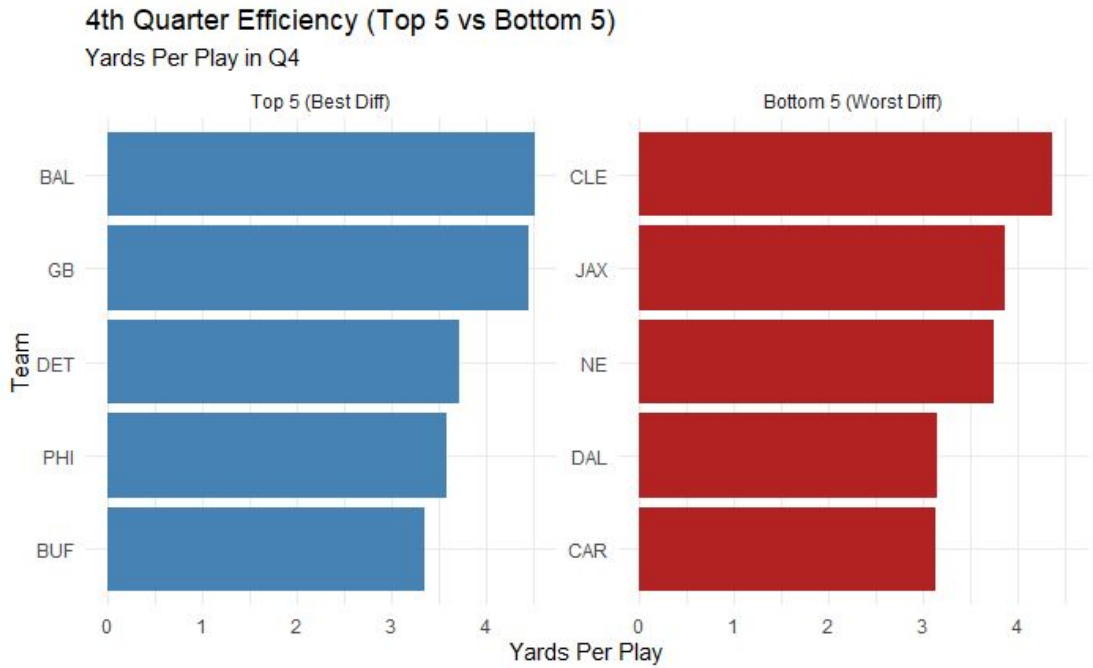
3rd down success rate is a strong indicator of team success.



4th Quarter Efficiency (Game End Pressure)

Baltimore continues to indicate offensive success

BAL ~4.5 yards per play in Q4



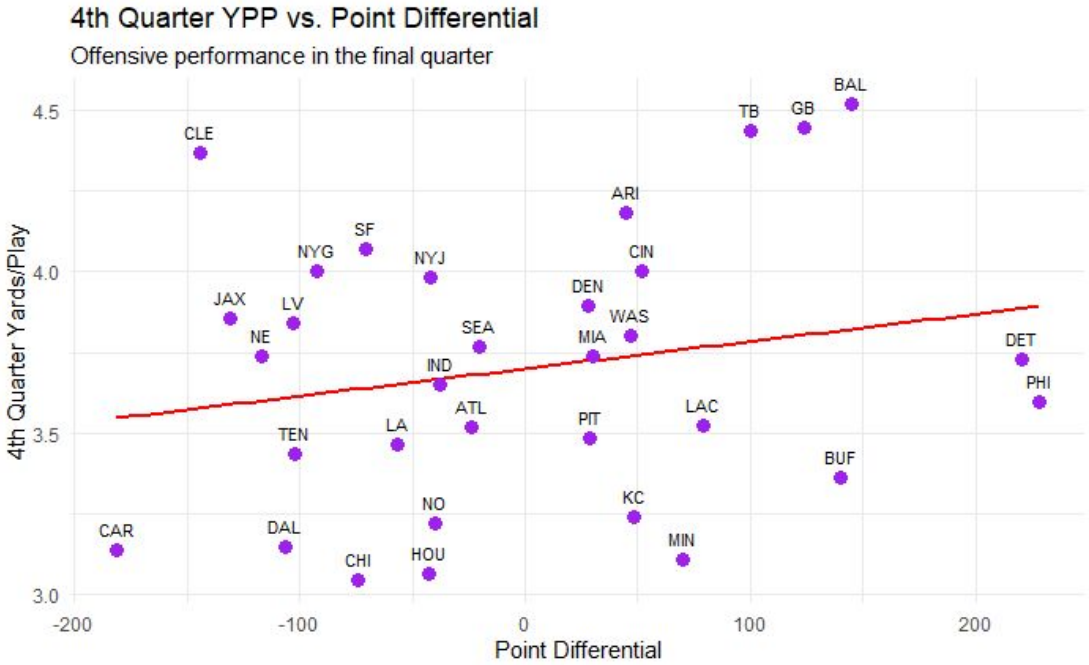
4th Quarter Efficiency (Game End Pressure)

Weaker correlation?

Good teams often already winning by a good margin

Correlation (r): 0.1695

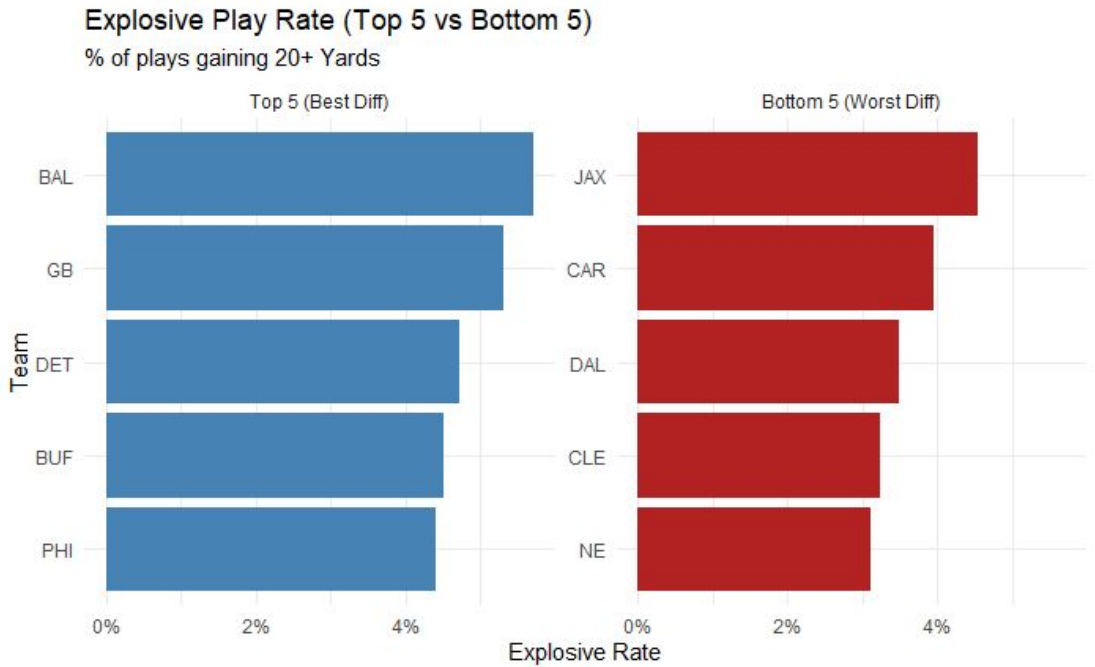
Overall, 4th quarter efficiency appears not to matter as much as the rest of the game.



Explosive Play Rate (Big Play Ability)

Better teams have bigger plays

Effective quarterbacks



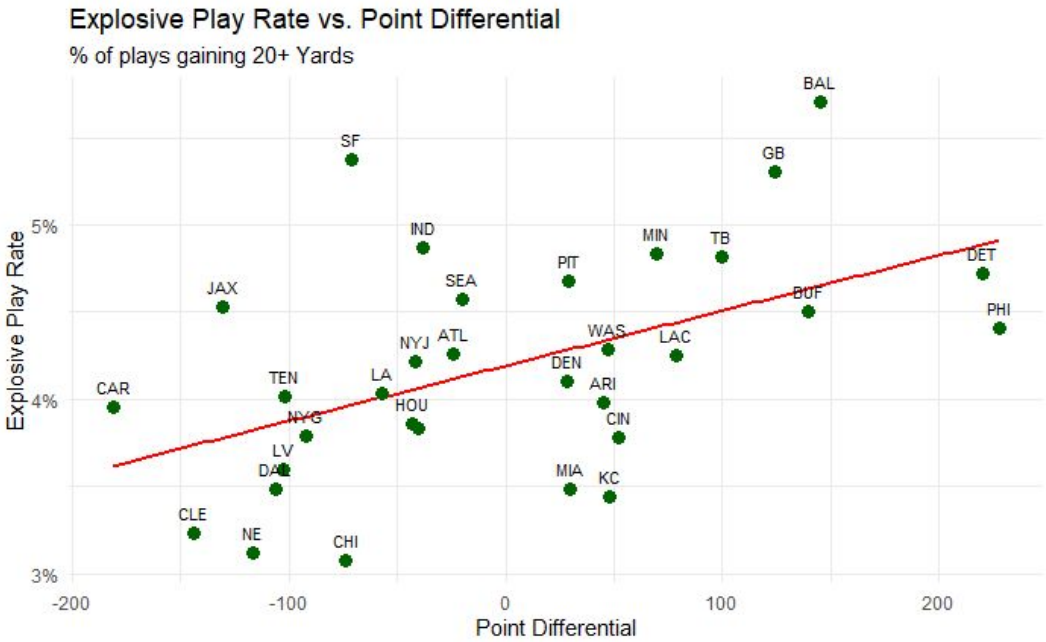
Explosive Play Rate (Big Play Ability)

Baltimore is a very explosive team

Effective quarterbacks drive good teams

Correlation (r): 0.5223

Teams that are more explosive offensively are better.

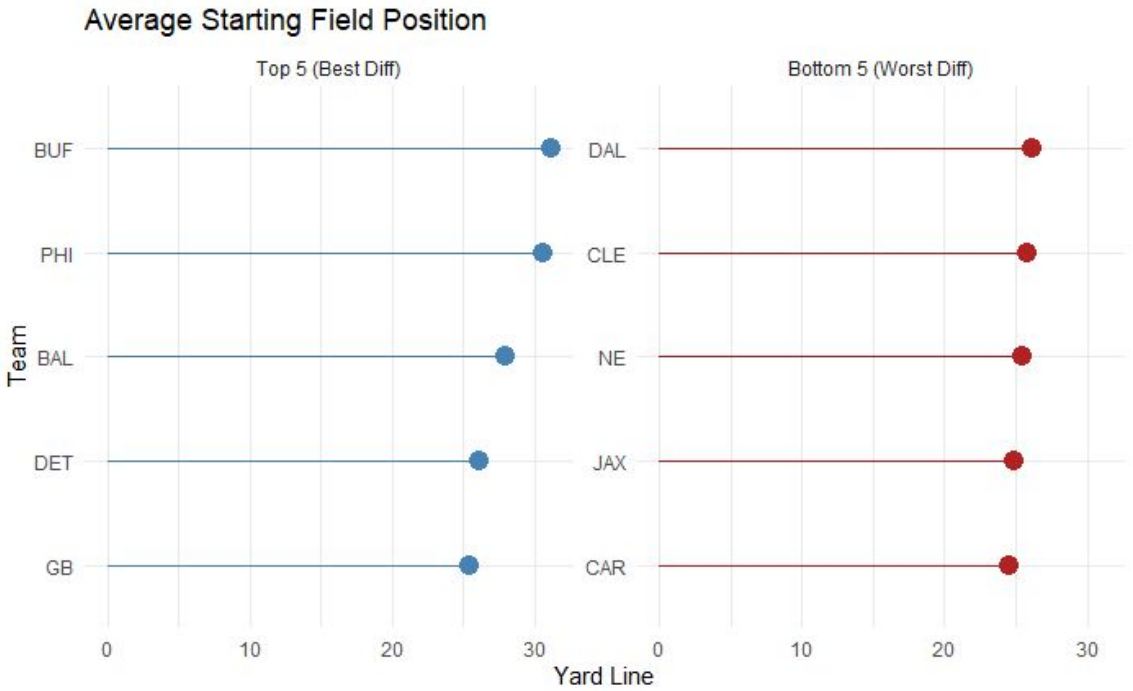


Defensive Stats

Switching Focus to Defense: Starting Field Position

Starting field position is a strong indicator on a team's defensive effectiveness

Teams like Buffalo and Philadelphia start 5 yards further upfield than the worst teams.

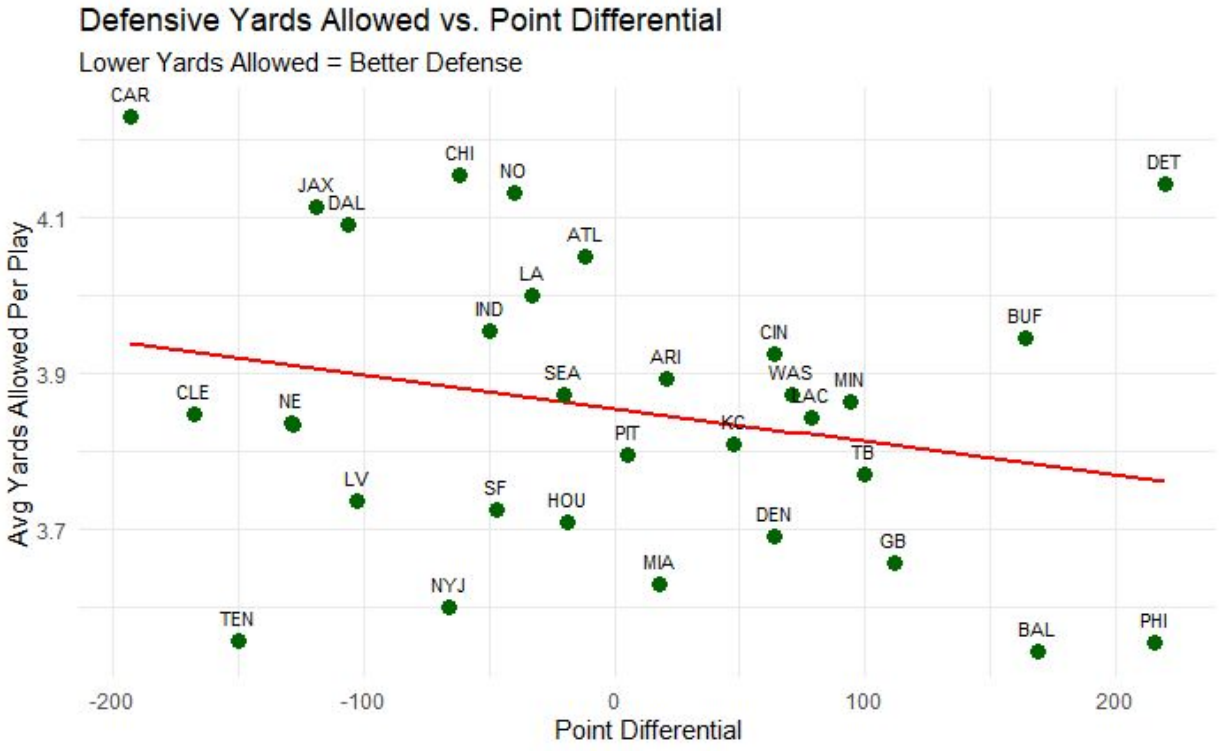


How much does having a Good Defense actually help you?

Detroit and Buffalo still have very good point differential even though they allow more yards per play than average

Correlation here is much weaker than the Offensive yards per play

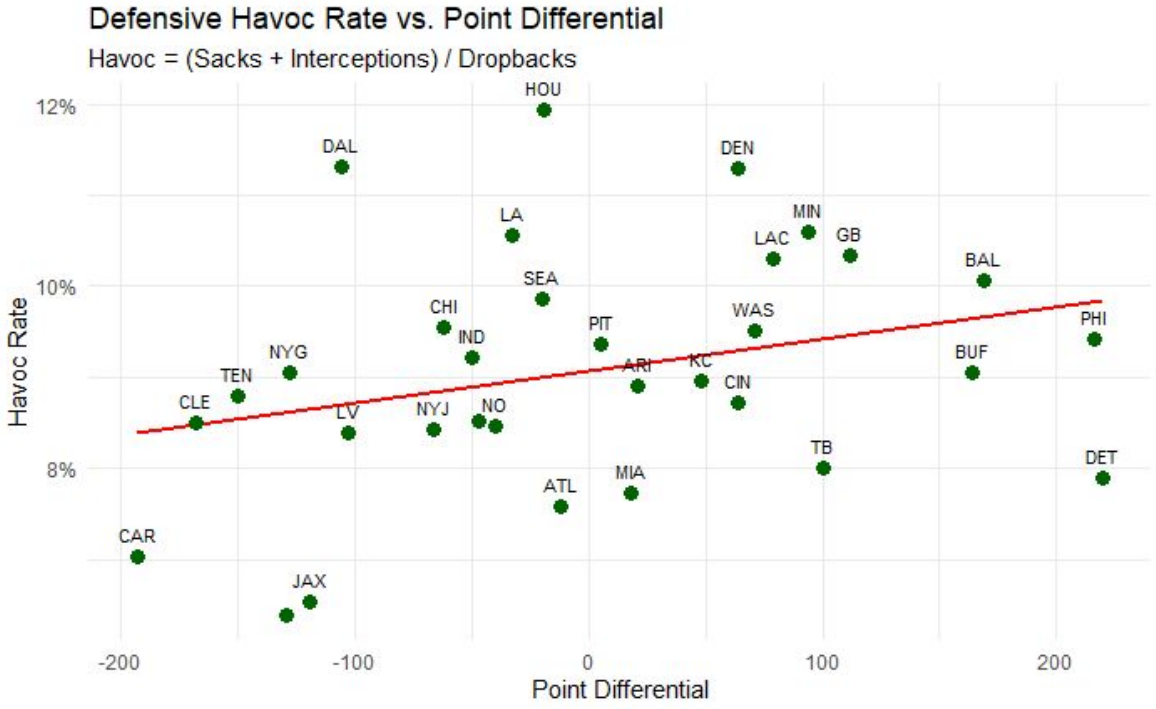
Correlation (r): -0.2489



How much does having a Good Defense actually help you?

Correlation (r): 0.2952

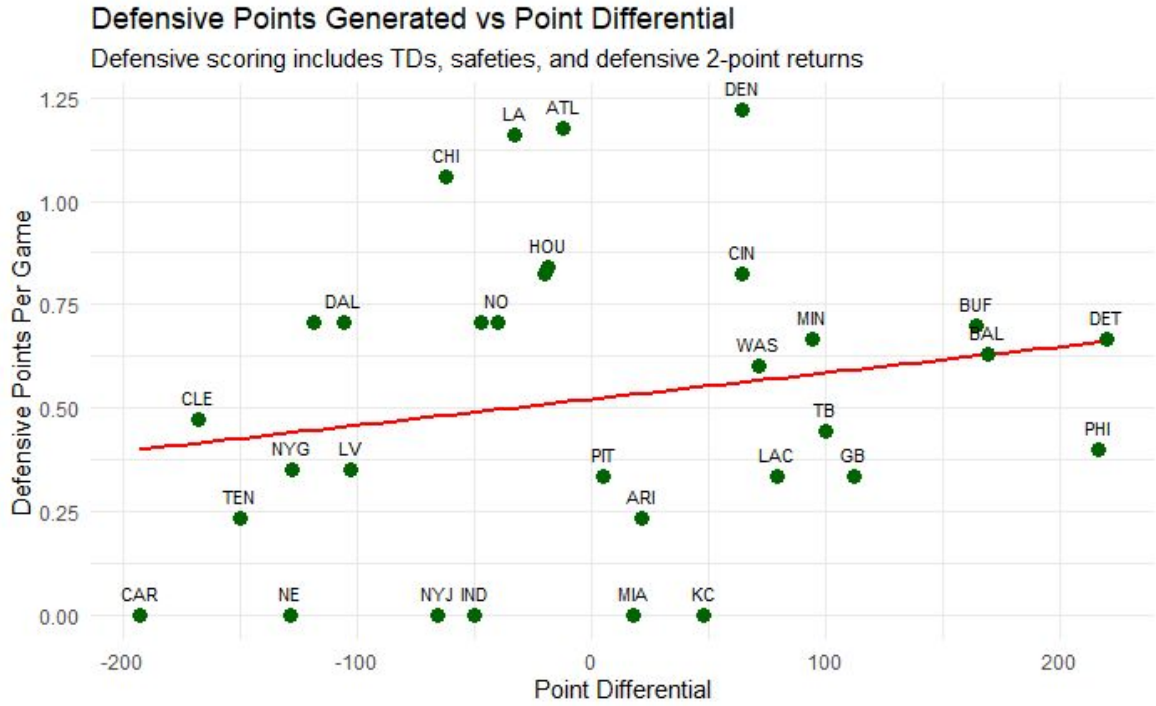
Defense matters, but offense still explains success more strongly.



How much does having a Good Defense actually help you?

Correlation (r): 0.1917

Once again, there is a slight correlation but not significant like most offensive stats.



Conclusions

1. Offensive success is much more indicative of point differential than defensive success
2. The strongest indicator of a team's success that we explored is 3rd down success rate
3. The weakest indicator of success was 4th quarter effectiveness.

Further Questions

- How would changing our metric of success from point differential to overall record affect the statistics?
- Why is the correlation between 4th quarter efficiency and a team's success so loose?
- How do different defensive schemes work against the best teams?