Sports Info Solutions Analytics Challenge

JOSEPH CHERNAK-GENERAL TRACK ENTRY

Questions to Answer

SIS posed two questions as a part of the general track entry:

1. Which route combinations were most popular in the NFL in 2020?
2. Of these route combinations, which perform best against each coverage type?

As a first step in answering these questions, I began by asking what constitutes a route combination? I decided that a route combo is:

- 2 or more routes that are designed within the structure of the play to interact with each other. This includes patterns that begin on opposite sides of the center (for example mesh).

So how do we incorporate this assumption into our work?

Route Classification Changes	
OLD ROUTE	**NEW ROUTE**
Fade - Back Shoulder	Go
Fade	Go
Screen - Bubble	Screen
Screen - TE	Screen
Screen - Quick	Screen
Screen - Quick	Screen
Screen - Tunnel	Screen
Screen - Beneath	Screen
Screen - Shovel	Screen
Screen - Drag	Screen
Screen - RB	Screen
Chip - Flat	Flat
Chip - Curl	Curl
Chip - Drag	Drag
Chip - Seam	Seam
Stick - Nod	Seam
Beneath	Drag
Hitch & Go	DoubleMove
Out & Up	DoubleMove
Go/Fly	Go
Over Ball	SitOverMiddle
Corner Post	COPO
Post Corner	POCO
Deep Cross	DeepCross
Jet Sweep Pass	JetSweepPass

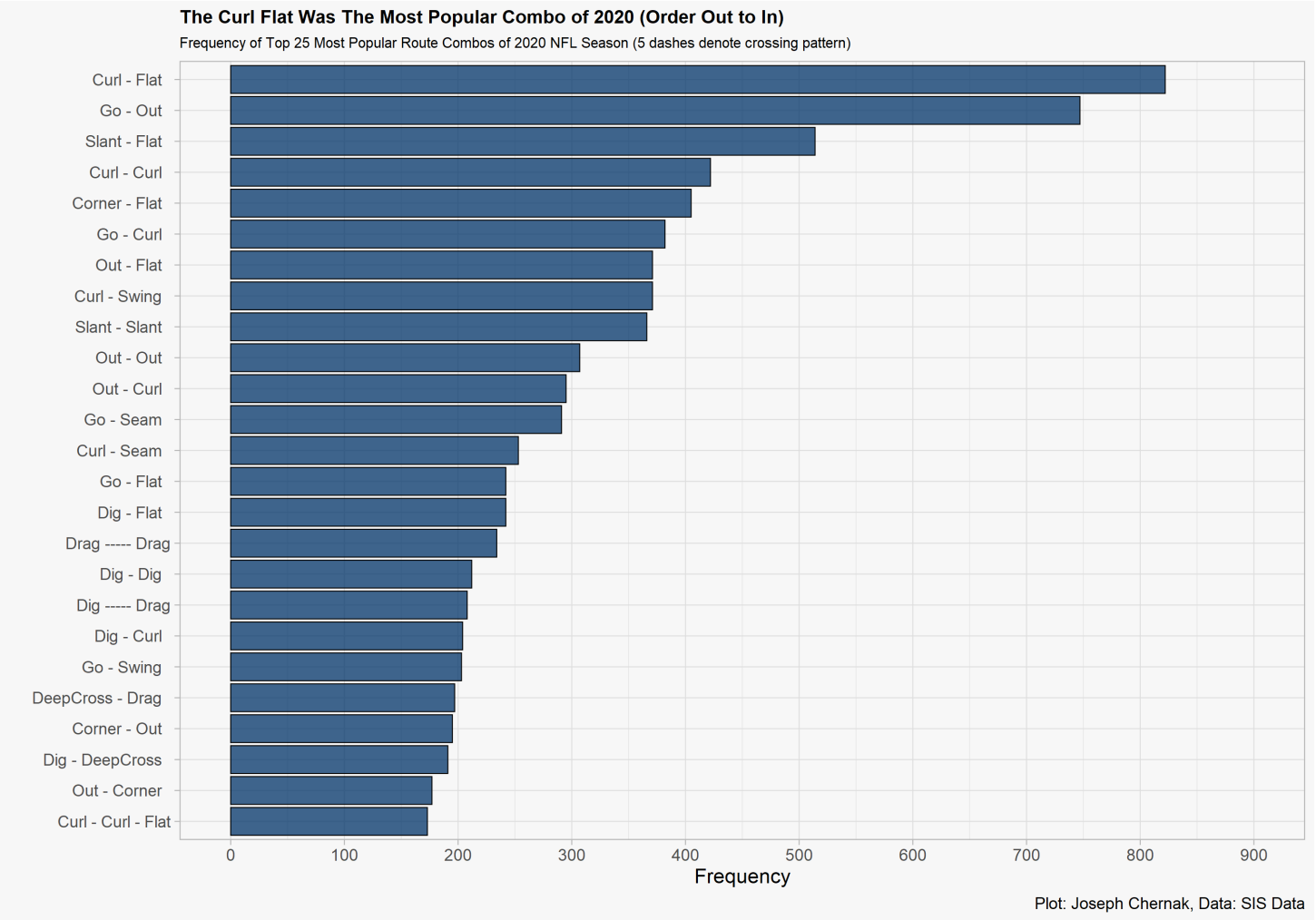
Route Combo Assumptions

1. I began by examining the unique routes in the dataset (50 different routes). To ensure the general route concept was captured in the analysis, I reclassified routes into a less specific classification.

2. Additionally, routes that were classified as “Flat – Left”, “Flat – Right”, “Swing – Left”, or “Swing – Right” and did not involve fast motion were reclassified to just “Flat” or “Swing”.
   - Those routes that did involve a fast motion were also reclassified to “Flat” or “Swing”, but their side of center was changed if their side of center was opposite from the described “left” or “right” action. (E.g., a player who: 1.) ran a “Flat – Left”, 2.) was initially marked as a “R” side of center & 3.) was marked as involved in fast motion, was reclassified as a “Flat” with an “L” side of center and their order from out to in was set as the innermost route runner.

Route Combo Assumptions Cont.

3. In theory, we could just split each play into routes run onto the left side of the center and right side of the center and analyze both sides as separate route combos. But this ignores route combos that involve crossing patterns.
4. To address this, I classified the following routes as “crossing routes” if run by a slot WR or tight end:
   - "Drag", "Dig", "Deep Cross", "Post", "Sit Over Middle”
5. If a play had a crossing route from a slot WR or TE on both sides of the field, then I considered this play a crosser and extracted only the crossing routes from the play.
   - For example: Left SOC WR #1 has a “Go” and is aligned at “WR”, WR #2 has a “Drag” is aligned at “SWR” , WR #3 has a “Dig” and is aligned at “SWR”. Right SOC WR # 1 has a “Drag” is aligned at “SWR”. We would mark this play as a crosser and extract the two drags and the dig while ignoring the “Go”.



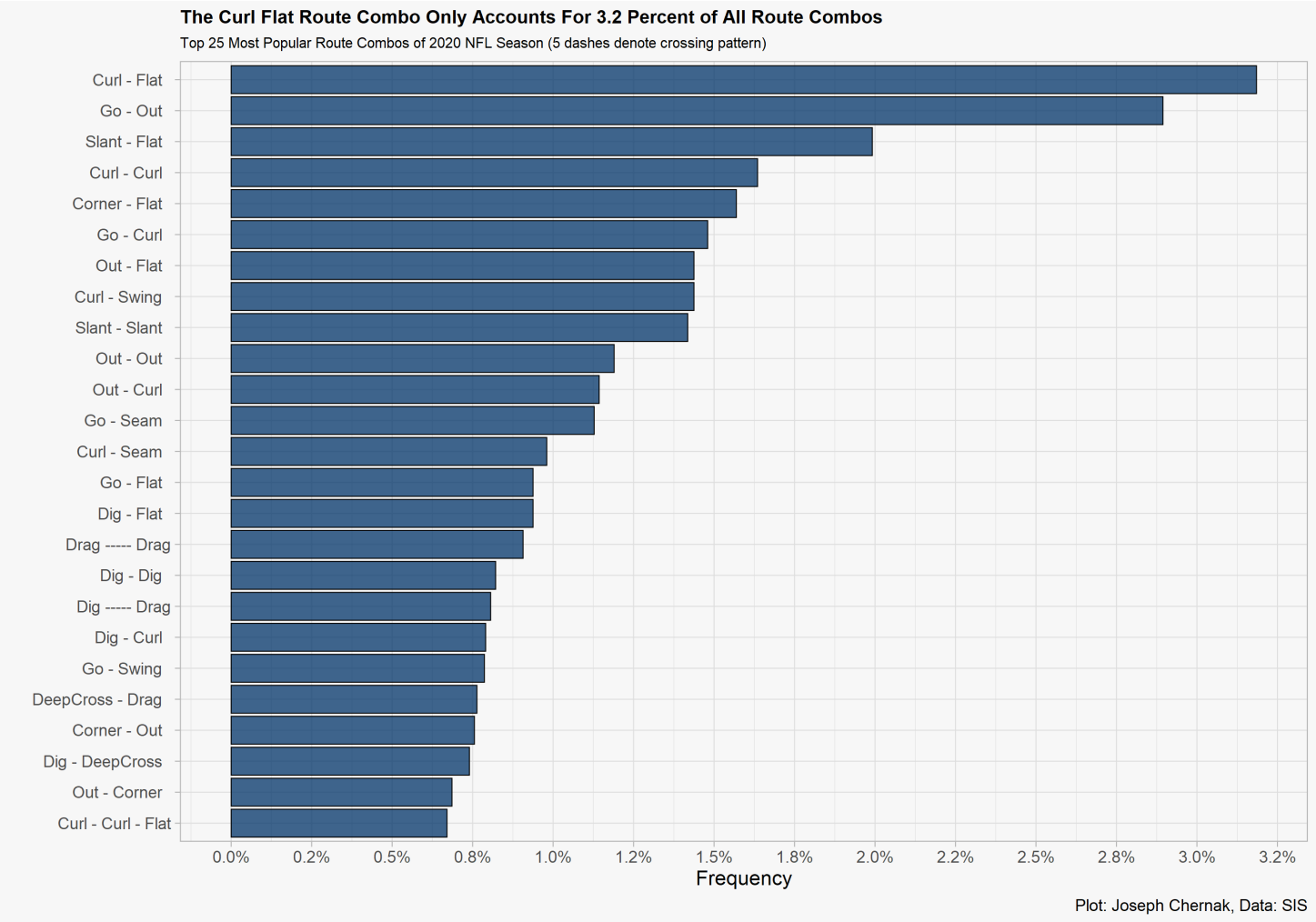
What Route Combo's Were Most Popular?

The most run route combination of 2020 was the "Curl – Flat" followed by the "Go – Out".

Two crossing route combos show up in our data as well (5 dashes indicate the offensive line) :

a. "Drag ----- Drag"

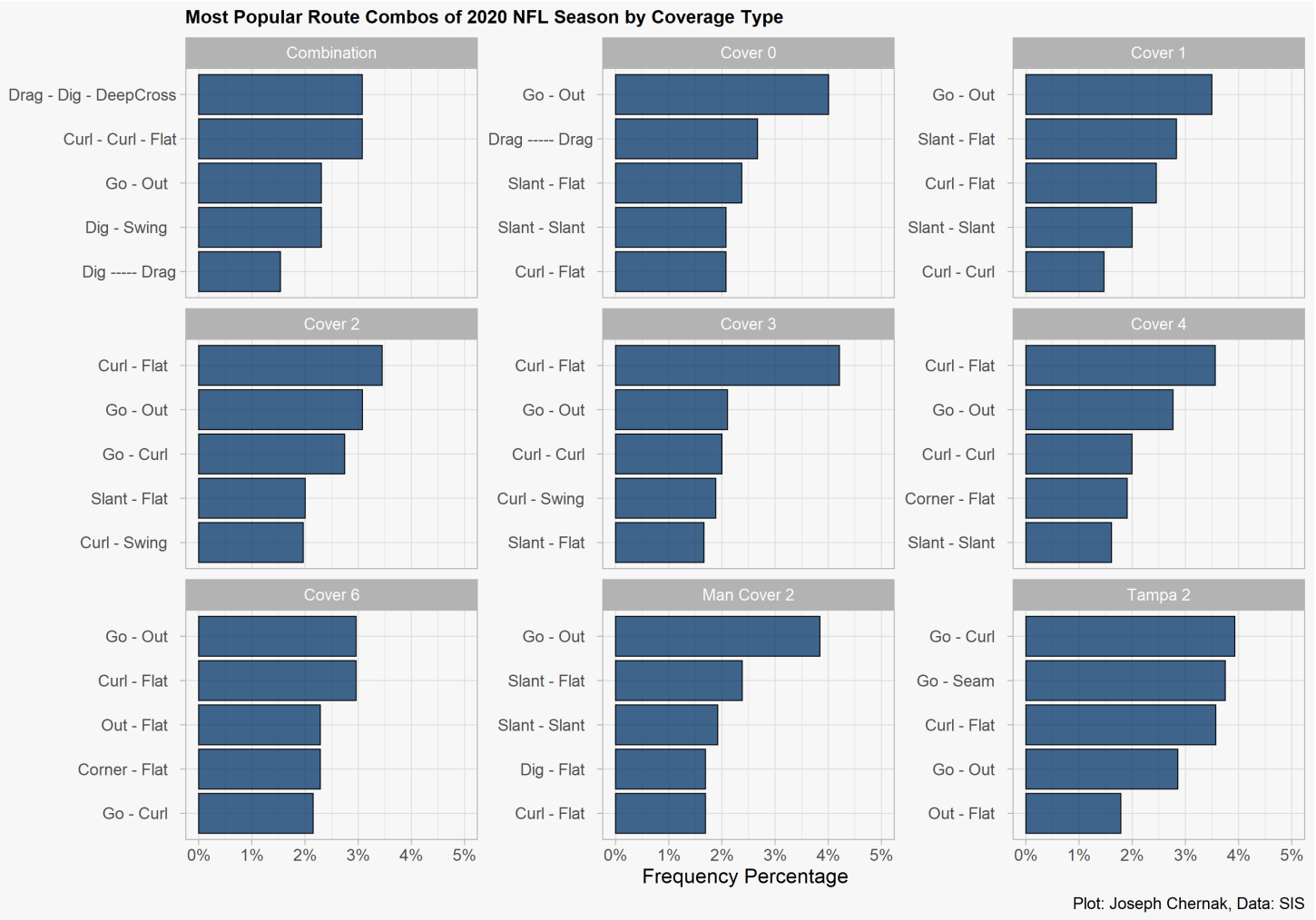
b. "Dig ----- Drag".



What Route Combo's Were Most Popular?

If we observe the frequency percentage of the same top 25 route combos, we can see that the most often run combos were not called at very high percent. This tells us that:

a. There is a lot of variation in route combos.
b. NFL play calling is hard to predict in terms of passing concepts.



What Route Combo's Were Most Popular Against Each Coverage?

If we observe the frequency percentage of the top 5 route combos by coverage scheme, we can observe some diversity in what is being called.

Against cover 3, 4 of the top 5 route combos attempt to put the "Flat/Curl" defender into conflict.

But what combos are most effective?

How Do We Define Route Combo "Success"?

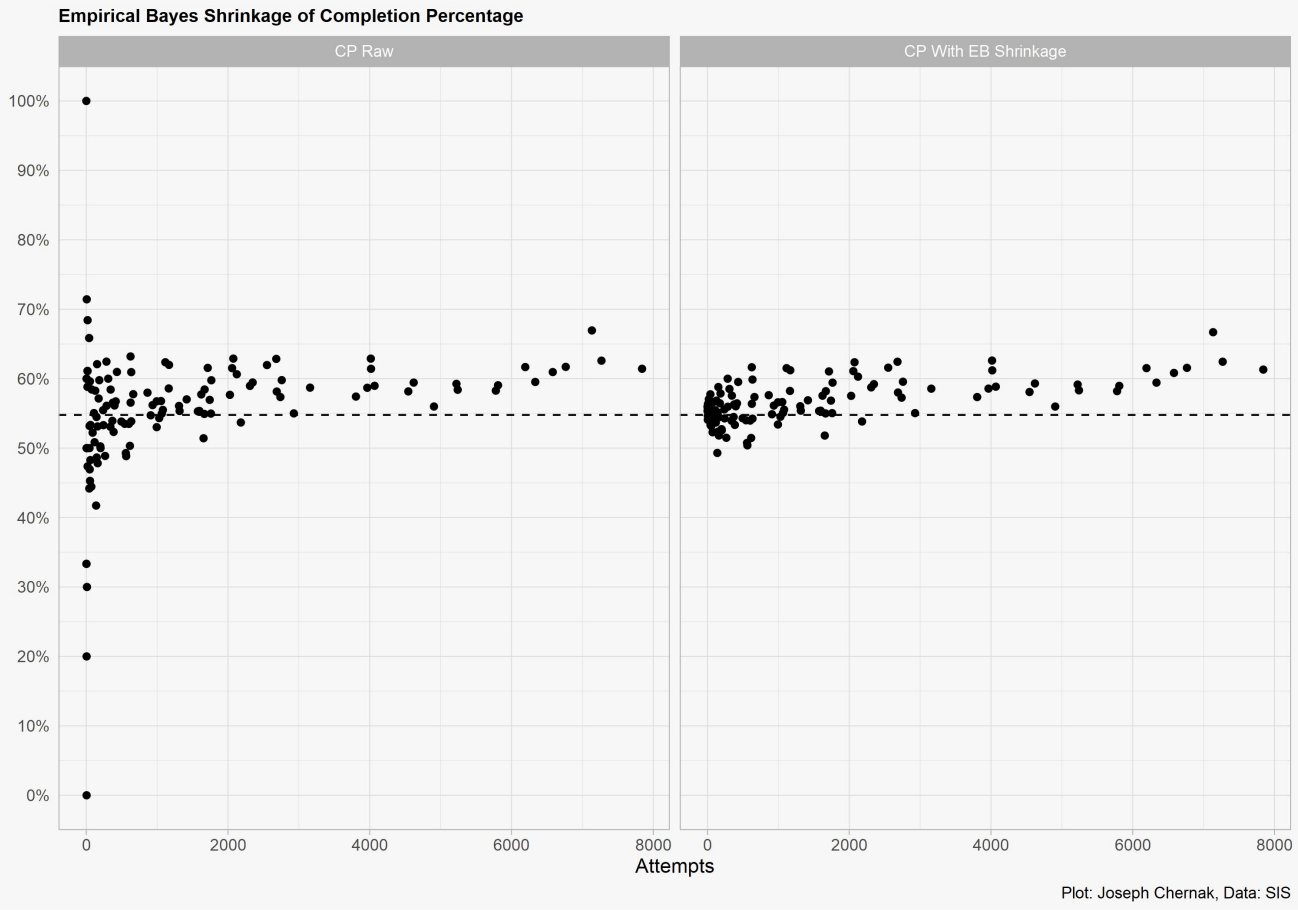
There are many metrics that are used to evaluate the passing game (EPA, CPOE, Success Rate, etc..), so which should we pick?

That's a trick questions, each metric tells us something different and in conjunction can answer our questions.

Thus, I decided to evaluate each route combo on their:

1. Completion Percentage – to evaluate how "easy" it is to complete a pass within the given route combo.
2. Expected Points Added – to evaluate how "effective" a route combo is.

When used together, these two metrics can identify route combos that are completed at a high level but are adding to the success of the offense. For this project, I used Bayesian methods and mixed effects methods.



Estimating Completion Percentage With Mixed Models (And Some Bayes)

To estimate the CP of a given route combo, we could just do:

$$Completion\ Percentage = \frac{Total\ Completions}{Total\ Attempts}$$

But then we'd be missing out on important factors that influence completion percentage such as air yards, who the QB is and other situation factors.

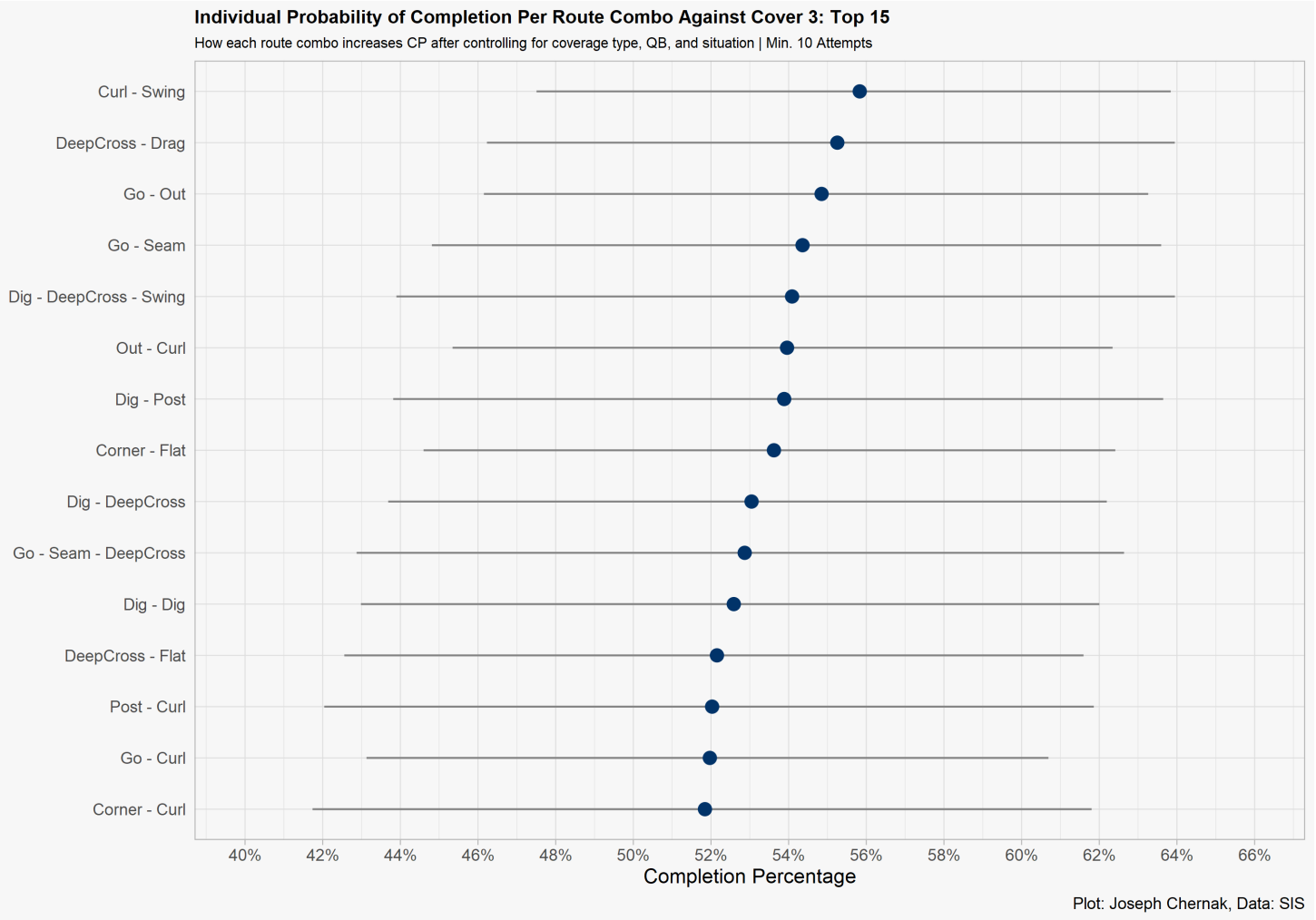
To overcome this, I fit a Generalized Additive Model (GAM) that accounts for situational factors (down, distance, pressure, air yards, and QB completion percentage).

The Bayesian aspect of this is that I estimated a given QB's completion percentage via empirical bayes beta binomial regression (so that QB's with very few attempts were shrunk to the average completion percentage, thus not distorting the model).

Mixed Model to Estimate Probability of Completion

With our EB estimates in hand, I used a mixed effect GAM to estimate the individual probability of completion for each route combo. I fit the following model which uses a smoothing spline on throw depth and accounts for situation factors that effect completion percentage.

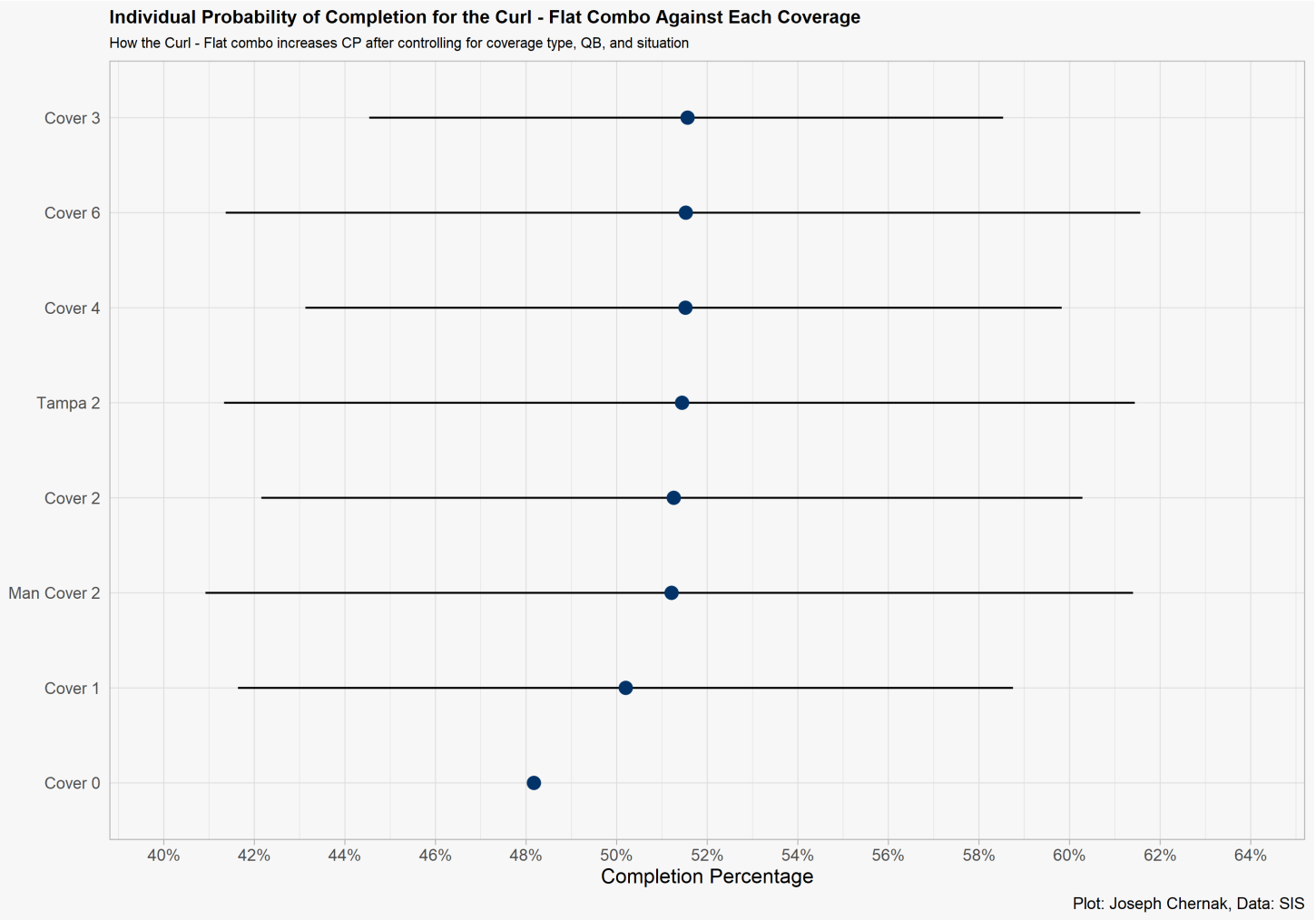
```
gam_model <- gamm4(
  Completion ~ #target
    ToGo + #yards to go
    StartYard + #yardline
    Down +
    PressureOnPlay + #if there was pressure on the play
    ayard_is_zero + #if the pass was at or behind line of scrimmage
    eb_estimate + #estimate of QB true completion percentage
    s(ThrowDepth), #spline on throw depth
  random = ~ (1 | combo_ID:Coverage_ID), #crossed mixed effects
  data = Data,
  nAGQ = 0,
  control = glmerControl(optimizer = "nloptwrap"),
  family = binomial(link = "logit")
)
```



Mixed Model Estimate of Completion Percentage

After modeling the individual probability of completion for each route combo, we can plot their estimate and confidence intervals against specific coverages.

For Cover 3, the “Curl – Swing” has the highest completion percentage probability but there is a high degree of uncertainty in our estimate. This is because of the relatively small sample (1 season).



How the Popular "Curl – Flat" Fairs Against Different Coverages

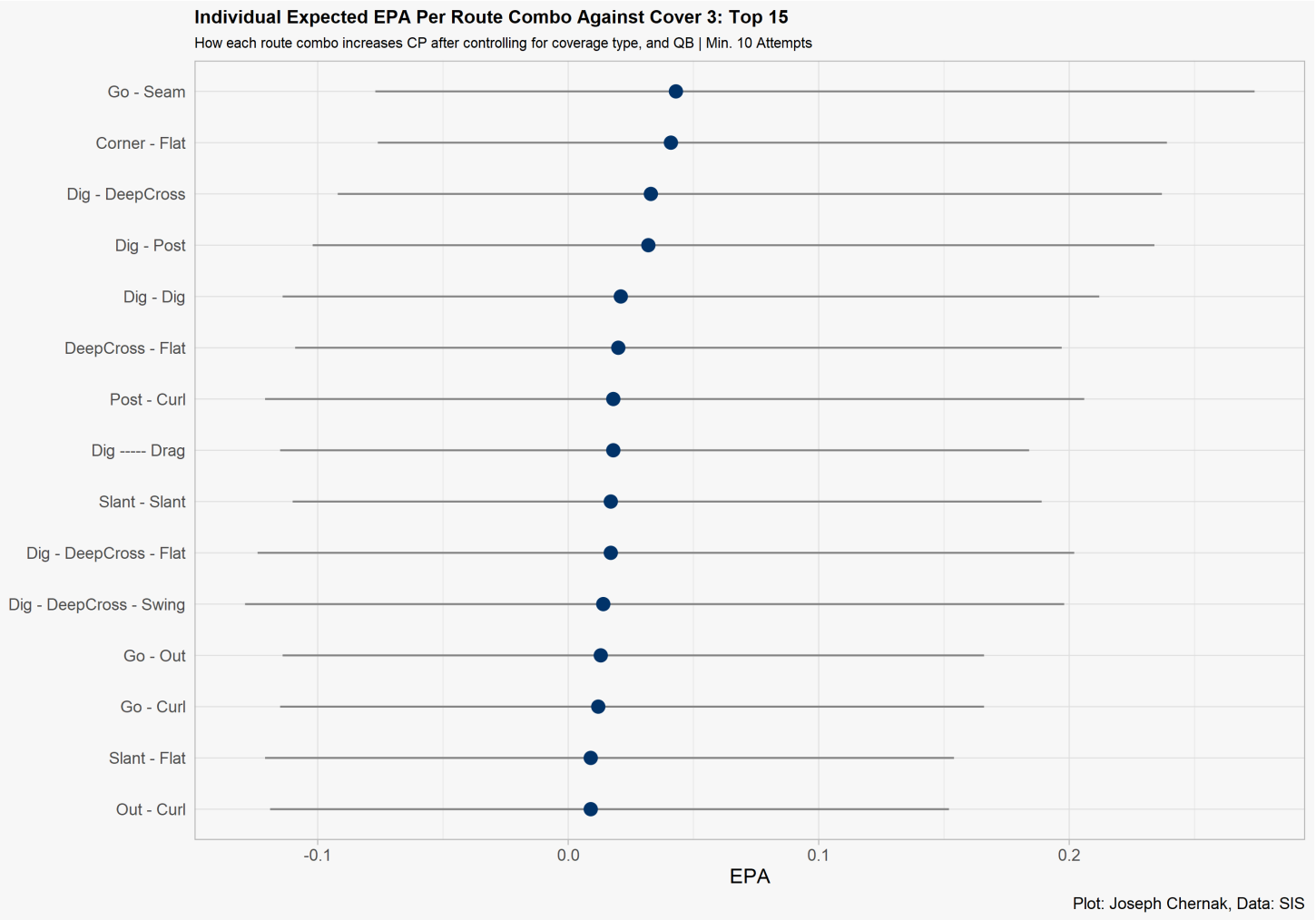
We can examine the completion percentage of the "Curl – Flat" route combo against each coverage in the data set. It is most effective against Cover 3.

Estimating EPA with Mixed Modeling

Using brms in R, we can build a simple multilevel, mixed model that estimates a route combo contribution to EPA after controlling for who the QB is and what coverage was called.

This model uses a Student T distribution to account for the unusual distribution of EPA. I further set the delta to .99 to avoid divergent transitions.

```
EPA_Mixed <- brms::brm(
  EPA_Final ~ Career_EPA + (1 | combo_ID:Coverage_ID),
  data   = Data,
  control = list(adapt_delta = 0.99),
      iter = 3000, warmup = 1000, chains = 3,
      seed = 9)
```

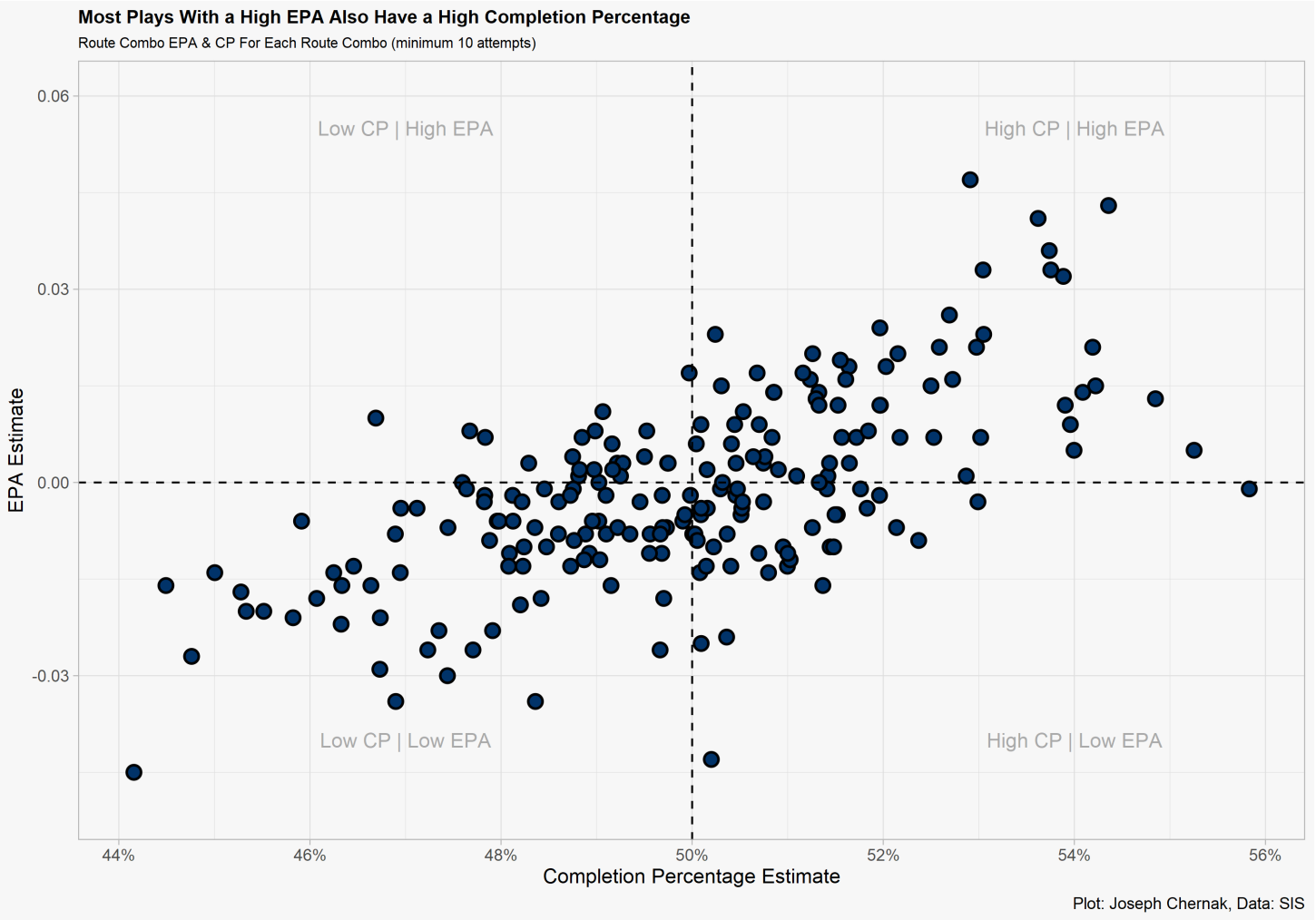


EPA Results

After setting a minimum attempt threshold on each route combo (min. 10 attempts) we can observe the top 15 route combos against cover 3 according to the model.

Most estimates are close to 0 and have a wide range of uncertainty.

Unsurprisingly, the “Go– Seam” has the highest EPA against Cover 3.



EPA Results Continued

We can further plot each route combo – coverage scheme estimate of completion percentage and EPA to see how using two metrics to rank route combos filters out combos with a high CP but a low EPA or vice versa.

Top 5 Route Combos Against Coverage

ROUTE COMBO	ESTIMATES EPA	ESTIMATES COMP %	OVERALL EPA	OVERALL COMP %	OCCURANCES
COMBINATION (MIN 1 ATTEMPT)					
Dig ----- Drag	0.011	50.88%	+	+	2
Whip - Corner - Out	0.006	51.01%	+	+	2
DoubleMove - Go - Out	0.005	50.88%	+	+	1
Curl - Curl - Swing	0.005	50.75%	+	+	1
Drag - Corner - Dig	0.007	50.47%	+	+	1
TAMPA 2 (MIN 1 ATTEMPT)					
Go - Curl	0.024	52.92%	+	+	9
Go - Seam	0.014	51.25%	+	+	7
Dig ----- Post	0.012	51.49%	+	+	4
Go - Post	0.011	51.36%	+	+	2
Slant - Flat	0.009	51.18%	+	+	4
MAN COVER 2 (MIN 5 ATTEMPTS)					
Curl - Flat	0.010	51.21%	+	+	7
Slant - Slant	0.011	50.54%	+	+	15
Curl - Curl	0.003	50.95%	+	+	6
Go - Curl - Curl	0.008	50.07%	+	+	5
Go - Out - Out	0.003	50.72%	+	+	9
COVER 0 (MIN 5 ATTEMPTS)					
Slant - Flat	0.017	50.58%	+	+	8
Slant - Out	0.003	52.22%	+	+	7
Corner - Flat	0.011	50.98%	+	+	7
Pick - Flat	0.003	51.61%	+	+	6
Dig ----- Drag	0.005	50.10%	+	+	7

Table: Joseph Chernak, Data = SIS

Top 5 Route Combos Against Each Coverage

ROUTE COMBO	ESTIMATES EPA	ESTIMATES COMP %	OVERALL EPA	OVERALL COMP %	OCCURANCES
COVER 6 (MIN 5 ATTEMPTS)					
Go - Curl	0.026	51.75%	+	+	7
Corner - Flat	0.018	52.10%	+	+	9
Go - Out - Flat	0.015	51.56%	+	+	5
Curl - Flat	0.012	51.53%	+	+	12
Curl - Out	0.007	50.08%	+	+	6
COVER 4 (MIN 10 ATTEMPTS)					
Go - Seam - Seam	0.033	53.75%	+	+	10
Go - Out	0.021	54.19%	+	+	48
Dig - Curl	0.026	52.69%	+	+	10
Curl - Go - Curl	0.012	53.90%	+	+	11
Corner - Flat	0.016	52.73%	+	+	44
COVER 3 (MIN 10 ATTEMPTS)					
Go - Seam	0.043	54.36%	+	+	22
Corner - Flat	0.041	53.62%	+	+	47
Dig - Post	0.032	53.88%	+	+	11
Dig - DeepCross	0.033	53.04%	+	+	27
Go - Out	0.013	54.85%	+	+	57
COVER 2 (MIN 10 ATTEMPTS)					
Go - Out	0.047	52.91%	+	+	43
Dig ----- Drag	0.015	54.22%	+	+	23
Go - Curl	0.005	54.00%	+	+	35
Out - Flat	0.007	53.02%	+	+	19
Curl - Flat	0.020	51.26%	+	+	41
COVER 1 (MIN 10 ATTEMPTS)					
Dig ----- Dig	0.036	53.74%	+	+	16
Out - Out	0.023	53.05%	+	+	36
Drag - Dig	0.021	52.98%	+	+	18
Out - Corner	0.024	51.97%	+	+	27
Go - DeepCross	0.019	51.55%	+	+	16

Table: Joseph Chernak, Data = SIS

Our Best Route Combos

We can observe the best route combos against each coverage by looking at the mixed model estimates of completion percentage and EPA in table form.

Against Cover 3, the “Go – Seam” is the best route combo according to its estimated EPA and Completion Percentage. For certain coverages, the attempts threshold is lowered to 5 or 1 because of a lack of qualifying route combos.

Limitations

Although the mixed models estimated the most effective and least effective route combos against each coverage, the project has notable limitations:

1. Route combos that are open but not targeted are not estimated in these models.
2. Without tracking data, we must judge each throw on its outcome, not process (Was it a bad throw by the quarterback? How much separation did the route combo create?).
3. One season of data is not enough to reach conclusions about each route combo. In 2020, the leaguewide completion percentage was 61%, the highest over the last decade, and leaguewide EPA/play was .10, also the highest over the last decade. 2020 may be overinflating our estimates.

Thank you for reading and putting on this competition!