Examining the NFL Combine

Kevin Petruzzelli

3/20/2018

Does the NFL Combine matter?

- In a league often dominated by which team successfully drafts incoming collegiate athletes, should NFL team scouts focus on combine performance?
- Does a strong combine performance warrant an athlete to move up in the NFL Draft order?
- Which combine event should a scout focus on by position?

Defining NFL Success

- Focus on three offensive positions ability to move the ball downfield and/or score touchdowns.
  - Quarterback
    - Average Throwing Yards per Season
    - Average Touchdowns per Season
  - Running Back
    - Average Rushing Yards per Season
    - Average Touchdowns per Season
  - Wide Receiver
    - Average Reception Yards per Season
    - Average Touchdowns per Season

NFL Combine Tests

- **40-yard dash**
- **225 lb. Bench press**
- **Vertical jump**
- **Broad jump**
- **3 cone drill**
  The 3 cone drill tests an athlete's ability to change directions at a high speed. Three cones in an L-shape.
- **Height (Inches)**
- **Weight (lbs.)**

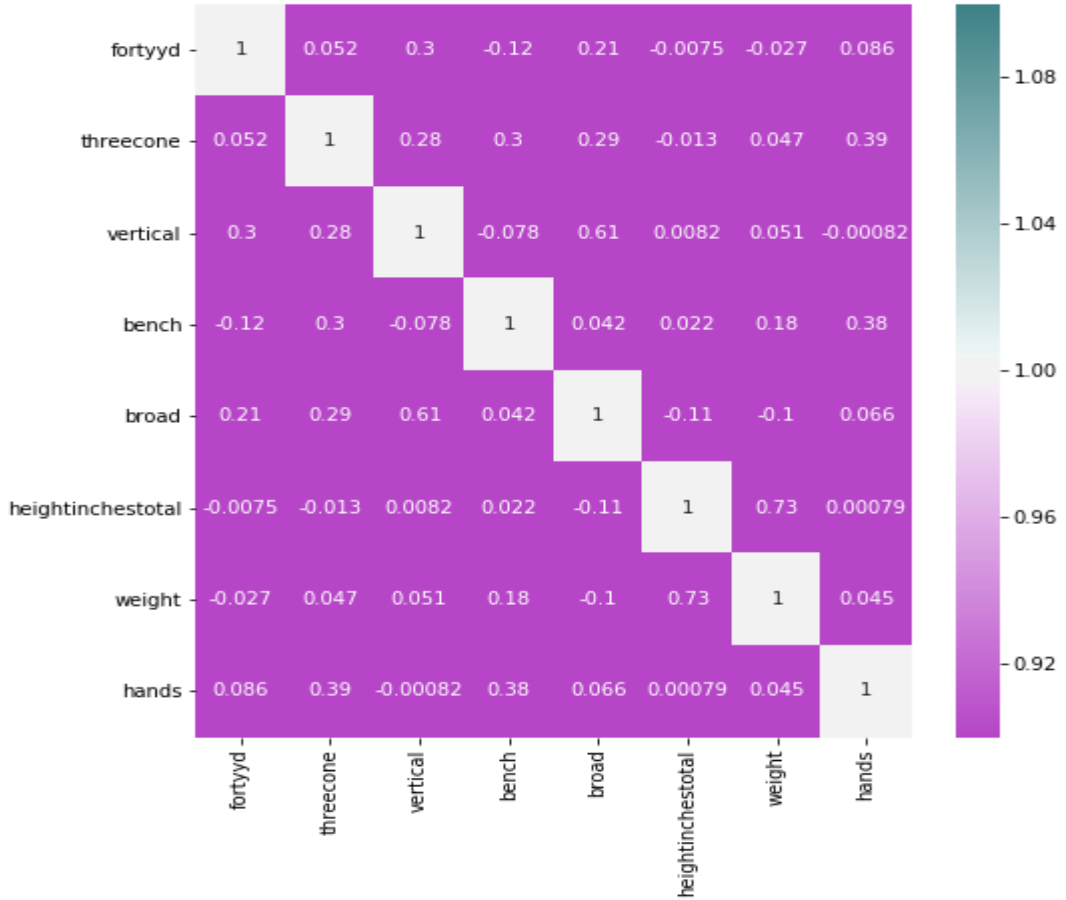
Exploring The Data

- Data is split between WR, QB, and RB positions.
- The combine data is joined to career player stats, such as reception yards.
- Data is confined to the years 1999 to 2015.

	year	name	firstname	lastname	position	heightfeet	heightinches	heightinchestotal	weight
0	2015	Agholor, Nelson	Nelson	Agholor	WR	6	0	72.0	198
1	2015	Alford, Mario	Mario	Alford	WR	5	8	68.0	180
2	2015	Coates, Sammie	Sammie	Coates	WR	6	1	73.0	212
3	2015	Conley, Chris	Chris	Conley	WR	6	2	74.0	213
4	2015	Cooper, Amari	Amari	Cooper	WR	6	1	73.0	211
5	2015	Crowder, Jamison	Jamison	Crowder	WR	5	8	68.0	185

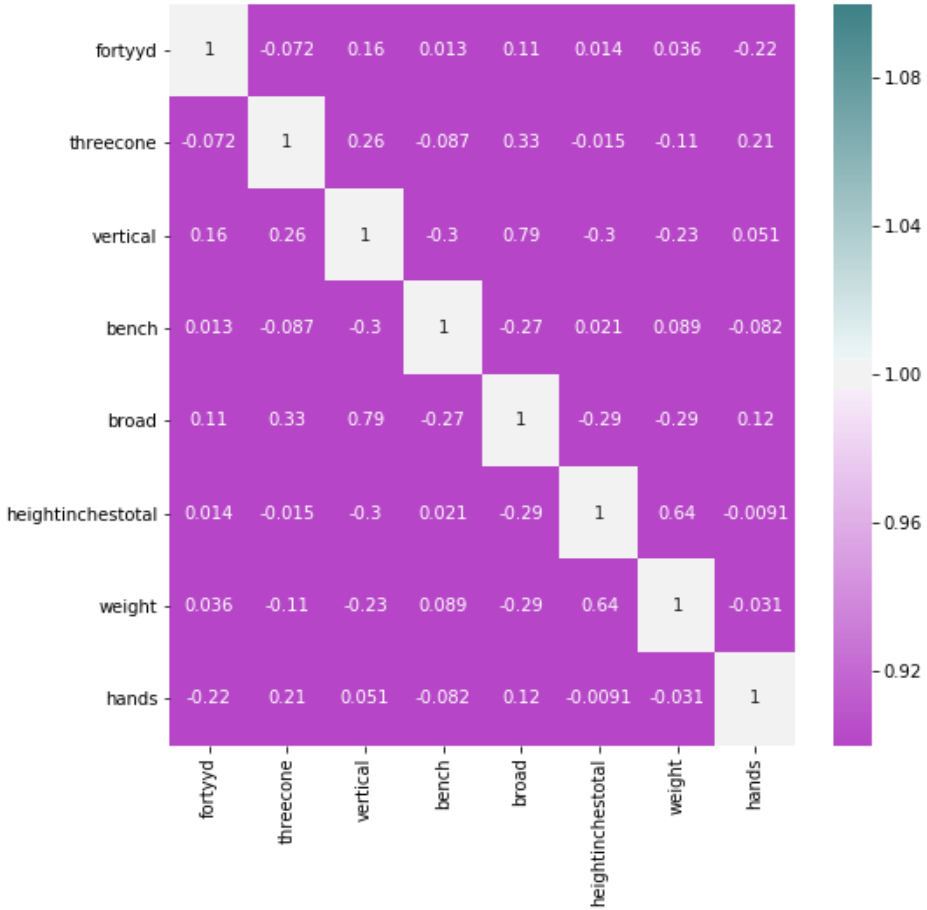
Exploring The Data

- Wide Receiver combine variables examined for multicollinearity
- Highest correlated tests are broad jump and vertical

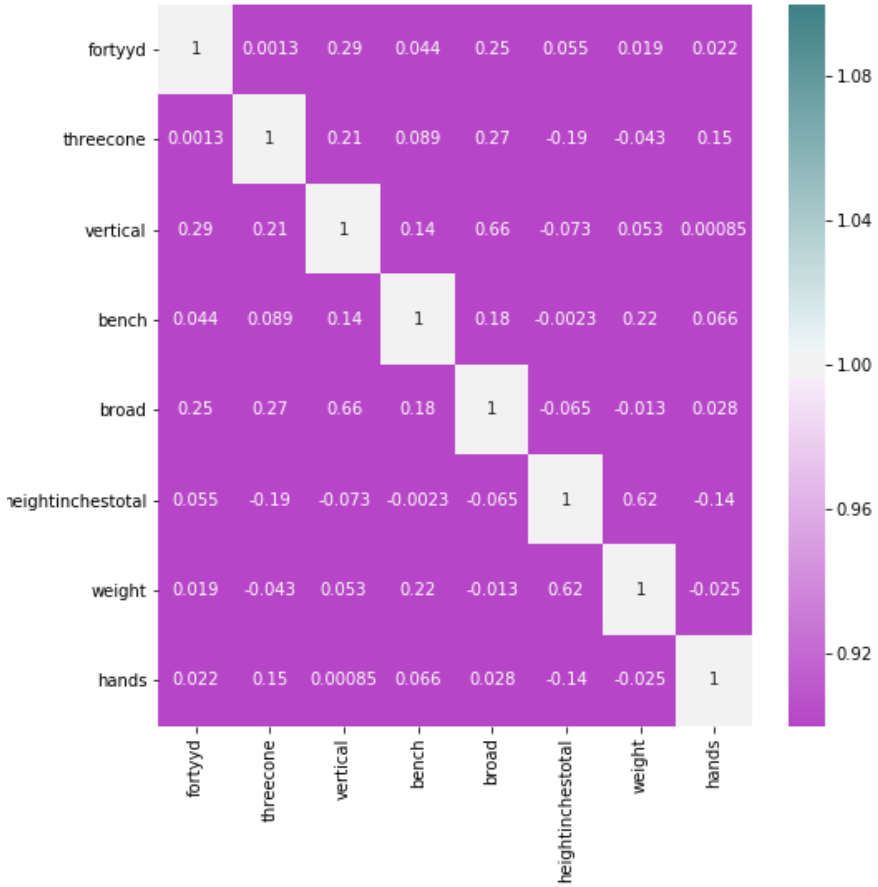


Exploring The Data

QB



RB



Model Selection

- In order to predict season averages in statistical categories such as receiving yards, a multilinear regression model was developed
- Utilizing the sklearn library, six models were developed to predict statistical averages based upon an athletes performance in the aforementioned combine events.
- Each model was tuned using regularization techniques
    - Lasso Regularization
    - Ridge Regularization

Examining coefficients – RB Rush Yards

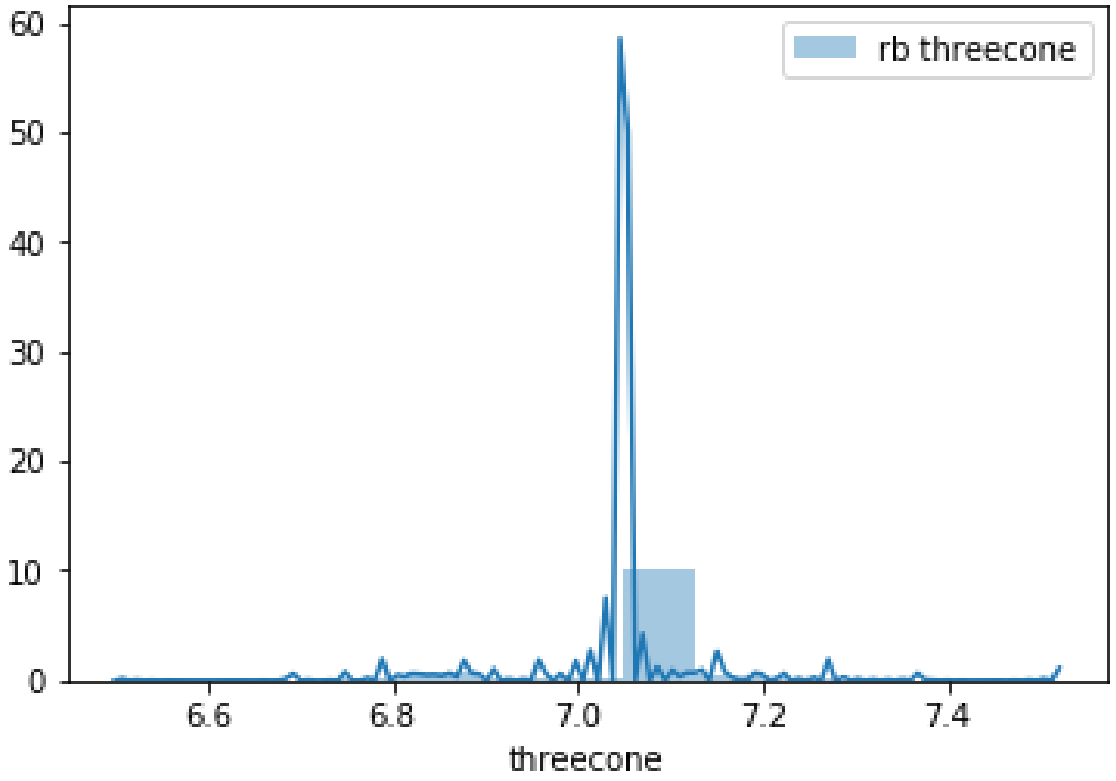
OLS Regression Results

Dep. Variable:	SeasonAvgRushYds	R-squared:	0.651
Model:	OLS	Adj. R-squared:	0.637
Method:	Least Squares	F-statistic:	46.89
Date:	Tue, 20 Mar 2018	Prob (F-statistic):	4.73e-37
Time:	10:36:10	Log-Likelihood:	-1237.7
No. Observations:	183	AIC:	2489.
Df Residuals:	176	BIC:	2512.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
fortyyd	-436.3435	140.901	-3.097	0.002	-714.417	-158.270
threecone	98.8864	113.498	0.871	0.385	-125.105	322.878
vertical	-6.8642	6.968	-0.985	0.326	-20.615	6.886
bench	-2.1236	4.307	-0.493	0.623	-10.623	6.376
broad	5.9609	3.693	1.614	0.108	-1.328	13.249
heightinchestotal	3.2453	10.408	0.312	0.756	-17.295	23.786
weight	4.2313	1.646	2.571	0.011	0.983	7.480

- Significant Variables
  - Three cone
  - Vertical
  - Height
  - Weight
- Three cone drill has the highest coefficient
- Model when created utilizing sklearn indicates that the variables are not provide an adequate indication of rushing yards.
  - R Squared Values less than .15

Examining coefficients - RB



A data gap in the combine scores resulted in a large concentration around the mean. A significant number of athletes do not perform the three cone drill.

Examining coefficients – QB Passing Yards

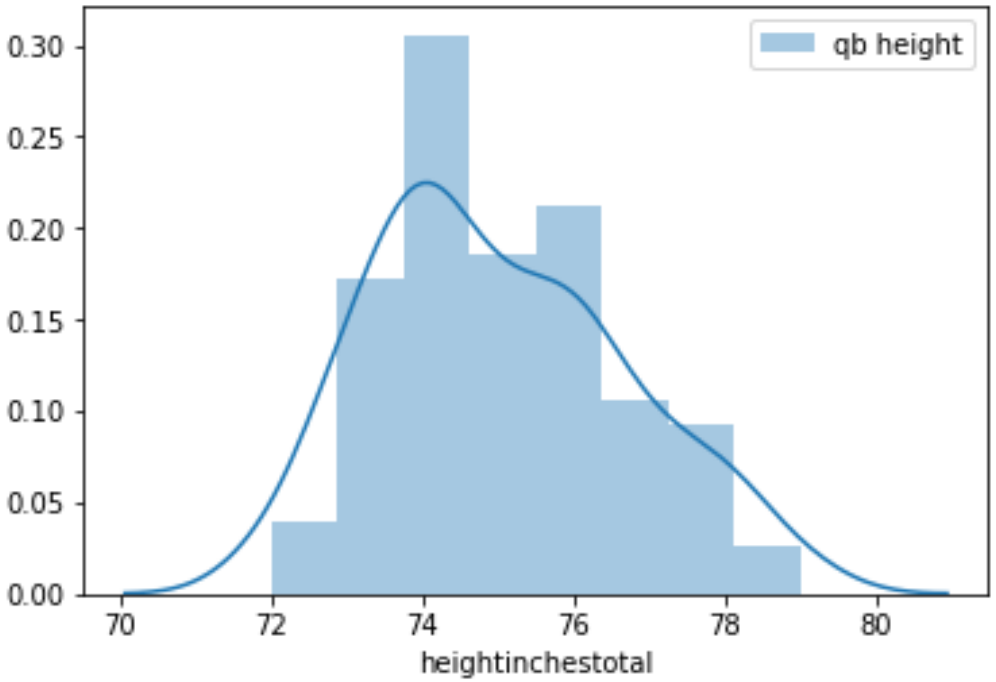
OLS Regression Results

Dep. Variable:	SeasonAvgPassYds	R-squared:	0.652
Model:	OLS	Adj. R-squared:	0.621
Method:	Least Squares	F-statistic:	21.11
Date:	Tue, 20 Mar 2018	Prob (F-statistic):	9.51e-16
Time:	10:44:52	Log-Likelihood:	-581.43
No. Observations:	86	AIC:	1177.
Df Residuals:	79	BIC:	1194.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
fortyyd	-132.7430	164.258	-0.808	0.421	-459.690	194.203
threecone	241.0981	171.654	1.405	0.164	-100.572	582.768
vertical	-19.3659	10.935	-1.771	0.080	-41.131	2.400
bench	1.7788	25.444	0.070	0.944	-48.866	52.423
broad	3.6217	4.933	0.734	0.465	-6.197	13.440
heightinchestotal	4.4196	17.409	0.254	0.800	-30.233	39.072
weight	-4.2050	2.863	-1.469	0.146	-9.904	1.494

- Most significant variable
  - Height has a positive correlation to passing yards and contains the strongest P value indicating significance.
- Model when created utilizing sklearn indicates that the variables do not provide an adequate indication of passing yards.
  - R Squared Values less than .10

Examining coefficients - QB



- There is an indication that the heights of QBs is highly concentrated around 74 to 76 inches
- Do two inches of height truly indicate an increase in performance indicators such as total passing yards? No!

Examining coefficients – WR Receiving Yards

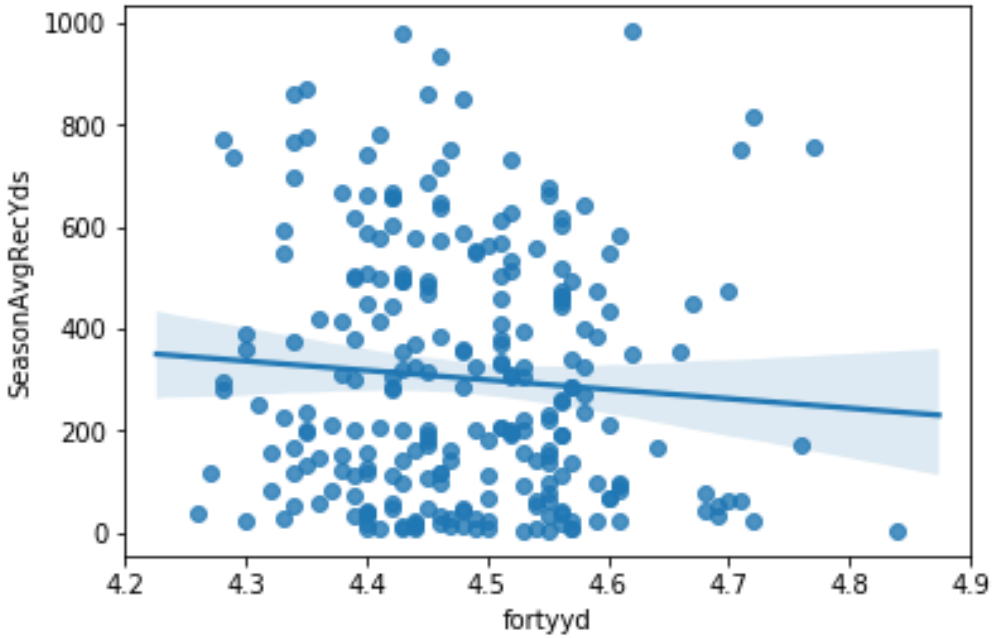
OLS Regression Results

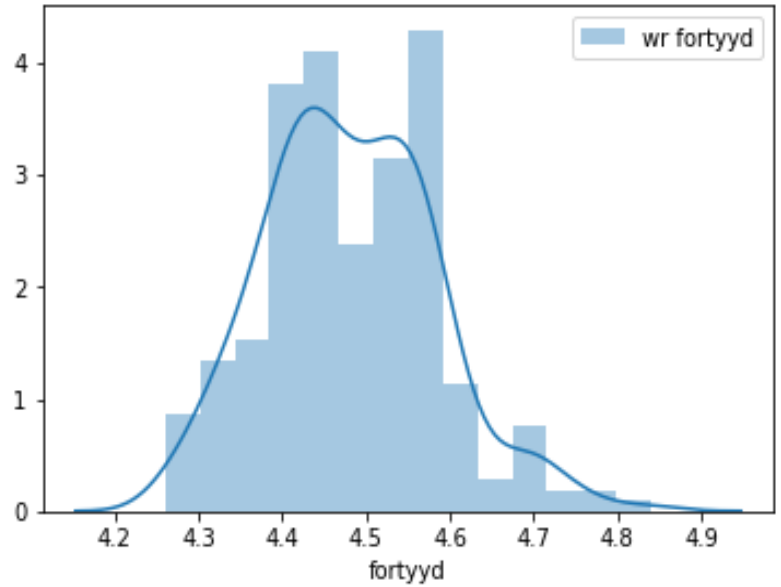
Dep. Variable:	SeasonAvgRecYds	R-squared:	0.617
Model:	OLS	Adj. R-squared:	0.606
Method:	Least Squares	F-statistic:	56.53
Date:	Tue, 20 Mar 2018	Prob (F-statistic):	9.64e-48
Time:	10:36:00	Log-Likelihood:	-1747.5
No. Observations:	253	AIC:	3509.
Df Residuals:	246	BIC:	3534.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
fortyyd	-225.0070	138.519	-1.624	0.106	-497.841	47.827
threecone	189.2936	94.520	2.003	0.046	3.122	375.465
vertical	-1.2952	6.518	-0.199	0.843	-14.132	11.542
bench	-1.4090	5.346	-0.264	0.792	-11.938	9.120
broad	1.7483	3.415	0.512	0.609	-4.978	8.475
heightinchestotal	-7.9704	9.513	-0.838	0.403	-26.707	10.767
weight	2.1846	1.328	1.645	0.101	-0.431	4.800

- Significant Variables
  - Forty Yd
  - Vertical
  - Broad
  - Height
  - Weight
- 40 yard drill has the highest coefficient
- Model when created utilizing sklearn indicates that the variables are not provide an adequate indication of receiving yards.
  - R Squared Values less than .15

Examining coefficients - WR





Conclusions

- Combine performance is not an indicator of NFL performance.
- Why?
  - One consideration is that once an athlete is considered an NFL prospect, the combine results are tightly clustered. They do not provide significant separation or correlation to performance.
  - Simply graphing a combine result relative to performance ,such as average receiving yards ,indicates that there is no line of best fit.
  - Athletes are drafted and never play a single snap. This results in 0 on field performance indicators.

Conclusions

- Next Steps:
  - Add an additional categorical variable “starter”
  - Limit the data to only athletes that are in a starting role as QB, RB, or WR
- Create a new model
- An interesting correlation to view is the relationship between a combine event and number of injuries.
- If the combine results have no true correlation to on field performance, why are we risking athlete's health?

Billy Price suffers incomplete tear of left pectoral at NFL scouting combine

Cameron Wolfe
ESPN Staff Writer

Mar 2, 2018

INDIANAPOLIS -- Ohio State center Billy Price suffered an incomplete tear of his left pectoral muscle during his bench-press workout at the NFL combine, he said Friday.

Price, a two-time All-American who is projected as a possible first-round pick, said he felt "a little pop" on after his third rep and immediately stopped. Indianapolis Colts doctors and trainers then looked at him, and he received his MRI results Friday morning.

Conclusions

- Why does the NFL Combine still exist?

The NFL puts a greater emphasis on media generated dollars than athlete health.

INDIANAPOLIS -- A hamstring injury ended Vita Vea's 2018 NFL Scouting Combine performance a bit early, but not before the Washington Huskies' massive defensive tackle made a strong impression on NFL coaches and scouts.

Vea's early exit was reported by NFL Network's Kimberly Jones.

Vea ranked second among defensive linemen in the bench press event Saturday with 41 reps at 225 pounds. He then turned in a 40-yard dash of 5.10 seconds, outstanding for a player of his size (6-foot-4, 347 pounds). Entering the combine, he was rated the No. 9 player in the draft by NFL.com analyst Daniel Jeremiah.

Data Sources

- Yearly Position Stats
  - Kaggle: https://www.kaggle.com/kendallgillies/nflstatistics
- Combine Results
  - Nflcombineresults.com