

TOCQ: Tackle Opportunity Containment Quotient in the 2022 National Football League Season (Weeks 1-9)

Duke University Masters of Statistical Science Portfolio Report

Eli Gnesin

April 5, 2024

1 Introduction and Motivation

During a play in a football game, such as in the National Football League (NFL), the primary goal of the defenders is to stop the ball carrier from advancing. This can be by tackle, out of bounds, forcing a fumble, or forcing a slide. The box score, however, only credits the player(s) who actually “make the tackle”, as if they did so alone. A player who is often in the “right place” but who is not credited for the tackle receives no credit for being “in the right place at the right time”. Conversely, a player who sees tackle “opportunities” slip away does not see negative impact for it in the box score. With this in mind, I define **TOCQ**, or **Tackle Opportunity Containment Quotient**, which reflects a defender’s ability to “contain” the “tackle opportunities” with which he interacts. I will first define **TOCQ** and explain the calculation process before discussing **TOCQ** in the first half of the 2022 NFL Season, using data provided during the 2024 NFL Big Data Bowl [2].

2 Defining TOCQ

TOCQ is the quotient of “tackles contained” and “tackle opportunities.” Consider a defender X . The *tackle zone* of defender X is the sector of a circle, centered on the player with radius $r = \max(1.25, 1.25 + \frac{s}{10} + \frac{1}{2 \cdot 100}a)$ yards, with a 150° sector angle bisected by a vector in defender X ’s orientation direction. The reasoning behind this radius is that, in this range, it is feasible the defender, facing the ball, could both see and react to the ball carrier. Given NFL speeds rarely exceed 22 mph or ~ 10.5 yards/second, and a frame rate of 10 frames per second (fps), most players would take ≈ 3 frames to pass through a tackle zone if completely unencumbered, at top speed, and the defender were static [4]. Further, if a player is moving forward, the tackle zone is adjusted by the displacement he would have before the next frame, using the standard displacement formula $\Delta x = st + \frac{1}{2}at^2$ for speed s in yards per second, acceleration a in yards per second squared, and time t , in order to recognize that the defender has increased range if he is moving in the direction of the ball as opposed to being static. Since the frame rate is 10 fps, the displacement formula is evaluated at $t = 0.1$, which gives the above radius calculation.

Now, consider an arbitrary play. For each frame between the ball snap + 2 frames (for run plays, adjusted to avoid unfairly penalizing defensive linemen who line up over the ball) or the pass completion (for pass plays), first calculate each defender’s tackle zone, and then check if the football is inside the zone. If this is true for defender X , he is credited with a *tackle opportunity*. Importantly, tackle opportunities are binary in a given play; even if the ball carrier leaves the tackle zone and later returns to it, he receives one tackle opportunity. When a play ends, whether by tackle, fumble, sack, slide, or out of bounds, that frame is the “action frame,” which is used to count “tackles contained.” If the ball is within a defender’s tackle zone during this frame, the defender is credited with a *tackle contained*. How that involvement occurs is ignored; “tackler” is the same as all other players with the ball in their zones. Any tackle contained is also simultaneously a tackle opportunity.

For defender X , there are three outcomes:

1. Defender X has no *tackle opportunity*. This is neutral for **TOCQ**.
2. Defender X has a *tackle opportunity* but no *contained tackle*. This is negative for **TOCQ**.
3. Defender X has a *tackle contained* and corresponding *tackle opportunity*. This is positive for **TOCQ**.

I now define **TOCQ_X** for defender X as:

$$TOCQ_X = \frac{\# \text{ of Tackles Contained by Defender } X}{\# \text{ of Tackle Opportunities for Defender } X}$$

Similarly, I define **TOCQ_T** for a team as:

$$TOCQ_T = \frac{\# \text{ of Tackles Contained by Team } T}{\# \text{ of Tackle Opportunities for Team } T}$$

For consistency, **TOCQ** without a subscript refers to the general metric and the subscript is for disambiguation in discussing actual values between team and players. The two metric forms above have slightly different interpretations. For the individual player, a high **TOCQ_X** reflects a defender who “makes the most of his opportunities” and is often “in the right place at the right time.” Conversely, a low **TOCQ_X** could reflect a defender who has opportunities pass but is not involved in as many plays. For a team, a high **TOCQ_T** reflects quick team containment of ball carriers and a lower **TOCQ_T** indicates a team which struggles to quickly contain ball carriers or has more “hero” plays where one defender makes a play solo.

Consider an example play from the Week 2 game between the New York Jets (NYJ) and Cleveland Browns (CLE). Table 3 shows an output table for the **TOCQ_X** scores for that single play. There are 11 rows, one for each of the 11 players on defense on that play. In this case, 4 players had a tackle opportunity, and 2 players had a containment, so they are denoted as such and every other player has a 0 in both column. Players are tracked across the season by an index of a unique player ID (provided in the dataset) and their team abbreviation.

3 Sensitivity Analysis of Tackle Zone Radius and Standard Errors

Before exploring the real outcomes for specific players and teams during the first half of the 2022 season, it is worth considering the sensitivity of **TOCQ** to the choice of tackle radius, as well as the standard error in the half-season **TOCQ_X** and game **TOCQ_T** scores. For this, I considered varying tackle radii of $r = \max(c, c + st + \frac{1}{2}at^2)$ for $c = 1, 1.25, 1.5$ (where $c = 1.25$ was the value used for **TOCQ** in the analysis stage), denoted as “cyd Varying,” as well as a constant radius $r = 1.5$, denoted as “1.5yd Constant,” for comparison. Likewise, it is worth considering the density of half-season **TOCQ_X** scores and half-season tackle opportunity counts at these 4 radii, as well as the mean and standard error of the mean. For all 4 radii, there are 869 players with **TOCQ_X** scores, and 136 games, resulting in 272 **TOCQ_T** scores (since each game has a **TOCQ_T** score for each team).

First, looking at the densities in Figure 1, the density of half-season **TOCQ_X** scores are relatively similar, with the smaller varying radius of 1.0 yards being slightly left-shifted and the larger varying radius of 1.5 yards being slightly right-shifted. All four radii show similar unimodal density with small clusters of data at both 0 and 1 due to players with small sample size (if a player had zero opportunities, he was also recorded as having **TOCQ_X** = 0). Looking at the densities of the tackle opportunities in Figure 2 reinforces Figure 1; the smaller varying radius of 1.0 yards has its density slightly more concentrated towards zero than the larger radii, and the larger radius of 1.5 yards has a fatter right tail than the other radii.

Mean and SE of TOCQ_X score and opportunity with varying radius				
Radius	Mean TOCQ_X	SE TOCQ_X	Mean Player Opportunities	SE Player Opportunities
1.0yd Varying	0.4031020	0.1994383	33.47181	30.45966
1.25yd Varying	0.4378373	0.1971925	37.98504	34.02091
1.5yd Constant	0.4414003	0.1981126	37.57077	33.39981
1.5yd Varying	0.4648202	0.2003979	42.03337	37.17232

Table 1: Table of mean and standard error of half-season **TOCQ_X** and tackle opportunities with varying tackle zone radius.

Looking at Table 1 provides further support to the argument that the choice of tackle zone radius has minimal impact on the overall distribution of **TOCQ_X** scores. For all four radii, the mean, half-season **TOCQ_X** score is within 0.062, which is significantly less than the standard error of all four radii, which hovers around 0.20. Since **TOCQ** scores are bounded between 0 and 1, this is a large standard error, and suggests that there is no statistically significant difference between the means of the densities. Conversely, though there is a seemingly large difference in the mean number of opportunities based on the radius, with the 1.0 yard varying radius having a mean number of opportunities ≈ 8.5 lower than the 1.5 yard varying radius, the standard error of all the opportunity counts is over 30, again indicating that there is not a statistically significant difference between the radii.

Mean and SE of TOCQ_T score and opportunity with varying radius				
Radius	Mean TOCQ_T	SE TOCQ_T	Mean Team Opportunities	SE Team Opportunities
1.0yd Varying	0.4197093	0.06903573	107.3321	22.52391
1.25yd Varying	0.4515254	0.06621810	121.8044	25.66440
1.5yd Constant	0.4542192	0.06463206	120.4760	26.04978
1.5yd Varying	0.4746512	0.06292535	134.7860	28.61944

Table 2: Table of mean and standard error of team single game **TOCQ_T** and tackle opportunities with varying tackle zone radius.

Similarly, it is worthwhile to compare the choice of tackle zone radius with respect to the team game **TOCQ_T** scores and the team number of opportunities per game, as presented in Table 2. Due to the much larger sample size per game, even as opposed to an individual player over 8-9 games, the standard error of the **TOCQ_T** scores is much lower than the individual player level (there are no 0 or 1 values at the game level, and as noted in 5 all single game **TOCQ_T** scores fall between 0.22 and 0.63). Further, with the mean team opportunities, there is again a positive correlation between the size of the varying radii and the mean team opportunities, though again the high standard error in both the **TOCQ_T** scores and the team opportunities indicates that these differences are not statistically significant.

The rest of the analysis proceeds with the 1.25 yard varying radius, that is $r = \max(1.25, 1.25 + st + 0.5at)$.

Since *TOCQ* is explanatory, rather than predictive, there is no metric by which to measure a “optimal” radius. As explained before, the 1.25 yard varying radius is intuitively driven given player speeds and a simple accounting of player motion, though further exploration of a “optimal” radius is a potential future step.

4 TOCQ for Weeks 1-9, 2022

Using the metric defined in Section 2, as well as the radius established in Section 3, I calculated **TOCQ** for individual players and teams in the first half of the 2022 Season. There are 136 games and 869 players who played a defensive snap (including 4 offensive players). First, I calculated **TOCQ_X** for single games and for the half-season. To focus on regular players and to rule out players who may have only played a few plays or a single game, I am only presenting **TOCQ_X** scores for players with 30 or more opportunities for the half-season. Further, given the small sample size of any individual game, and acknowledging the large variance in even half-season scores in Figure 1, single game scores for individual players are not considered, since no conclusions can be drawn from them.

Table 4 shows the top 10 and bottom 10 players by half-season **TOCQ_X** score, of players with a minimum 30 opportunities, of which 442 players, or 50.9%, qualify. The top 10 players by half-season **TOCQ_X** are mostly defensive backs, led by Pittsburgh Steelers Cornerback (CB) Levi Wallace. The bottom 10 players are all considered defensive linemen, with the exception of Za'Darius Smith of the Minnesota Vikings, who is marked as an Outside Linebacker but also regularly acts as a Defensive End. This is reasonable; on run plays, a defensive lineman, especially at the center of the line, can pick up a tackle opportunity if the running back cuts from one side of the line to the other, even if the defensive lineman is, himself, engaged with an offensive lineman as a blocker. Since tackle opportunities ignore blocker position, defensive linemen can pick up these “phantom” opportunities when, though they are in range, actually containing the rusher is not nearly as clear-cut. Secondary players, conversely, likely only have opportunities when they are “in on the play,” and may often be the primary defender responsible on a pass play, which defensive linemen rarely are.

With this in mind, I then considered the distribution of **TOCQ_X** scores (among players with 10 or more opportunities in the half season, of which 657 players, or 75.6%, qualify) by position and position group. Figure 3 shows these distributions by boxplots by position, colored by the three “position groups,” defensive line, linebackers, and secondary. Figure 3 reinforces some of the trends seen in Table 4. Among position groups, secondary players mostly have higher mean **TOCQ_X** scores, while defensive linemen mostly have lower mean **TOCQ_X** scores. Acknowledging the large standard errors discussed in Table 1, and based on the boxplots of Figure 3, however, no statistically significant conclusions can be drawn from these data.

I then move to consider **TOCQ** at the team level. As noted, **TOCQ** at the team level is denoted as **TOCQ_T**. As with **TOCQ_X** for individual players, **TOCQ_T** can be calculated for teams in single games and over the half-season. Unlike **TOCQ_X** above, **TOCQ_T** includes all players on each team in each game, since the calculation of **TOCQ_T** sums over all players, and small sample sizes are not an issue. As noted in Table 2, the mean number of opportunities per game per team is over 120.

Table 5 shows the top and bottom 10 games by **TOCQ_T** scores. As noted in Table 2, **TOCQ_T** scores have much less variance than the **TOCQ_X** scores in table 4. Looking at Table 5, there are few patterns or trends, but the Carolina Panthers (CAR) has the 2 worst team performances in the dataset in weeks 5, and 7. Given Carolina was poor in 2022, this is unsurprising, and suggests their defense struggled to finish plays when they were close to ball carriers. On the flipside, the top 10 team games are represented by 10 different teams, and 4 teams appear on both lists in different weeks. Looking at Figure 4 shows several different trends in the mean **TOCQ_T** scores by team. Carolina is far and away the worst team by **TOCQ_T** in the dataset, and appears to almost single-handedly drag the “league average” below the league median. There are fewer noticeable trends in the teams than in the player **TOCQ_X** scores. The “below average teams” include playoff teams such as the Dallas Cowboys (DAL), Jacksonville Jaguars (JAX), and Tampa Bay Buccaneers (TB), as well as the Houston Texans (HOU), and Chicago Bears (CHI), two of the three worst teams in 2022 [1]. The best team by **TOCQ_T** is the Minnesota Vikings (MIN), which went 8-1 in the dataset and made the playoffs, as well as the Super Bowl runner-up Philadelphia Eagles (PHI), and the Pittsburgh Steelers (PIT) and Detroit Lions (DET), both of which started the season poorly and then turned their seasons around in the second half of the season (outside the dataset) [1].

Finally, Figure 5 shows the ratio of the average half-season TOCQ_T to league average in that position group, broken down into the three major position groups. This is a team level equivalent to Figure 3, though it demonstrates which teams overperformed and underperformed league average. The struggles of Carolina are most apparent on this figure, given the bright red marks for their defensive line and linebackers, while Philadelphia stands out for their defensive line overperformance, which is unsurprising considering they were at the top of the NFL in Sacks and Total Defense during this span [3]. Overall, Figure 5 helps identify which position groups of teams are contributing most strongly positively or negatively towards the team scores seen in Figure 4, relative to the league average of those position groups.

5 Conclusions, Limitations, Future Steps

This paper introduces the metric **TOCQ: Tackle Opportunity Containment Quotient**, parameterized for both individual players and full teams. **TOCQ** also introduces the *tackle zone*, the *tackle opportunity* when the football enters the tackle zone, and the *tackle containment* when the play ends within a defender’s tackle zone. At the individual player level, defensive linemen, specifically defensive tackles, have generally lower TOCQ_X scores than linebackers and secondary players. At the team level, the top teams by TOCQ_T include several teams that underperformed early in the season, such as Pittsburgh and Detroit, as well as strong playoff teams in Minnesota and Philadelphia, and the lowest teams include several teams with middling records such as Carolina, Denver, and Cleveland.

There are several limitations to this analysis and **TOCQ** as currently presented. With the dataset provided, only nine weeks of games from a single NFL season are available. Though `plays.csv` contains approximately 12,500 plays, restricting to only these weeks puts many players under the number of opportunities I thought sufficient to draw conclusions from (I chose 30 as the traditional statistical threshold for “small sample size”). As such, I chose not to analyze individual players from single game TOCQ_X scores, which would be a more granular level of analysis. This does not affect TOCQ_T , as noted, because a full game has a large enough sample of opportunities, but comparing teams on a sample of 8-9 games (depending on bye week) is also likely inadvisable due to the small sample size.

Within the dataset, I made several decisions which may have had adverse effects on the results. First, at least three plays were removed for missing player tracking data, and it is possible that are other plays have missing, incomplete, or incorrect player tracking data. More importantly, the choices I made regarding the frame events in the player tracking data may have significantly impacted the **TOCQ** scores. I chose to only consider frames between two frames after the “ball snap” (on run plays) or from the “pass outcome caught” frame (on pass plays) and through the tackle/out of bounds/slide/sack/fumble frame. I chose this because, for pass plays, not every frame from the ball snap is available, so I sought consistency. This, however, creates inconsistency between pass and run plays, which cannot be resolved because data from the snap is not provided on passing plays. Relatedly, while the decision to adjust to 2 frames after the snap does avoid penalizing defensive linemen for lining up over the snap (i.e. across from the center), the choice of 2 frames is heuristically chosen and may be improved upon. Further, there may be errors in the event tagging which would cause the incorrect frames to be used in individual plays and could, if numerous enough, impact the results.

With changing the **TOCQ** metric and definition or calculation of tackle zones and tackle opportunities in the future, while player speed and acceleration were accounted for in the tackle zone measurement, there may be more complex, elegant ways of measuring a player’s motion to calculate an increased range. Further, the tackle zone is still based off a constant, $c = 1.25$, and it could be better to have that constant vary based on an individual player’s attributes, such as wingspan or height. Likewise, as noted previously, exploring **TOCQ** in a predictive sense could lend itself towards an “optimal” radius for each player, as measured by some convex function.

Finally, the tackle zone currently ignores the presence of blockers within the zone. This was deliberate; on many run plays, for example, an interior defensive lineman may be “obstructed” by a blocker and still make a tackle, but this could be changed, and it is possible that this decision is part of the reason defensive linemen have lower **TOCQ** scores. **TOCQ** would, in this vein, benefit from a more consistent set of frames in each play, instead of separate between run and pass plays.

References

- [1] *2022 NFL Standings & Team Stats*. URL: <https://www.pro-football-reference.com/years/2022/index.htm>.
- [2] Michael Lopez, Thompson Bliss, Ally Blake, Andrew Patton, Jonathan McWilliams, Addison Howard, Will Cukierski. *NFL Big Data Bowl 2024*. 2023. URL: <https://kaggle.com/competitions/nfl-big-data-bowl-2024>.
- [3] *Stathead Team Span Finder*. URL: https://stathead.com/football/team-span-finder.cgi?request=1&order_by=pass_sacked_opp&timeframe=seasons&year_min=2022&year_max=2022&span_length=9&span_type=season_start.
- [4] Edward Sutelán. *Who is the fastest player in the NFL? Tyreek Hill, other Dolphins dominate list of top 2023 speeds so far*. Oct. 2023. URL: <https://www.sportingnews.com/us/nfl/news/nfl-fastest-player-top-2023-speeds/80dde018741ad620c97859e5>.

6 Appendix

All code for calculating TOCQ and generating figures is found in [this github repository](#). Data is found at the Big Data Bowl link on [Kaggle](#).

6.1 Figures and Tables

TOCQ_X Output of Example Play			
Index	Player Name	Tackle Opportunity	Tackle Containment
41243_NYJ	C.J. Mosley	0	0
44815_NYJ	Solomon Thomas	0	0
46141_NYJ	Nathan Shepherd	0	0
46186_NYJ	Jordan Whitehead	0	0
46211_NYJ	D.J. Reed	0	0
46255_NYJ	Jake Martin	0	0
47881_NYJ	Quincy Williams	0	0
42467_NYJ	Kwon Alexander	1	0
54491_NYJ	Jermaine Johnson	1	0
41270_NYJ	Lamarcus Joyner	1	1
54469_NYJ	Ahmad Gardner	1	1

Table 3: **TOCQ** for a single play.

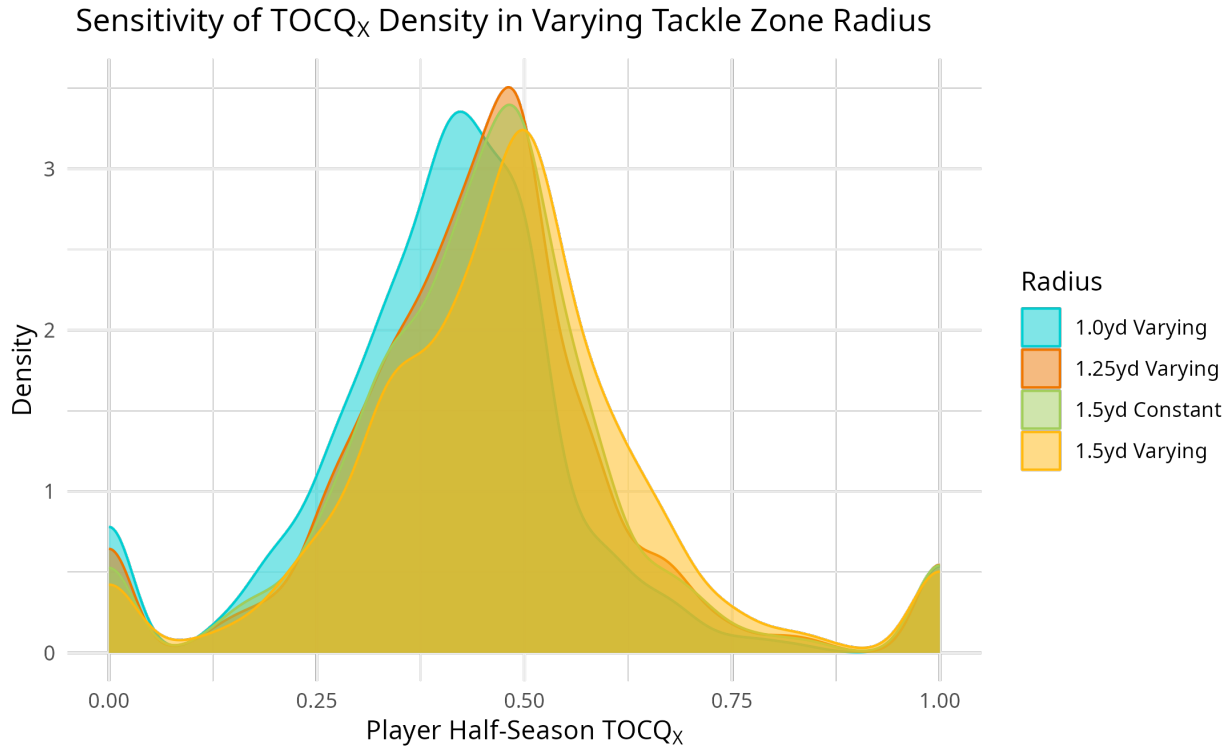


Figure 1: Density of half-season TOCQ_X scores with varying tackle zone radius.

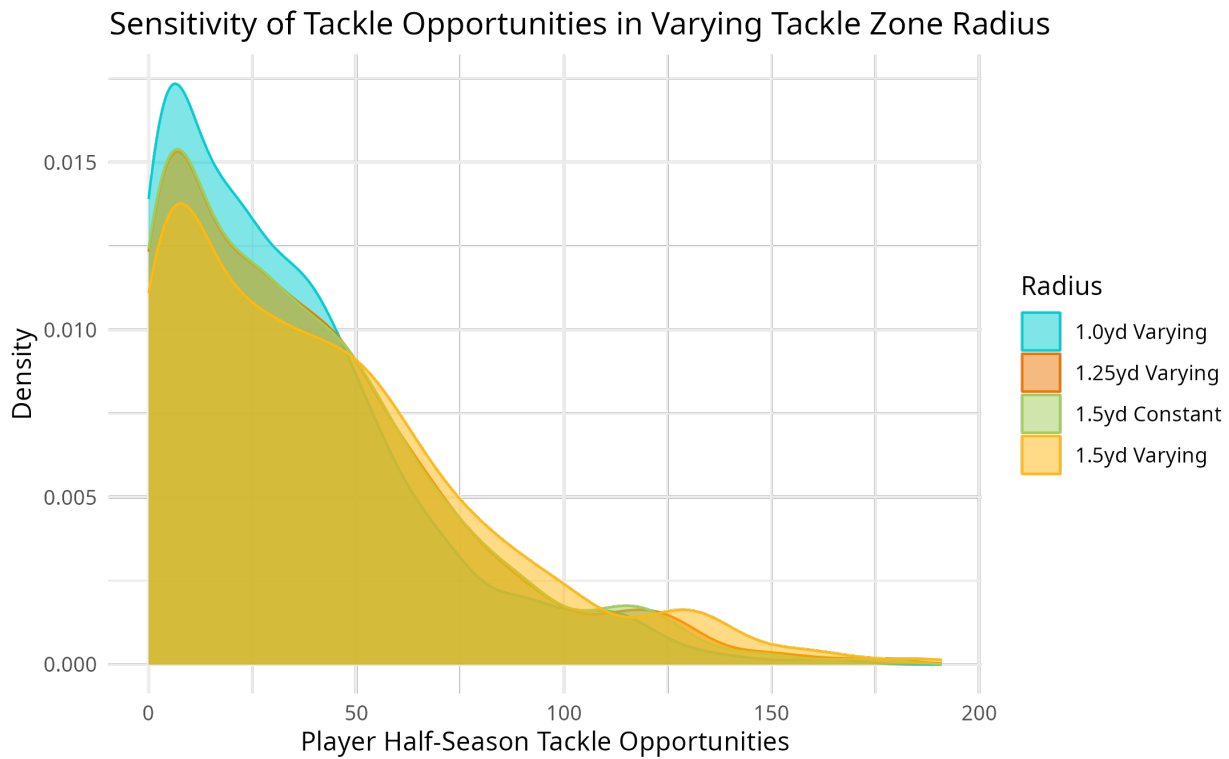


Figure 2: Density of half-season player tackle opportunities with varying tackle zone radius.

Half Season TOCQ_X Scores (minimum 30 Opportunites)							
Top 10 Individual Players				Bottom 10 Individual Players			
Player Name	Team	Position	TOCQ_X	Player Name	Team	Position	TOCQ_X
Levi Wallace	PIT	CB	0.7142857	Za'Darius Smith	MIN	OLB	0.1515152
Derrick Barnes	DET	ILB	0.7142857	Marquan McCall	CAR	NT	0.1538462
Terrance Mitchell	TEN	CB	0.7142857	Roy Lopez	HOU	DT	0.1587302
Darrick Forrest	WAS	SS	0.6956522	Vita Vea	TB	NT	0.1857143
D.J. Wonnum	MIN	OLB	0.6666667	Greg Gaines	LA	NT	0.1944444
Charvarius Ward	SF	CB	0.6603774	Carl Davis	NE	NT	0.2000000
Kyle Hamilton	BAL	FS	0.6571429	Tommy Togiai	CLE	DT	0.2051282
Michael Davis	LAC	CB	0.6410256	Abdullah Anderson	ATL	DE	0.2285714
Cameron Dantzler	MIN	CB	0.6376812	Teair Tart	TEN	DT	0.2340426
Derek Stingley	HOU	CB	0.6166667	Daron Payne	WAS	DT	0.2346939

Table 4: Player **TOCQ_X** scores for the half-season.

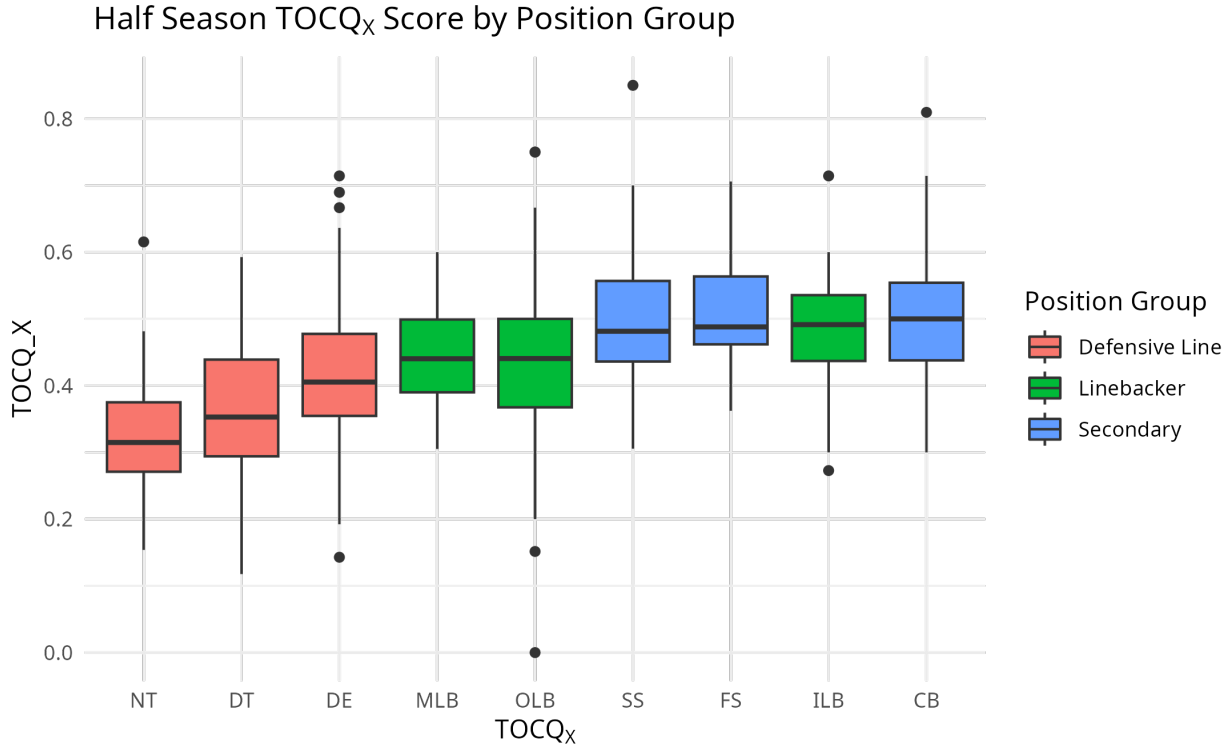


Figure 3: Distribution of half-season TOCQ_X scores by position group.

Team TOCQ_T Single Games						
Top 10 Team Games				Bottom 10 Team Games		
Team	Week	Opponent	TOCQ_T	Team	Week	TOCQ_T
BUF	6	KC	0.6276596	CAR	5	0.2258065
SF	8	LA	0.5982143	CAR	7	0.2519084
CHI	3	HOU	0.5973154	CLE	1	0.2830189
PIT	5	BUF	0.5964912	NYG	2	0.3027523
TB	2	NO	0.5940594	BUF	4	0.3046875
KC	6	BUF	0.5905512	HOU	5	0.3047619
NO	8	LV	0.5862069	TB	8	0.3172414
BAL	9	NO	0.5769231	CHI	1	0.3190184
TEN	9	KC	0.5743243	NO	3	0.3258427
NYJ	4	PIT	0.5684932	DEN	2	0.3260870

Table 5: Best and worst single game team TOCQ_T scores.

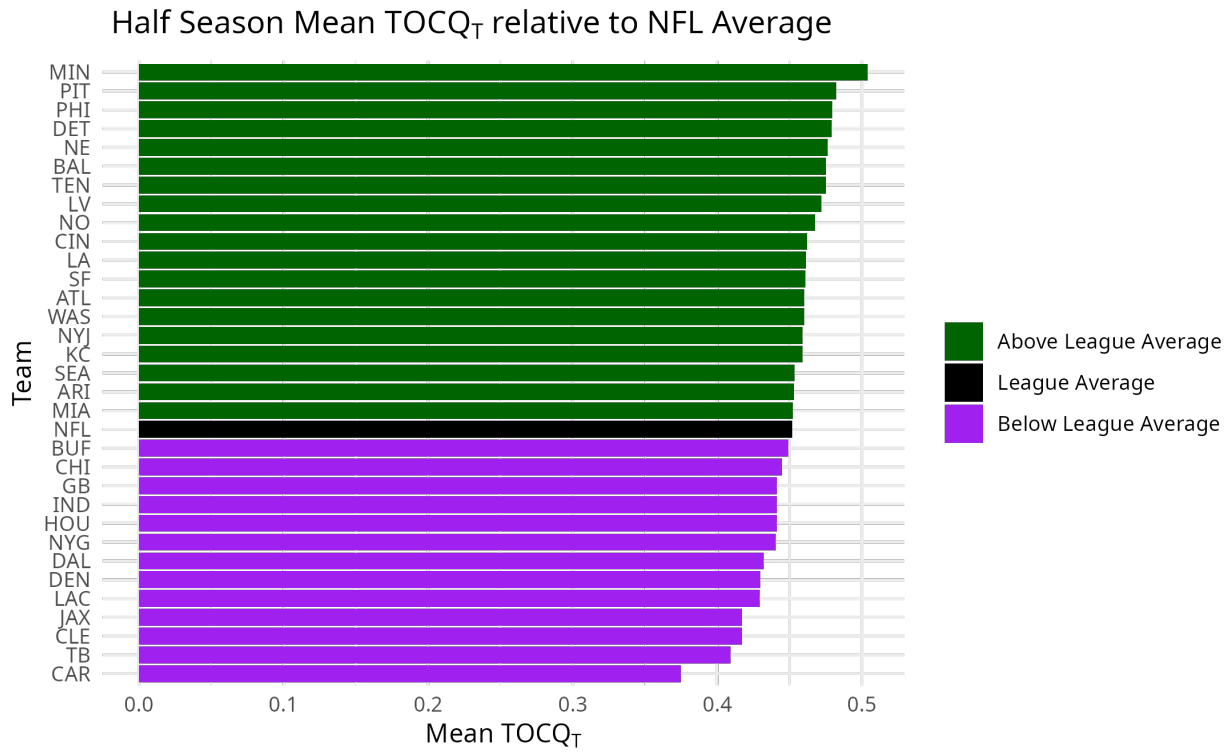


Figure 4: Barplot of half-season TOCQ_T scores relative to league average.

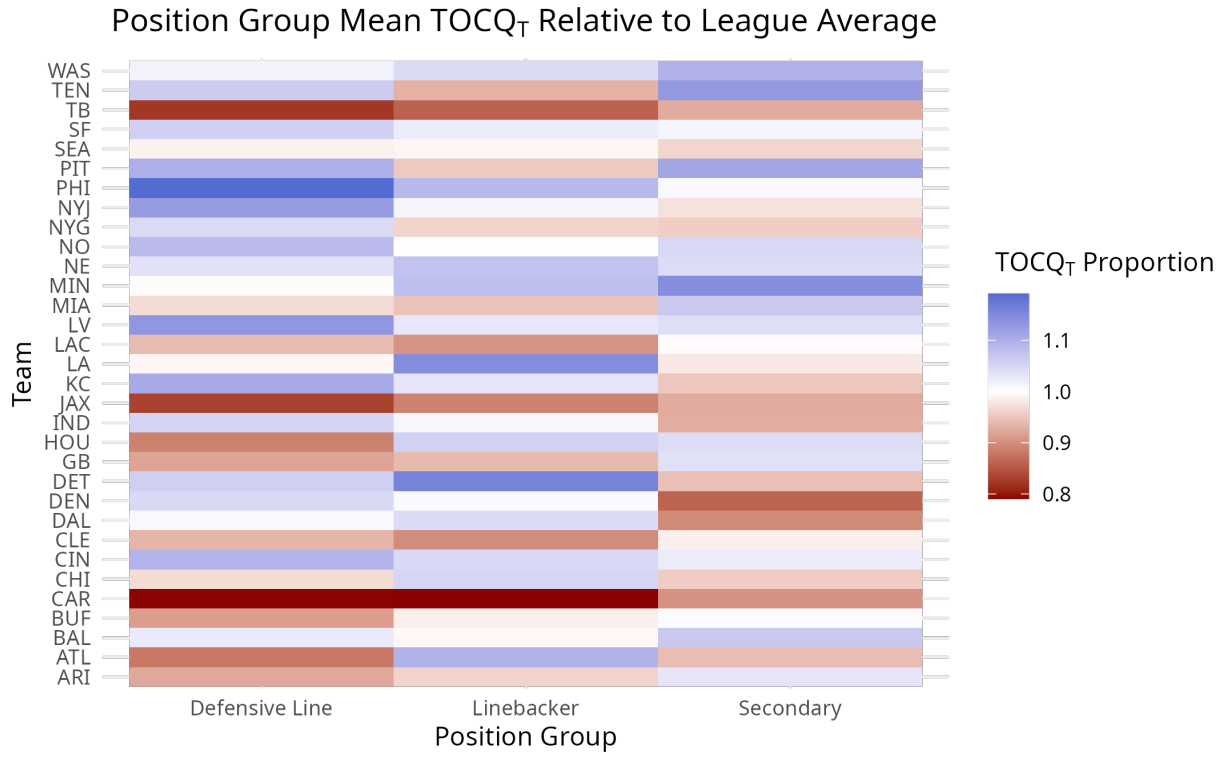


Figure 5: Heatmap of half-season TOCQ_T scores broken down by position group, as a ratio relative to position group league average