LOCATING AND EXPLOITING PERSONNEL MISMATCHES IN POPULAR ROUTE COMBOS AND DEFENSIVE COVERAGE SCHEMES: A DATA-DRIVEN ANALYSIS

Bradley J. Congelio, Ph.D.

@BradCongelio

KUTZTOWN UNIVERSITY

Sports Info Solutions
Football Analytics Challenge

INTRODUCTION

- The NFL is a game of strategy wherein locating and exploiting personnel mismatches is often the difference between winning and losing.
- Case in point: this piece from The Ringer where Danny Kelly outlines how "Josh McDaniels and the Patriots Create Mismatches to Burn Defenses."
  - In it, Kelly explains that the Patriots' offense – under Daniels – "adapt their game plan from week to week" to "best attack the specific vulnerabilities of an opposing defense." In short, Kelly argues, the Patriots are **"better than anyone at finding a weakness and attacking it mercilessly."**
- This research builds upon this idea, and in doing so, will do the following:
  - Showcase how a team's most used route combinations can be identified.
  - Show how to determine the success of each route combination against all defensive schemes.
  - Determine which wide receiver has the highest EPA for each route combination against each defensive scheme.
  - Determine where each WR should line up on the field, and which side of center, for the best chance of net-positive EPA.
  - Highlight how to identify the weakest pass defender for each route combination out of each defensive scheme to assist in creating strategic optimal matchups.

THEORETICAL STANCE

- It is "easy" to take data, code advanced models, and say "look at what I've created!"
- But to what point?
  - Meaning: **Does the cool, advanced model you created have any practical purpose? Would it *actually* be able to help an NFL team win games?**
- There is no purpose to analytics if the results cannot be transferred to the football field.
- Because of this, all my analytical work is grounded by William Barry's approach to action research: **all work must create ontological weight.**

Fancy way of saying research must "challenge the status quo" and seek to "improve what is taking place."

THEORETICAL STANCE

- The impact of creating ontological weight on this research:
  - The results must be able to contribute, directly, to helping NFL teams win games.
  - It needs to do more than discover popular routes combinations and which perform best against coverage types.
  - It needs to explain why we should care about that information and how NFL teams can use it to gain an advantage over their opponents.

ASSUMPTIONS

- The core assumptions built into this research derive from the following:
  - Because of the amount of parity in the NFL, coaching staffs are continually looking for even the smallest strategic advantage on a week-by-week basis.
  - Football is ultimately a game of strategically creating personnel mismatches.
  - As a result of those two assumptions, this research displays a strategic and data-driven approach to locating, isolating, and successfully implementing these personnel mismatches in the offensive passing game.

FIRST GLANCE

One of the core goals of this analytics challenge was to discover "which route combinations were most popular in the NFL in 2020."

By utilizing the `*stringr*` function in R, I was able to combine each route run per `EventID` (ie, play) and then find the most frequent of those combinations, as seen below:

Route Combos	Number
Curl, Curl, Curl, Flat – Left, Flat – Right	50
Chip – Flat, Chip – Flat, Curl, Curl, Curl	45
Beneath, Corner, Deep Cross, Run Fake	32
Flat – Left, Flat – Right, Over Ball, Slant, Slant	30
Curl, Curl, Flat – Left, Flat - Right	25
Curl, Curl, Flat – Left, Flat – Right, Over Ball	25
Out, Run Fake, Screen - Bubble	23

*** Route combinations that included: **(1.) only screen passes** or **(2.) one route and a run fake** were excluded. For example, the "Run Fake, Screen – Bubble" combination was run 143 times in 2020. However, given the scope of this Analytics Challenge, I do not believe those truly constitute route combinations.*

```
###########
##GATHERING ALL ROUTE COMBINATIONS FROM 2020 SEASON
###########

most.common <- complete.data %>%
  separate_rows(Route, sep = ',\\s*') %>%
  group_by(GameID, EventID) %>%
  summarize(route.combos = toString(Route)) %>%
  ungroup()


###########
##FINDING MOST FREQUENT COMBINATIONS
###########

most.common.combos <- most.common %>%
  mutate(rn = row_number()) %>%
  separate_rows(route.combos, sep=",\\s+") %>%
  arrange(rn, route.combos) %>%
  group_by(rn, GameID, EventID) %>%
  summarise(route.combos = toString(route.combos)) %>%
  ungroup %>%
  count(route.combos)
```

BUT, ROUTES DO NOT EXIST IN A VACUUM

- Examining *just* those plays with, for example, three (3) curl routes and two (2) flat routes, is **missing the forest for the trees.**

- Rather than examining just those 50 instances, this study looks at routes that include at least one (1) of each of the concepts, regardless if other routes are included.

- For example, the following route combination is included in this expanded definition:

  Dig, Flat – Left, Curl, Flat – Right, Curl

- The core components (highlighted in yellow) of the most popular route combinations from the previous slide are included, as well as "wrinkles" (ie., the 'Dig' route).

- This allows for this study to compensate for the creativity that we often see from offensive coordinators.

UPDATED ROUTE NUMBERS

After filtering with `*str_detect*` to find those offensive plays that include at least the core components of the route combos in the below chart, we can fully calculate the number of times the route combinations were run in the 2020 season:

Route Combos	Number
Chip – Flat, Chip – Flat, Curl, Curl, Curl	802
Curl, Curl, Curl, Flat – Left, Flat – Right	313
Beneath, Corner, Deep Cross, Run Fake	125
Flat – Left, Flat – Right, Over Ball, Slant, Slant	51
Curl, Curl, Flat – Left, Flat – Right, Over Ball	48
Out, Run Fake, Screen - Bubble	27
Curl, Curl, Flat – Left, Flat - Right	25

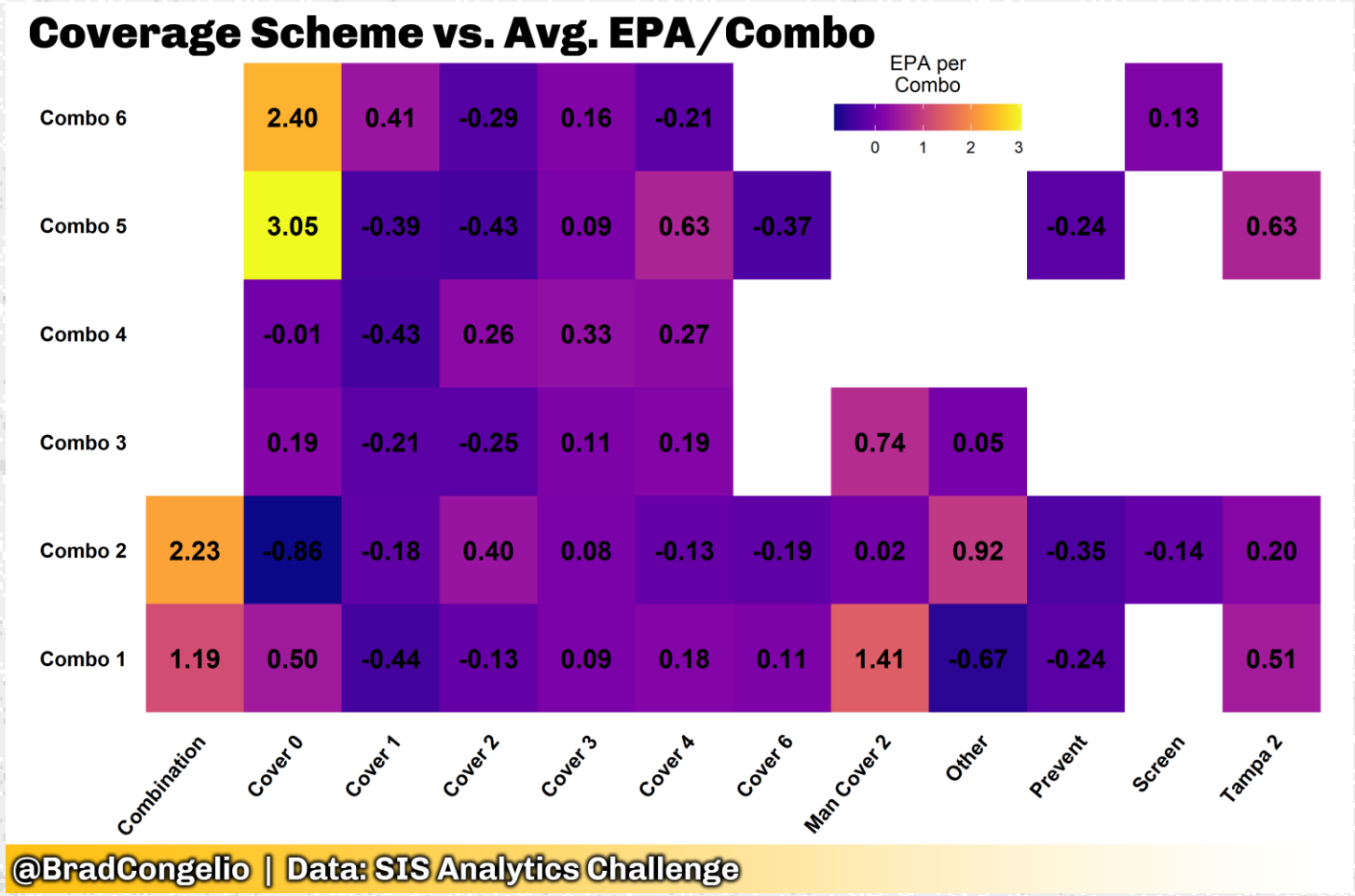
*** Route combinations that included: **(1.) only screen passes** or **(2.) one route and a run fake** were excluded. For example, the "Run Fake, Screen – Bubble" combination was run 143 times in 2020. However, given the scope of this Analytics Challenge, I do not believe those truly constitute route combinations.*

```
##########
##PUTTING TOGETHER ALL PLAYS THAT INCLUDE AT LEAST THE COMMON COMBOS
##########

route.epas <- most.common %>%
  filter(str_detect(route.combos, "\\bCurl\\b"),
         str_detect(route.combos, "\\bFlat - Left\\b"),
         str_detect(route.combos, "\\bFlat - Right\\b"))

routes.combo2 <- most.common %>%
  filter(str_detect(route.combos, "\\bChip - Flat\\b"),
         str_detect(route.combos, "\\bCurl\\b"))

routes.combo3 <- most.common %>%
  filter(str_detect(route.combos, "\\bBeneath\\b"),
         str_detect(route.combos, "\\bCorner\\b"),
         str_detect(route.combos, "\\bDeep Cross\\b"),
         str_detect(route.combos, "\\bRun Fake\\b"))

routes.combo4 <- most.common %>%
  filter(str_detect(route.combos, "\\bFlat - Left\\b"),
         str_detect(route.combos, "\\bFlat - Right\\b"),
         str_detect(route.combos, "\\bOver Ball\\b"),
         str_detect(route.combos, "\\bSlant\\b"))

routes.combo5 <- most.common %>%
  filter(str_detect(route.combos, "\\bCurl\\b"),
         str_detect(route.combos, "\\bFlat - Left\\b"),
         str_detect(route.combos, "\\bFlat - Right\\b"),
         str_detect(route.combos, "\\bOver Ball\\b"))

routes.combo6 <- most.common %>%
  filter(str_detect(route.combos, "\\bOut\\b"),
         str_detect(route.combos, "\\bRun Fake\\b"),
         str_detect(route.combos, "\\bScreen - Bubble\\b"))
```

TOP- LEVEL RESULTS:

AVG. EPA/COMBO VS. COVERAGE SCHEME

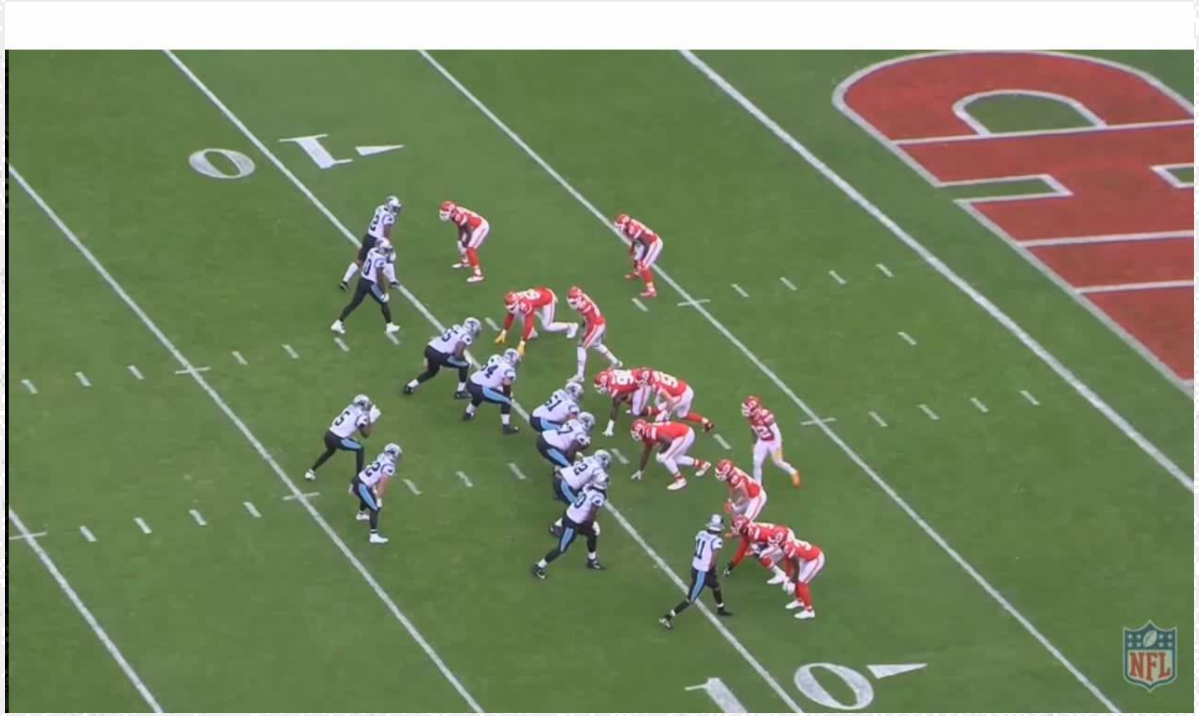
- To start our examination, we will use EPA (expected points added) per route combo against the coverage scheme it was run against.



- The most successful route combination was 'Combo 5' against a Cover 0 defensive scheme with an EPA of +3.05.
- Combo 5 is:
  - Curl
  - Flat – Left
  - Flat – Right
  - Over Ball

AN EPA OF +3 IS CAUSE FOR CONCERN

- Unfortunately, an EPA of +3 is immediately suspect.
- The unusually high number is the result of route combo 5 being employed against Cover 0 just <u>once</u> during all of the 2020 season.
- What happened? A T. Bridgewater TD pass.

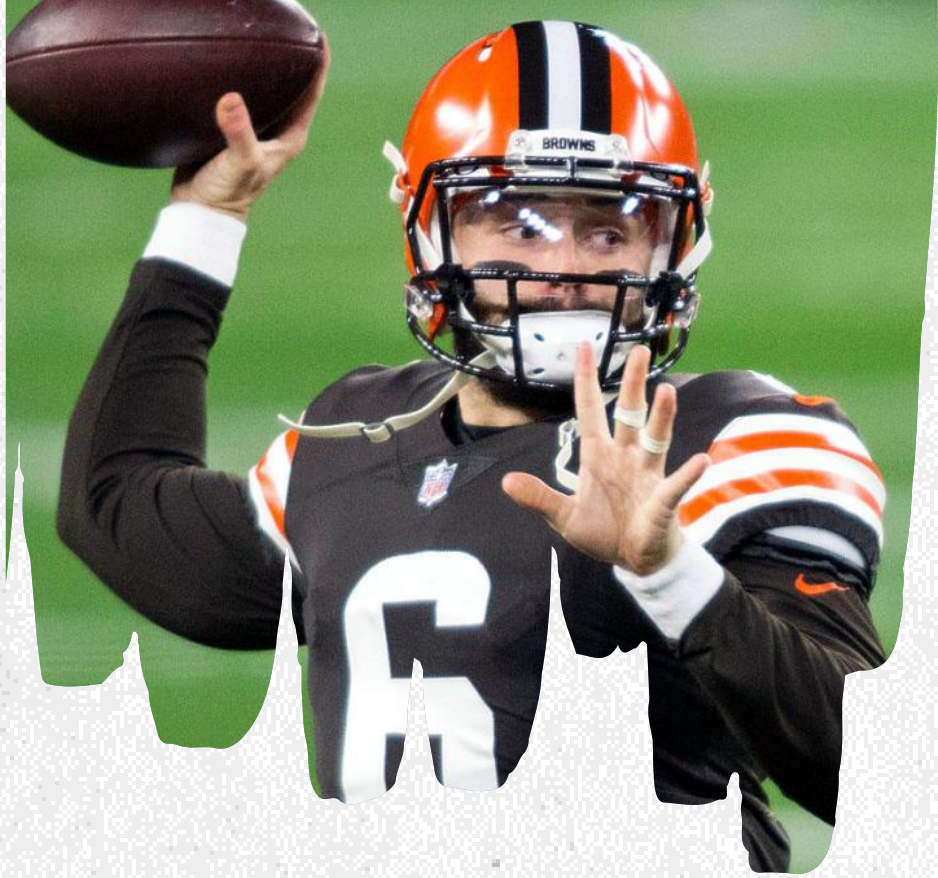


CREATING AN ADJUSTED EPA VIA WEIGHTED MEANS

$$\frac{\sum_{i=1}^{n}(x_i \cdot w_i)}{\sum_{i=1}^{n} w_i}$$

- To correct this issue, we must take into consideration the **total number of times any one combo was run against a coverage scheme.**

- To do this, I created an `*Adjusted EPA*` metric that values those combos/scheme combinations with higher amounts of attempts.

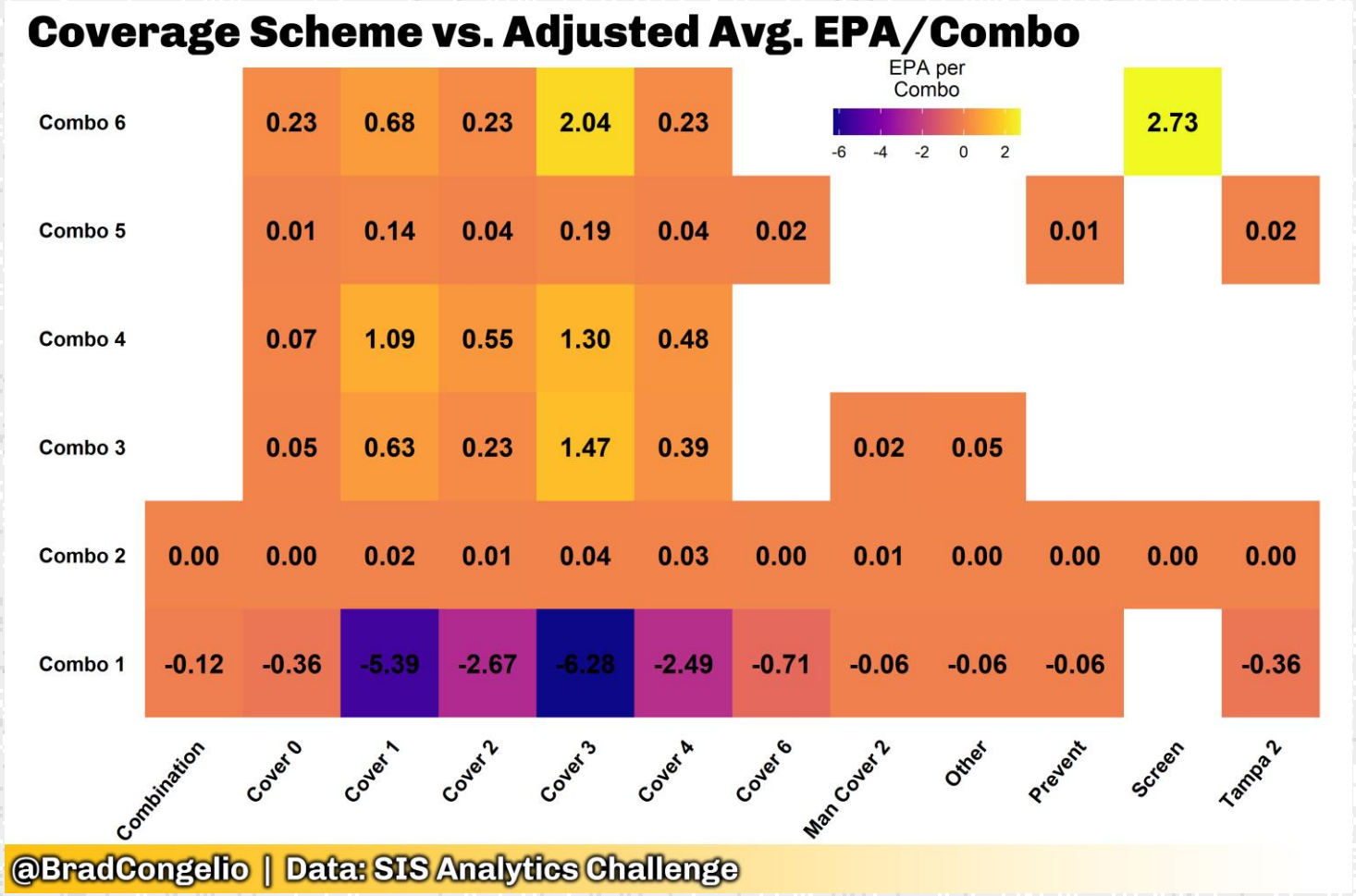


ADJUSTED TOP-LEVEL RESULTS:

ADJUSTED EPA/COMBO VS. COVERAGE SCHEME

- The correlation after creating the adjusted EPA is drastically different – in a good way.

- We still have two 2+ EPAs, but I am now comfortable with them as they are weighted by total plays.



- The most successful route combination was 'Combo 6' against a "Screen" defensive scheme.
- Combo 6 is:
  - Out
  - Run Fake
  - Screen - Bubble

FINAL ADJUSTED RESULTS:
WHICH ROUTE COMBOS PERFORM BEST AGAINST COVERAGE TYPE?

Defensive Scheme	Most Successful Route Combo	Adj. EPA
Combination	Chip – Flat, Curl	0.00
Cover 0	Out, Run Fake, Screen - Bubble	0.23
Cover 1	Flat – Left, Flat – Right, Over Ball, Slant	1.09
Cover 2	Flat – Left, Flat – Right, Over Ball, Slant	0.55
Cover 3	Out, Run Fake, Screen - Bubble	2.04
Cover 4	Flat – Left, Flat – Right, Over Ball, Slant	0.48
Cover 6	Chip – Flat, Curl	0.00
Man Cover 2	Beneath, Corner, Deep Cross, Run Fake	0.02
Other	Beneath, Corner, Deep Cross, Run Fake	0.05
Prevent	Curl, Flat – Left, Flat – Right, Over Ball	0.01
Screen	Out, Run Fake, Screen - Bubble	2.73
Tampa 2	Curl, Flat – Left, Flat – Right, Over Ball	0.02

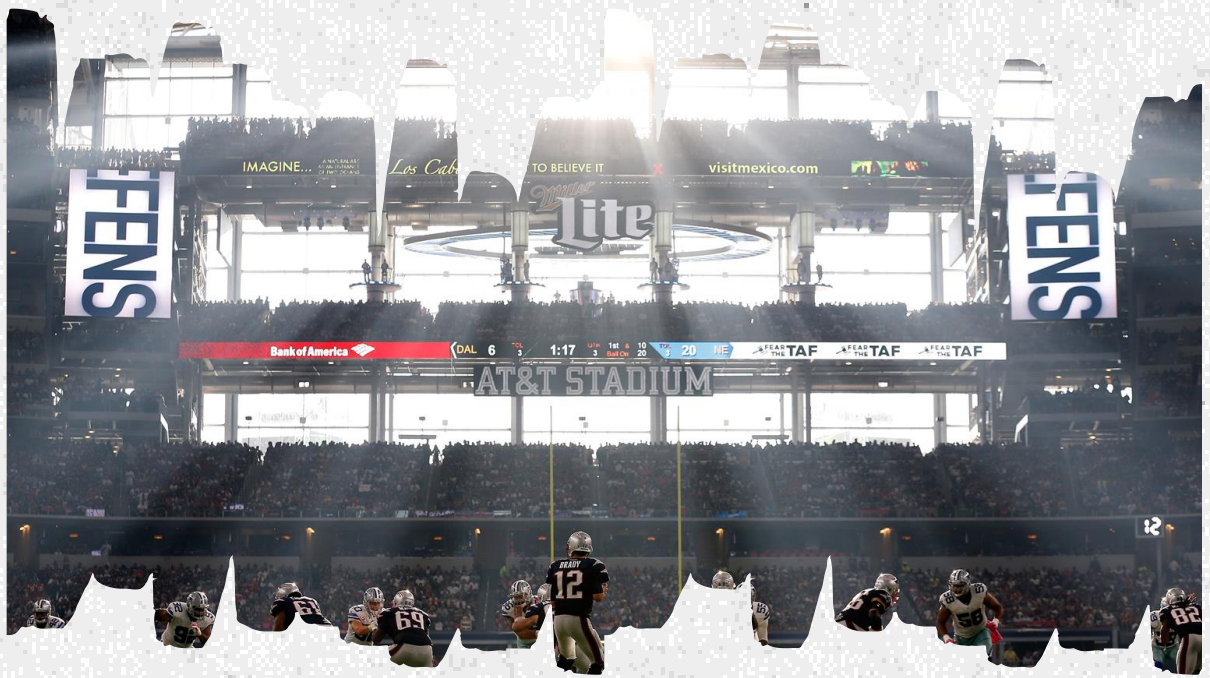
WHERE WE STAND

- We have identified the most popular route combinations from the 2020 NFL season.
- We have determined which of the route combinations performs best against each NFL coverage scheme.
- *But have we created ontological weight?*

Yes.

- From a schematic standpoint, with our findings thus far:
  - Offensive coordinators can call plays with specific route combinations based on defensive coverage scheme.
  - Quarterbacks can audible/hot route to successful route combos once they have diagnosed the defensive coverage scheme.

This is a great example of "action research" as it can be transferred to the field of play and readily assist teams in winning games.



WHAT NOW?

- With the basis framework of the 2021 SIS Football Analytics Challenge complete, we can:
  - Explore the role of wide receivers in the success of route combos.
  - Explore the role of corner/defensive backs in the success of route combos.

… and most importantly:

  - Use this information to create optimal mismatches against the opponent on passing plays.

THE IMPACT OF WIDE RECEIVER POSITIONING

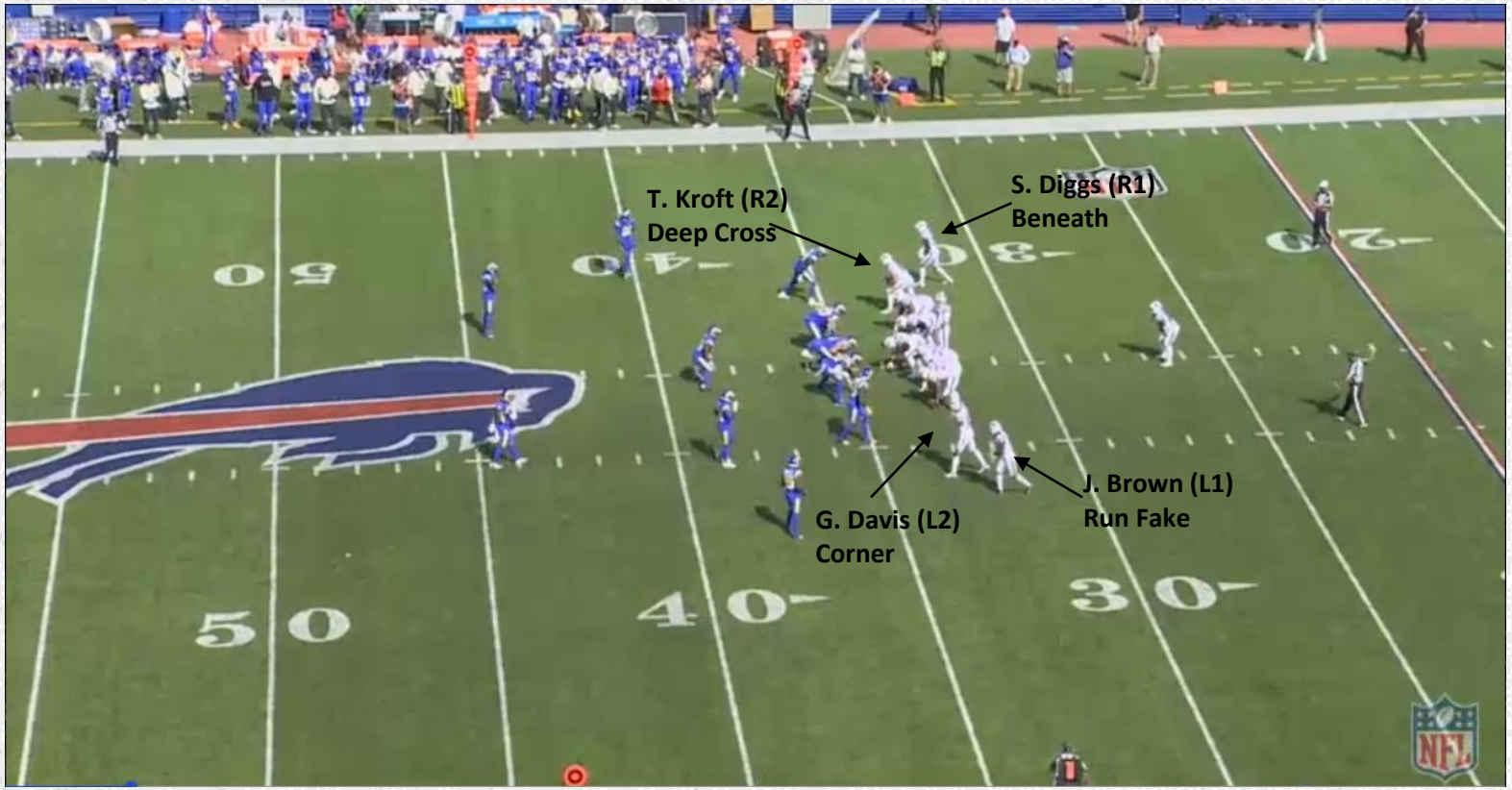
- Which side of center (SOC) and order creates the most success against various coverages?

For example: *Bills vs. Rams (Week 3)*

Bills: Route Combo #3 (Run Fake, Deep Cross, Beneath, Corner, Run Fake)

Rams: Cover 3

Result: 39-yard pass to G. Davis (EPA = 2.18)



Sports Info Solutions
Football Analytics Challenge

AVG. WIDE RECEIVER EPA

SOC VS. FIELD ORDER | ROUTE COMBO 3 VS. COVER 3

L-Set WRs **Offensive Line** **R-Set WRs**

SOC	Order	Adj. EPA
L	1	0.10
L	2	0.11
L	3	0.89
R	1	0.22
R	2	-0.02
R	3	0.09

- When running Route Combo #3 (Run Fake, Deep Cross, Beneath, Corner) against a Cover 3 defensive scheme (such as the Bills against the Rams in week #3 of 2020), the left-side wide receiver set closest to the offensive line (L3) provided the most adjusted EPA during the 2020 NFL season.
- This is proof of theory that we can determine which receiver is likely to produce the most adjusted EPA based on not only side of center and field order, but also route combination and defensive scheme.

MERGING DATA

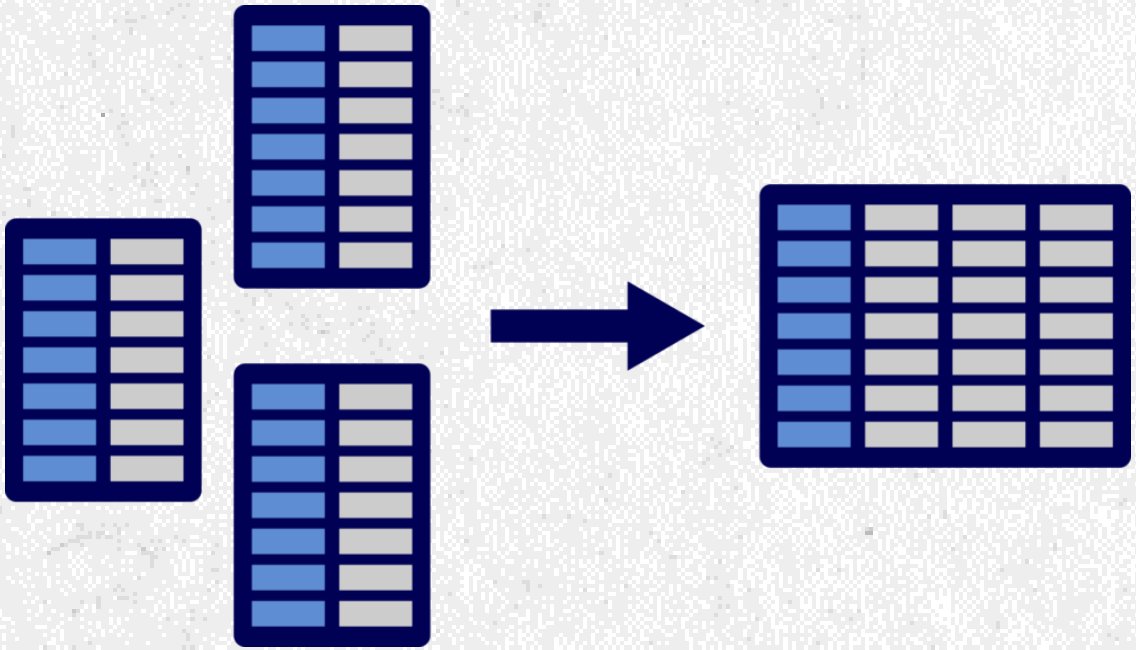
In order to take our analysis even further, I'd like to examine which defensive players perform worst against what route combos and out of which coverage scheme.

To do this, I will be merging in 2020 NFL play-by-play data from the nflfastR project. However, there are significant issues to take care of first:

1. The `GameID` in the SIS data and the `game_id` in `nflfastR` are not the same … so a straight `left_merge` was not possible.

2. `nflfastR` uses "posteam" and "defteam" while the SIS data uses "OffensiveTeam" and "DefensiveTeam."

After much work in attempting to mutate the data to line up correctly, it was discovered that the 'fuzzyjoin' function could easily merge the data together by comparing the 'play description' used in both datasets.



RESULTS:

WHICH DEFENSIVE PLAYERS PERFORM WORST OUT OF SPECIFC SCHEMES AGAINST ROUTE COMBOS?

Defensive Scheme	Route Combo	Player	Air EPA Allowed
Cover 0	Combo 1	M. Epps	0.10
Cover 1	Cover 1	D. Slay	4.95
Cover 2	Combo 2	A. Taylor	2.98
Cover 3	Combo 2	T. Diggs	4.25
Cover 4	Combo 2	X. Howard	3.01
Cover 6	Combo 2	T. Mullen	1.50
Man Cover 2	Combo 2	J. Ward	2.56
Prevent	Combo 2	M. Fitzpatrick	0.28
Screen	Combo 6	D. Savage	-1.02
Tampa 2	Combo 2	N. Kwiatkoski	0.42

** *Assumption:* *given the structure of the data provided via SIS and nflfastR, the Air EPA was calculated by determining the player who made the tackle on a completed pass. While not 100-percent accurate, it is reasonable to conclude that, in a statistically significant amount of the cases, the player who made the tackle was also assigned coverage to the receiver.*

Coach Tomlin is happy because we now have the needed information to attack the opposing defense's weakest link out of any coverage scheme they may show.

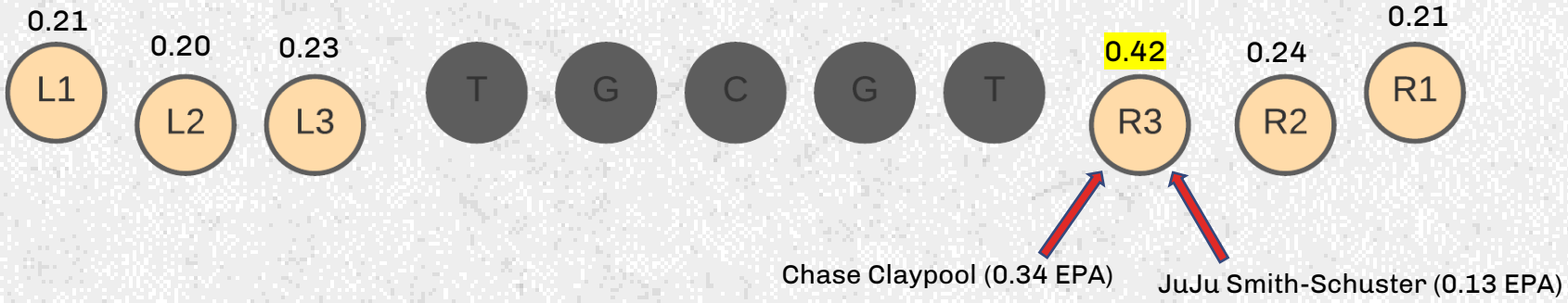
SO, WHAT NOW?

- We have identified the most popular route combinations from the 2020 NFL season.
- We have determined which of the route combinations performs best against each NFL coverage scheme.
- We can statistically determine which receiver is likely to provide the most EPA based on their location (side of center and out-to-inside) against specific coverage schemes.
- We can theoretically determine which defender is the "weakest" link in any coverage scheme against any of the popular route combos

NOW, TO PUT IT ALL TOGETHER.

STEELERS VS. EAGLES 2021 SEASON: WEEK #1

- During the 2020 season, the most used coverage scheme by the Philadelphia Eagles was the Cover 1.
- Against the Cover 1 defensive scheme in 2020, the Pittsburgh Steelers most often used popular NFL Route Combo #1 (as determined earlier in this presentation): Curl, Flat – Left, Flat – Right
- When running the Cover 1 defensive scheme, the weakest Eagles' defender in 2020 was Avonte Maddox with a mean air EPA of 0.87.
- When running NFL popular route combo 1 in 2020, the Steelers had the EPAs as listed in the below graph by SOC and outside-to-inside order.



- In this hypothetical scenario during week 1 of the upcoming 2021 season, the Steelers should be seeking to get either Claypool or Smith-Schuster into the R3 receiver position while covered by Avonte Maddox.
  - If Claypool is lined up in R3 in this situation, a "Flat – Right" is his highest-grossing EPA route.
  - If Smith-Schuster is lined up in R3 in this situation, a "Curl" is his highest-grossing EPA route.
- BASED ON THE ANALYSIS DONE IN THIS RESEARCH, THIS PROVIDES THE BEST CHANCE FOR NET-POSITIVE EXPECTED POINTS ADDED.

THIS WAS A LEAGUE-WIDE APPROACH

IT CAN JUST AS EASILY BE BROKEN DOWN BY TEAM

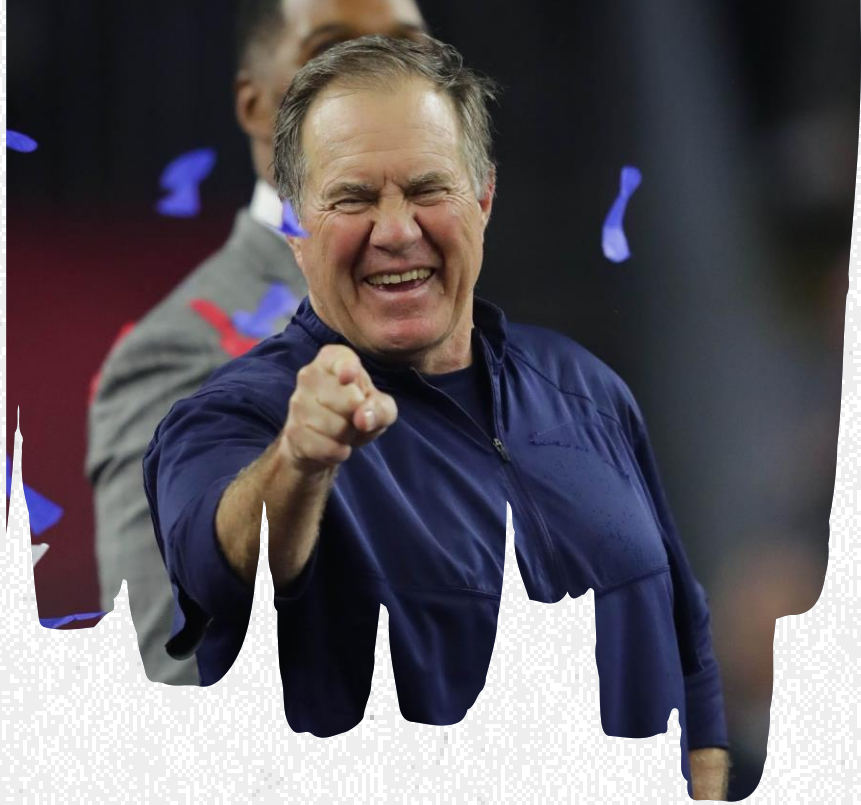
Steelers' Top 3 Combos

Route Combos	Best EPA	Defensive Scheme
Out, Run Fake, Slant	0.16	Cover 4
Flat – Left, Flat – Right, Go/Fly, Out, Slant	0.02	Cover 3
Flat – Right, Screen – Quick, Slant	0.11	Cover 2

- At this point, we can start breaking down the side-of-center and field position for each Steelers' WR in each popular route combo to determine where they should be lined up for best EPA (and against which defensive scheme).
- And, as seen in the previous slide, we can break down the weakest link on the opposing defense, on a week-by-week basis, by coverage scheme and alike combos to strategically create personnel mismatches.

CONCLUSION

- The is no denying that the NFL is a micro match-up driven league. The individual matchups that occur during each play can have a significant impact on the outcome of the game.
- In the end, teams that can better schematically create optimal matchups are more likely to win the game.
- With that in mind, this research has showcased:
  - *How a team's most used route combinations can be identified.*
  - *How to calculate the success of each route combination against all defensive schemes.*
  - *Which receiver has the highest EPA for each route against the defensive schemes and, importantly, where they should line up on the field and which side of center.*
  - *How to identify the weakest pass defender for each route out of each defensive scheme to assist in creating strategic optimal matchups.*
- **Combined, it has constructed a data-driven approach to win a micro-strategy battle. And, to that end, the more micro-strategy battles a team claims, the more likely it is they will win the game.**



LIMITATIONS

- With just one year of data to examine, the results are not likely as robust as they would be if multiple years were available.
- Offensive/Defensive Coordinators are not included in the data. Given that the personality of an offense and/or defense is largely represented by the coaching persona of the coordinator, including this information in the data (and weighing it) would be beneficial.
- Data did not include definitive names of defenders for targeted passes. Including such information would allow 100-percent accurate analysis as this research, as noted, defined the “closest defender” as the one that made the tackle.
- Without player tracking data, it is not possible to determine the distance from receiver-to-defender, or vice-vera, on pass plays. With this information, we could start calculating the precise depth a “Curl” route should be, or when to break on the “Flat” routes against specific defenders and against specific defensive coverages to further optimize personnel-based mismatches.

FUTURE RESEARCH

- Once more than one year of data is public, merging the publicly available list of offensive coordinators to team, by year, to better refine team route combos and personnel groupings.
- Conduct the study while also including situational-based metrics, specifically red zone route combinations and defensive schemes.
- Manipulate the data into a regression-based decision tree that includes offensive personnel, route combos, defensive scheme, and defensive personnel to more efficiently compute the optimal positioning and routes for matchup purposes.

THANK YOU

@BradCongelio

KUTZTOWN UNIVERSITY

DATA SOURCES:

1. 2021 SIS Football Analytics Challenge
2. nflfastR
3. NFL Game Pass (for video and video capture)