SIS Analytics Challenge

Sam Hoppen

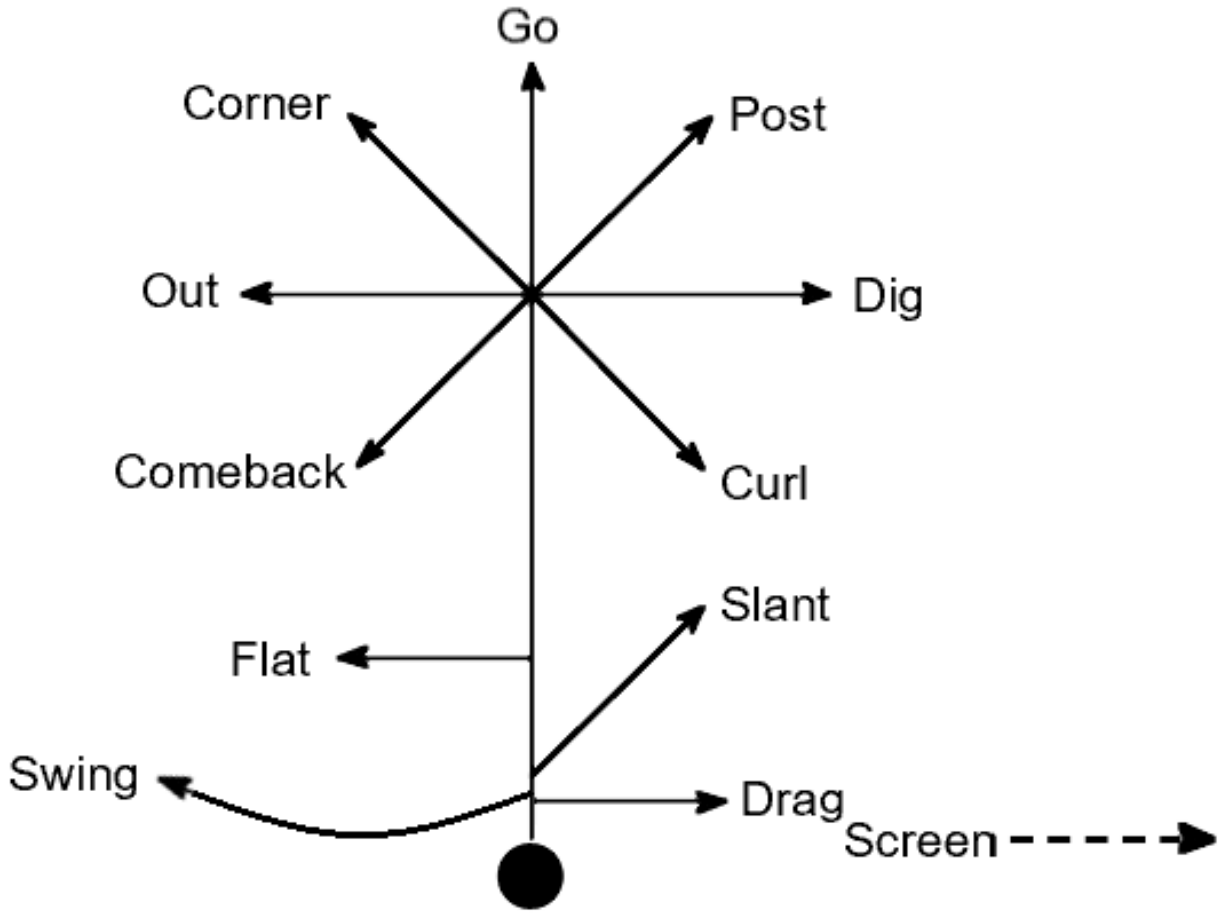
July 2021

Introduction

- Goals:
  - Identify which route combinations were most popular in the NFL in 2020
  - Of these route combinations, evaluate which perform best against each coverage type
- Evaluating route combinations is difficult: how do we measure the impact that they had on the success of a play?
- Since EPA (Expected Points Added) is a function of team-level success on a play, I used it as my primary method of measuring route success.
- I calculated success rates for each route and route combination, then built a model to measure “expected success” of a play and ultimately the expected success rate of various route combinations
- Results in Success Rate Over Expected (SROE) for each route and route combination

Assumptions: narrowing down the route combinations

- With original route descriptions, 12,000+ unique route combos exist
- Leveraged domain knowledge and input from industry experts to group routes into one of 13 categories based on a general route tree (see right image)
  - Reduced unique route combos to ~3,500
  - Included a 14th “other” category for variables that did not fit well into a category)
  - *See appendix for full list of route conversions*

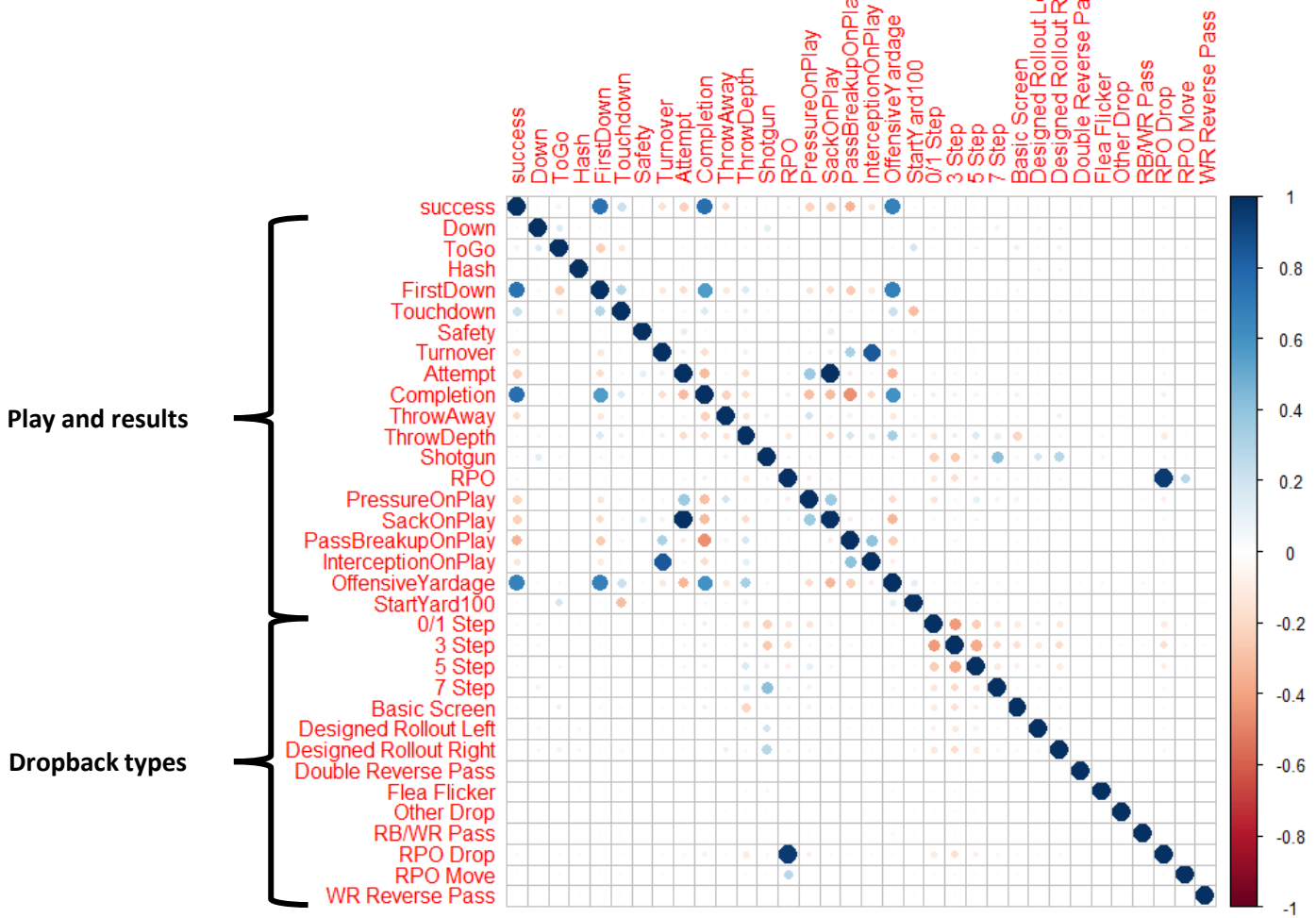


Final assumptions: plays used & evaluating success

- Only evaluated passing plays in which at least one route was run
- In calculating success rates, success defined as a play with an EPA > 0
- Removed plays from dataset in which any NA values existed in the variables I used
- Did not use the following variables in model building as they are more descriptive than predictive:
  - Season
  - Week
  - Offensive/Defensive Team
  - Score
  - Quarter
  - Time Left
  - Play Description

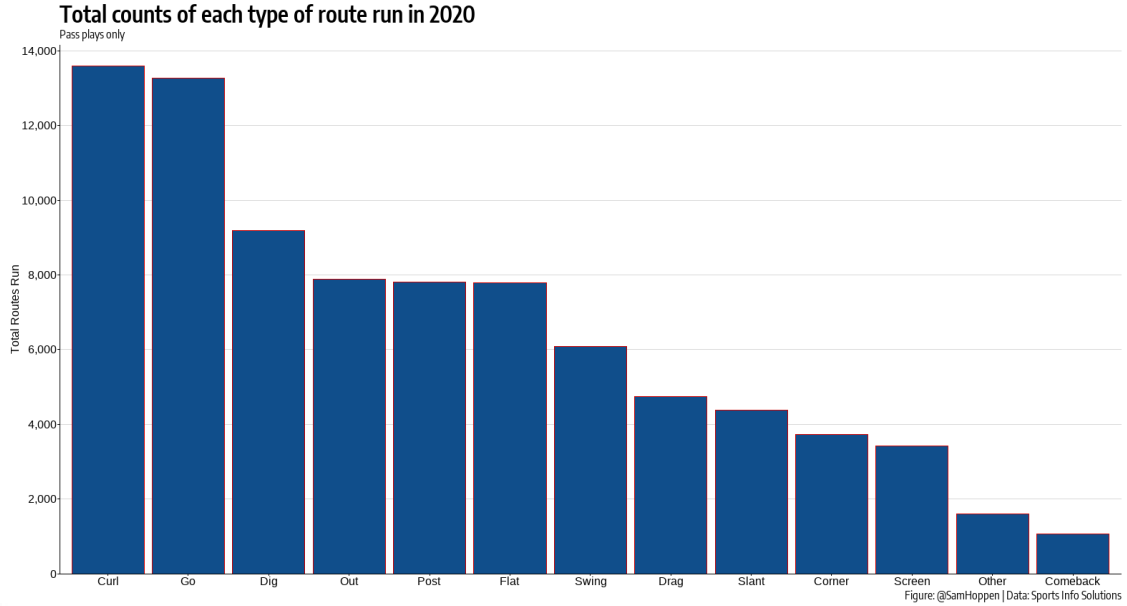
Exploratory data analysis: what do correlations tell us?

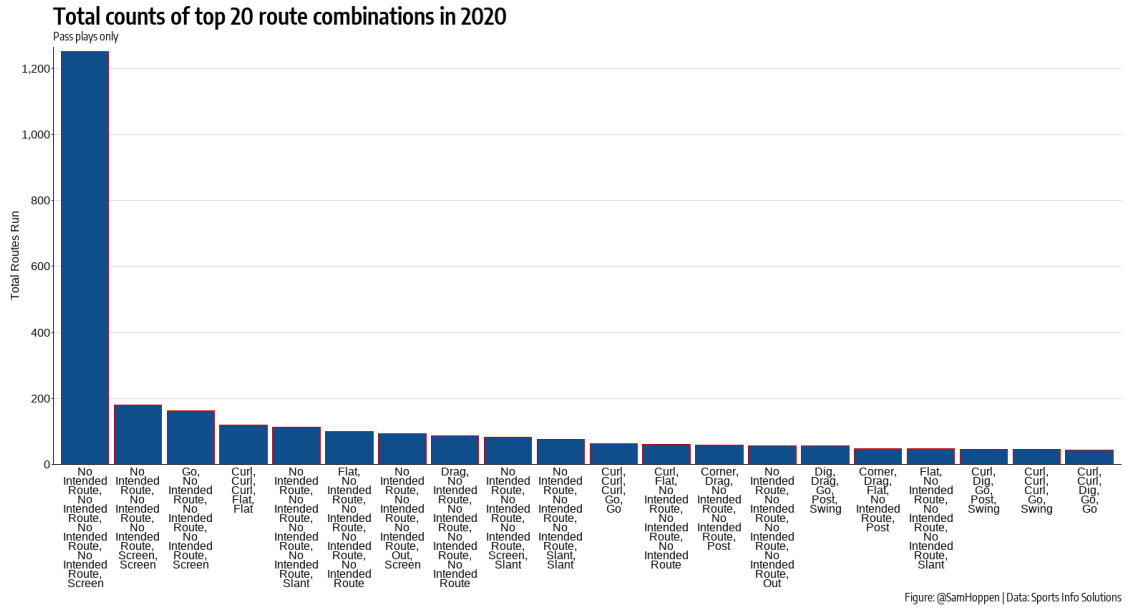
- Primarily evaluated correlation with "success," my target variable
- Difficult to glean major implications from this as many variables are binary, however, some things stood out:
  - Unsurprisingly, FirstDown, Completion, and OffensiveYardage are heavily correlated with success
  - Dropback correlations with success (and most other variables) are nonexistent, so **removed from analysis moving forward**



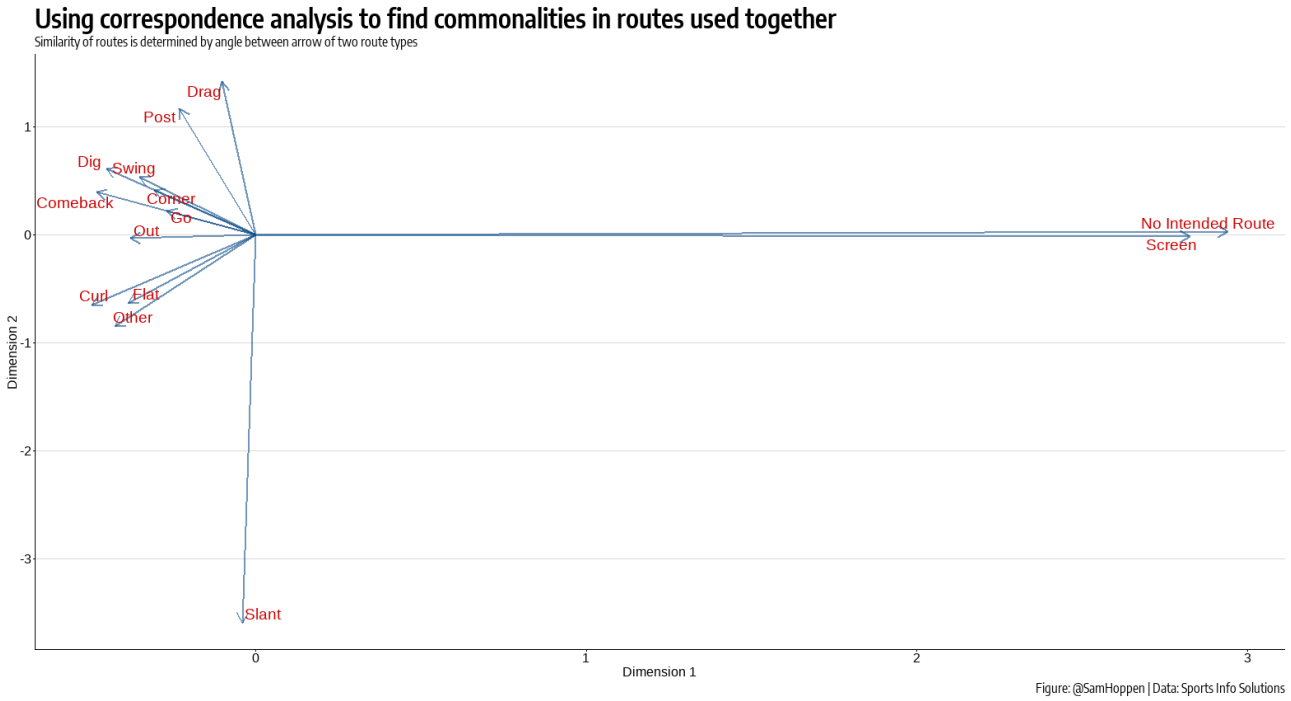
More exploratory data analysis: which routes and route combinations were used most often?

- Curl and Go routes are the most used routes, followed closely by a Dig route
- One player running a screen route with the remaining four receivers not running a route (likely blocking) is far and away the most common route combo
- Several of the most-used route combinations include multiple players that don't run an intended route
- Three receivers running a Curl route and two receivers running a Flat route is the most-used route combo in which all receivers are running a route





Which routes were used most frequently with other routes?



- Ran a correspondence analysis to identify which routes were run most regularly with other routes
- Key takeaways:
  - When a player runs a screen, most other players do not have a route (typically blocking instead)
  - Several high/low route combinations are run together (i.e. Post/Drag, Go/Comeback, Post/Dig)
  - Go routes almost always include another player running a comeback
- Note: similarity of routes is determined by angle between two route types

Initial steps in building the expected success rate model

- Used logistic regression to get expected success as value between 0-1
- Initially built model with all remaining 19 variables, but this resulted in too many insignificant variables
- Performed feature selection using both forward and backward selection methods, which both produced the same model suggestion:

```
success ~ Down + ToGo + FirstDown + Touchdown + Turnover + Attempt +
    Completion + OffensiveYardage + StartYard100
```

- However, after fitting that model, the beta values for Turnover and Touchdown were larger than the remaining
- Ultimately chose to use lasso regularized regression to counteract the shrink the beta values, coupled with forming feature selection

Lasso model results

- Performed 10-fold cross validation to find lambda values for model
- Used lambda.1se value to ensure most regularized model
- Lasso feature selection left me with 11 total variables
  - Slightly different model produced than forward/backward selection: Lasso includes ThrowDepth, RPO, and Shotgun variables, but excludes Attempt

Now… onto the results!

Down	4.927146e-01
ToGo	-1.853009e-01
Hash	.
FirstDown	2.117936e+00
Touchdown	2.407805e+00
Safety	.
Turnover	-1.229308e+01
Attempt	.
Completion	1.900071e+00
ThrowAway	.
ThrowDepth	3.558275e-03
Shotgun	2.875117e-01
RPO	2.220581e-01
PressureOnPlay	.
SackOnPlay	.
PassBreakupOnPlay	.
InterceptionOnPlay	.
OffensiveYardage	7.399542e-01
StartYard100	-9.091003e-05

Top targeted routes in Success Rate Over Expected (SROE)

Route	Plays Targeted	Success Rate	SROE
Out	1872	57.9%	1.7%
Curl	2647	61.9%	1.5%
Swing	1112	43.9%	0.2%
Corner	502	46.2%	0.1%
Dig	1470	55.4%	−0.0%
Go	1683	38.2%	−0.2%
Post	1194	52.5%	−0.3%
Slant	1259	59.6%	−0.4%
Drag	1324	53.2%	−1.0%
Flat	1501	48.9%	−1.6%
Screen	1746	46.1%	−2.1%

Table: @SamHoppen | Data: Sports Info Solutions

First, I evaluated which individual routes measured well in Success Rate Over Expected (SROE)

- Since targets on Screen routes typically come behind the line of scrimmage, it takes more yards after catch for a play to become a success, making it barely more effective than rushes, which had a Success Rate of 42.2% in 2020
- Meanwhile, Go routes offer more of a boom-or-bust result with a low success rate but median SROE
- When targeted, five route types – Swing, Corner, Go, Flat, and Screen – have a Success Rate worse than 50 percent

What if we layer in coverage schemes?

- A quick-pass route (Flat) shows a great deal of success (3.5% SROE, highest among individual routes) against a Cover 0 blitz
- Out routes show its versatility with above-average SROE against multiple coverage schemes (Cover 1, Cover 2, Cover 3, and Cover 4)
- Despite Screen routes being the worst targeted route in SROE, it shows it can work well against Cover 3 defense

Top 25 targeted routes in Success Rate Over Expected (SROE) versus various coverage schemes

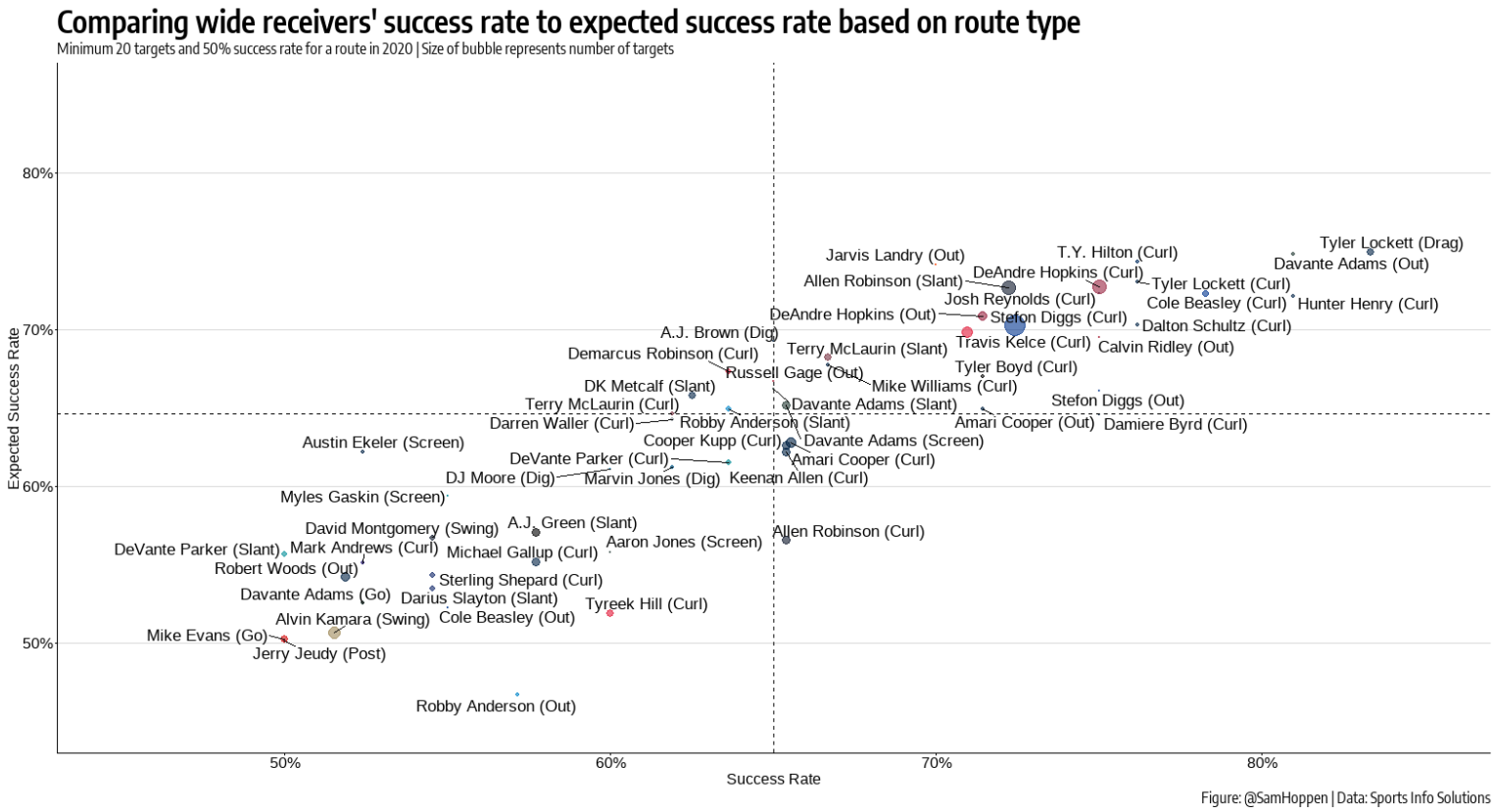
Minimum 50 targeted plays

Route	Coverage Scheme	Plays Targeted	Success Rate	SROE
Flat	Cover 0	64	67.2%	3.5%
Swing	Cover 2	154	42.9%	3.1%
Curl	Cover 3	992	65.7%	3.1%
Out	Cover 2	176	56.2%	2.9%
Screen	Cover 3	60	56.7%	2.7%
Out	Cover 3	599	60.4%	2.1%
Out	Cover 4	326	59.5%	1.9%
Curl	Cover 1	410	55.4%	1.9%
Out	Cover 1	482	57.5%	1.9%
Slant	Cover 3	263	65.8%	1.7%
Drag	Screen	109	49.5%	1.3%
Curl	Cover 2	370	64.1%	1.3%
Slant	Cover 2	101	62.4%	1.3%
Comeback	Cover 1	64	59.4%	1.1%
Swing	Cover 1	233	42.5%	1.1%
Dig	Man Cover 2	99	48.5%	1.0%
Curl	Tampa 2	78	67.9%	1.0%
Comeback	Cover 3	80	63.7%	0.6%
Flat	Cover 2	174	51.7%	0.4%
Out	Man Cover 2	122	54.1%	0.4%
Curl	Man Cover 2	71	49.3%	0.3%
Out	Cover 6	56	60.7%	0.0%
Dig	Cover 3	430	57.4%	0.0%
Dig	Cover 1	386	55.4%	−0.0%
Go	Man Cover 2	83	30.1%	−0.1%

Table: @SamHoppen | Data: Sports Info Solutions

Evaluating which receivers' routes score high in Success Rate

- Curl and Out routes dominate the top-right quadrant
- Davante Adams, Tyler Lockett, and Stefon Diggs with multiple routes among the top players
- Stefon Diggs was targeted on a curl 58 times, the league lead for a single route



Identifying top route combos in Success Rate Over Expected (SROE)

- Clearing out the defense with three receivers while two receivers come back on curls proves to be the most valuable route combination (Curl, Curl, Go, Go, Go)
- At least one curl and one go route is included in nine of the top 15 route combinations
- A Curl, Curl, Go, Other, Swing route combination offers a 64% Success Rate (2nd highest) but only 3.2% SROE
- Interestingly, two route combos that feature at least three receivers not running an intended route are among the top 20 route combos in SROE
- It's clear that a high Success Rate does not necessarily indicate a high SROE

Top 25 targeted route combinations in Success Rate Over Expected (SROE)

Minimum 25 plays used

Route Combo	Plays	Success Rate	SROE
Curl, Curl, Go, Go, Go	25	68.0%	12.6%
Curl, Curl, Go, Out, Swing	31	54.8%	10.5%
Flat, Go, Out, Out, Slant	26	61.5%	7.2%
Curl, Flat, Go, Go, Out	26	42.3%	6.6%
Curl, Go, Go, Go, Post	35	48.6%	6.3%
Curl, Dig, Go, Go, Swing	30	56.7%	5.9%
Flat, Flat, Other, Slant, Slant	30	63.3%	5.2%
Curl, Dig, Dig, Go, Out	29	44.8%	4.5%
Dig, Flat, Go, Out, Post	25	52.0%	4.1%
Curl, Curl, Curl, Curl, Curl	32	62.5%	3.6%
Curl, Curl, Curl, Go, Swing	44	56.8%	3.4%
Curl, Curl, Go, Other, Swing	25	64.0%	3.2%
Dig, Dig, Go, Out, Swing	29	41.4%	3.2%
Curl, Dig, Dig, Go, Go	25	48.0%	3.0%
Drag, Go, No Intended Route, No Intended Route, Post	35	54.3%	2.7%
Curl, Curl, Dig, Go, Go	42	54.8%	2.6%
Curl, Curl, Curl, Curl, Swing	26	53.8%	2.6%
Curl, Curl, Curl, Go, Post	31	54.8%	2.3%
No Intended Route, No Intended Route, No Intended Route, Slant, Slant	29	48.3%	2.1%
Curl, Dig, Flat, Go, Slant	27	59.3%	1.9%
Curl, Dig, Dig, Go, Swing	27	44.4%	1.9%
Curl, Dig, Go, Post, Swing	44	38.6%	1.9%
Curl, Curl, Other, Out, Swing	31	41.9%	1.7%
Dig, Go, Out, Post, Swing	33	51.5%	1.6%
Curl, Curl, Curl, Dig, Go	28	53.6%	1.5%

Table: @SamHoppen | Data: Sports Info Solutions

Again, adding coverage schemes provides more context for route combinations

- A Drag and Go (with no other receivers running routes) against a Screen defense ranks highest in both Success Rate and SROE
- Curl, Curl, Go, Out, Swing was 2nd in coverage-agnostic SROE, and is 2nd in coverage-specific SROE (against Cover 3)
  - This makes sense, given Cover 3 is weaker to shorter routes
- Curl routes are heavily featured among the top route combos in SROE
- Cover 3 defenses look like a sieve as it allows 10 of the 15 highest SROE, most of which feature a heavy dose of short-area route combos

Top 25 route combos in Success Rate Over Expected (SROE) versus various coverage schemes

Minimum 10 plays used

Route Combo	Coverage Scheme	Plays	Success Rate	SROE
Drag, Go, No Intended Route, No Intended Route, No Intended Route	Screen	10	90.0%	23.4%
Curl, Curl, Go, Out, Swing	Cover 3	10	60.0%	15.1%
Curl, Go, Go, Go, Post	Cover 3	14	50.0%	15.1%
Curl, Curl, Out, Out, Out	Cover 3	10	80.0%	13.3%
Curl, Dig, Flat, Go, Slant	Cover 1	12	66.7%	10.8%
Curl, Curl, Curl, Flat, Flat	Cover 2	18	77.8%	8.9%
Drag, No Intended Route, Post, Post, Swing	Cover 3	11	54.5%	8.6%
Curl, Dig, Flat, Go, Out	Cover 3	11	54.5%	8.5%
Dig, Flat, Go, Out, Post	Cover 3	10	50.0%	8.2%
Curl, Curl, Dig, Go, Out	Cover 3	12	75.0%	8.1%
Curl, Curl, Dig, Go, Go	Cover 3	18	72.2%	7.9%
Curl, Curl, Curl, Curl, Go	Cover 3	10	60.0%	7.0%
Curl, Curl, Flat, Go, Slant	Cover 1	10	50.0%	6.4%
No Intended Route, No Intended Route, No Intended Route, Out, Screen	Cover 3	16	68.8%	6.4%
Curl, No Intended Route, No Intended Route, No Intended Route, Screen	Screen	22	68.2%	6.4%
Curl, Curl, Curl, Out, Swing	Cover 3	10	70.0%	6.2%
Curl, Curl, Curl, No Intended Route, No Intended Route	Cover 3	14	71.4%	5.8%
Go, No Intended Route, No Intended Route, Screen, Slant	Screen	11	63.6%	5.6%
Curl, Curl, Curl, Dig, Go	Cover 4	10	70.0%	5.2%
Corner, Drag, No Intended Route, No Intended Route, Post	Cover 3	23	52.2%	4.5%
Curl, Dig, Flat, Go, Post	Cover 3	14	50.0%	4.3%
Curl, Curl, Curl, Out, Out	Cover 3	10	60.0%	4.2%
Drag, Go, No Intended Route, No Intended Route, Post	Cover 1	10	70.0%	3.8%
Curl, Curl, Curl, Go, Swing	Cover 3	11	81.8%	3.8%
Curl, Curl, Curl, Flat, Flat	Cover 3	38	52.6%	3.6%

Table: @SamHoppen | Data: Sports Info Solutions

Conclusions

- An Out route is the most effective route, regardless of coverage scheme
- One Drag and one Go route with the remaining three receivers not running a route against a Screen coverage scheme offers the highest Success Rate (90%) and SROE (23.4%) among all route combos
- Cover 3's weakness with shorter routes shows with high SROE values from individual routes (Curl, Screen) as well as route combos that feature many of those types of routes

Limitations 

- Only one year of data, which didn’t allow for better model calibration or more definitive results
- High number of unique route combinations didn’t allow for specificity in analysis
- Lack of player tracking data doesn’t allow for separation-based measurement

Future Analysis 

- With tracking data, evaluate success based on receiver separation by route and coverage scheme
- Measure situation-based (i.e., red zone, third and short) route success

Thank you!

Appendix

- Full coding script used for analysis:
  https://bit.ly/3Bu6eiv

New route descriptions:

Original Route Description	New Route Description
Curl	Curl
Out	Out
Slant	Slant
Dig	Dig
Drag	Drag
Deep Cross	Post
Flat - Right	Flat
Flat - Left	Flat
Screen - RB	Screen
Go/Fly	Go
Post	Post
Corner	Corner
Fade	Go
Screen - Tunnel	Screen
Chip - Flat	Flat
Screen - Bubble	Screen
Seam	Go
Beneath	Drag
Whip	Dig
Swing - Left	Swing
Over Ball	Other
Check & Release	Swing
Swing - Right	Swing
Comeback	Comeback
Wheel	Swing
Fade - Back Shoulder	Go
Angle	Angle
Screen - Quick	Screen
Chip - Curl	Curl
Quick	Screen
Screen - TE	Screen
Jet Sweep Pass	Drag
Hitch & Go	Go
Out & Up	Go
Chip - Drag	Drag
Post Corner	Corner
Corner Post	Post
Sluggo	Go
Jerk	Dig
Chip - Seam	Go
Screen - Drag	Screen
Screen - Beneath	Screen
Screen - Shovel	Screen
Stick - Nod	Curl
Leak	Dig
Chip	Other
Pick	Other
Blocking	No Intended Route
Run Fake	No Intended Route