

A screenshot from the video game Madden NFL 06 showing a Dallas Cowboys player (number 28) running with the football while being pursued by Chicago Bears players. The scene is set in a large stadium filled with spectators.

DS501 Project 4

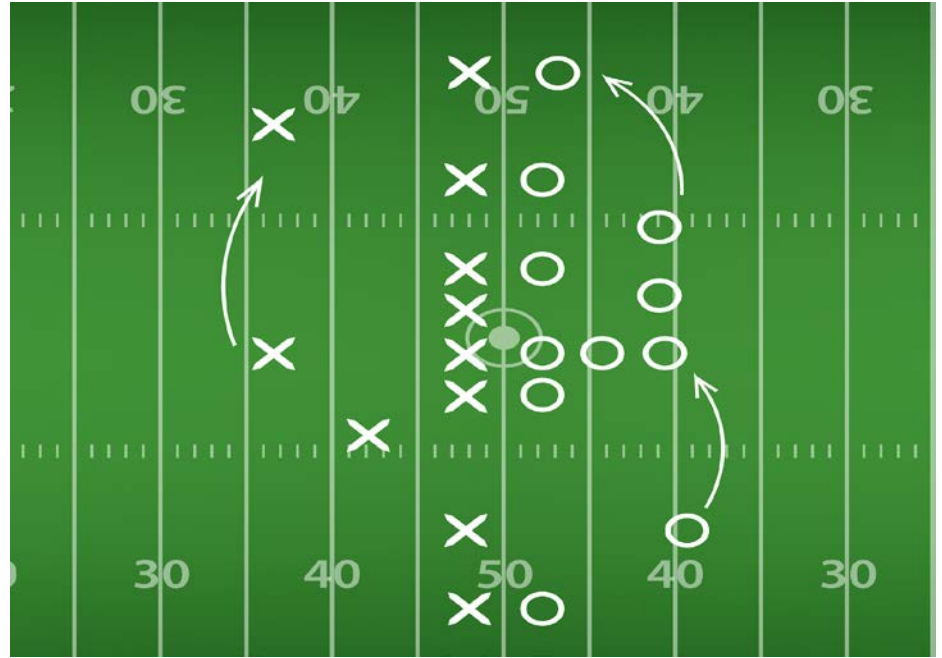
NFL Big Data Bowl

MIA BARGER • QUINCY HERSHEY • ALEXANDER MOORE • ETHAN PRIHAR



Kaggle Competition to Predict Yardage

Training data of 23,000 rushing plays across 49 features

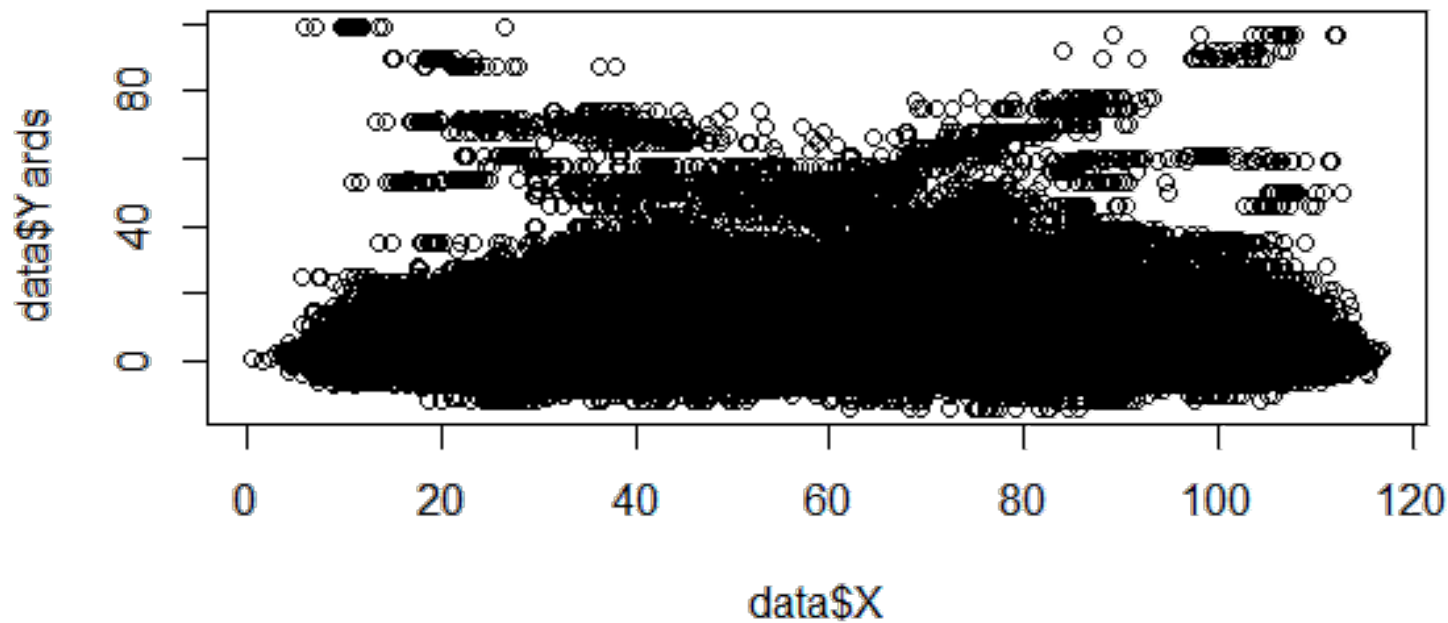


Data Problems

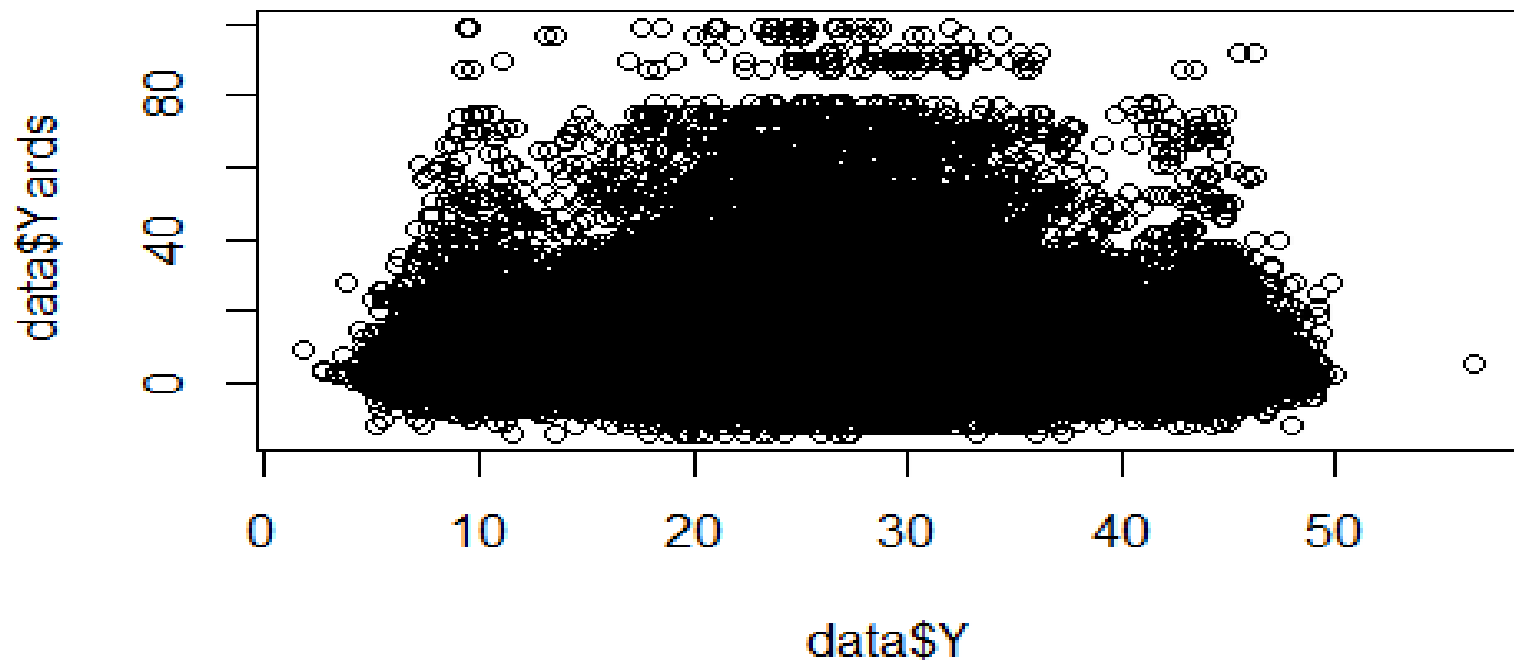
Single features independent to target, complex plays

GameId	PlayId	Team	X	Y	S	A	Dis	Orientation	Dir	NflId	DisplayNai	JerseyNum	Season	YardLine	Quarter	GameClock	Possessor
2.02E+09	2.02E+13	away	73.91	34.84	1.69	1.13	0.4	81.99	177.18	496723	Eric Berry	29	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	74.67	32.64	0.42	1.35	0.01	27.61	198.7	2495116	Allen Baile	97	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	74	33.2	1.22	0.59	0.31	3.01	202.73	2495493	Justin Hou	50	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	71.46	27.7	0.42	0.54	0.02	359.77	105.64	2506353	Derrick Jol	56	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	69.32	35.42	1.82	2.43	0.16	12.63	164.31	2530794	Ron Parke	38	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	75.06	24	1.01	0.32	0.18	308.34	95.01	2543494	Dee Ford	55	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	74.11	16.64	1.11	0.83	0.02	357.23	322.59	2543637	Terrance M	39	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	73.37	18.73	1.24	0.74	0.13	328.52	270.04	2543851	Phillip Gai	23	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	56.63	26.9	0.26	1.86	0.28	344.7	55.31	2550257	Daniel Sor	49	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	73.35	38.83	4.55	0.76	0.51	75.47	190.84	2552488	Marcus Pe	22	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	away	74.15	28.9	0.72	0.73	0.01	342.58	274.14	2556369	Chris Jone	95	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	home	75.82	17.56	2.3	1.39	0.55	178.97	284.15	2649	Danny Am	80	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	home	74.78	33.21	1.71	0.82	0.19	178.82	215.9	497240	Rob Gronk	87	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	home	75.43	32.41	1.5	1.36	0.32	207.08	222.76	2495131	Marcus Ca	61	2017	35	1	14:14:00	NE
2.02E+09	2.02E+13	home	75.9	25.12	1.38	0.8	0.19	133.01	198.55	2495232	Nate Solde	77	2017	35	1	14:14:00	NE

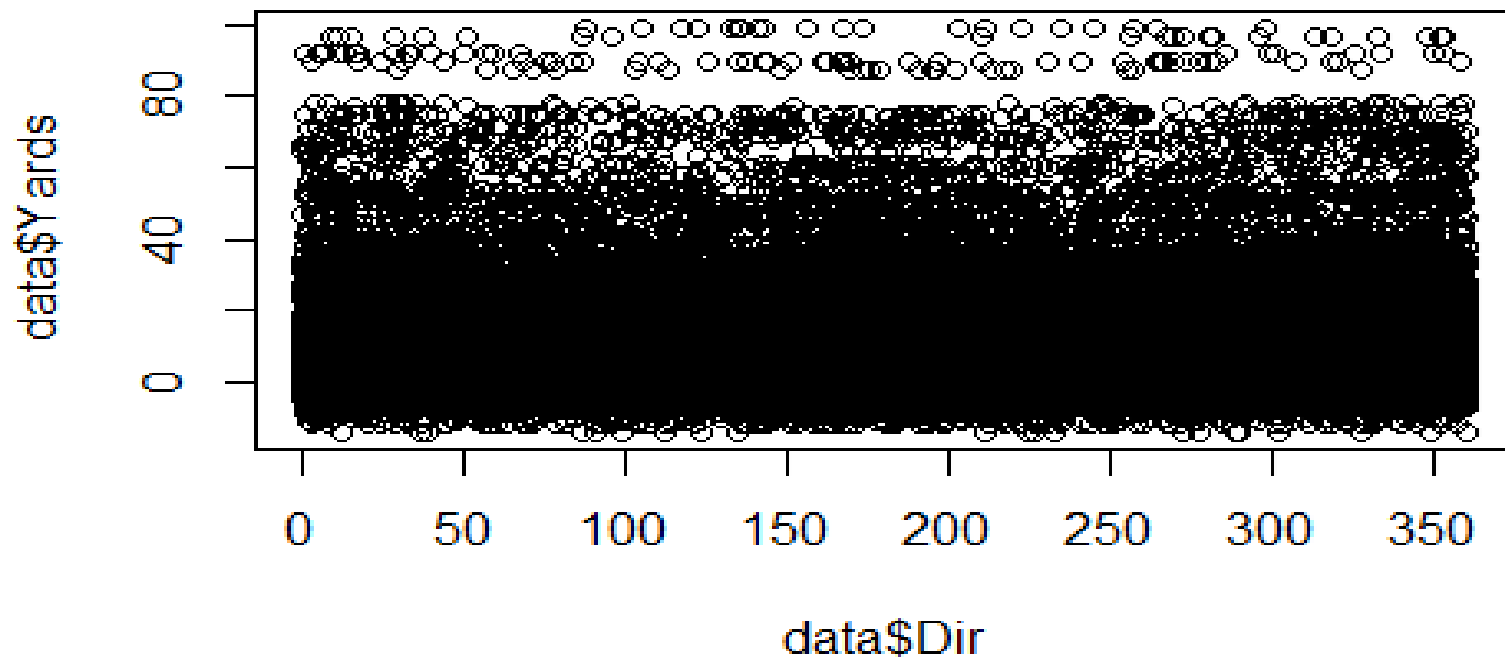
Data Analysis



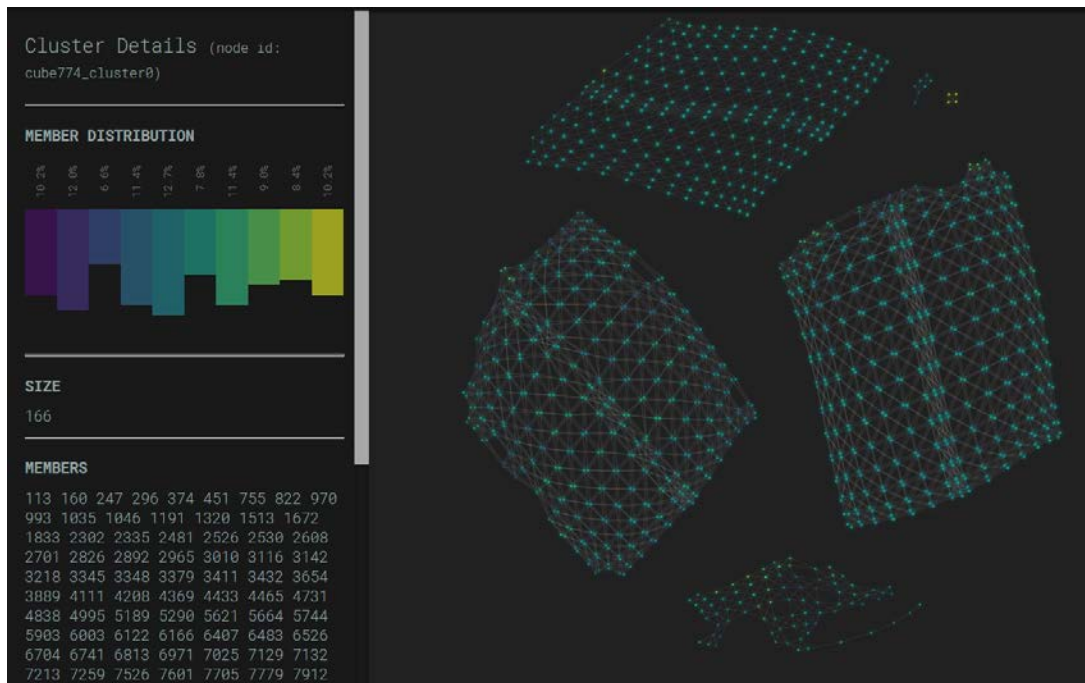
Data Analysis



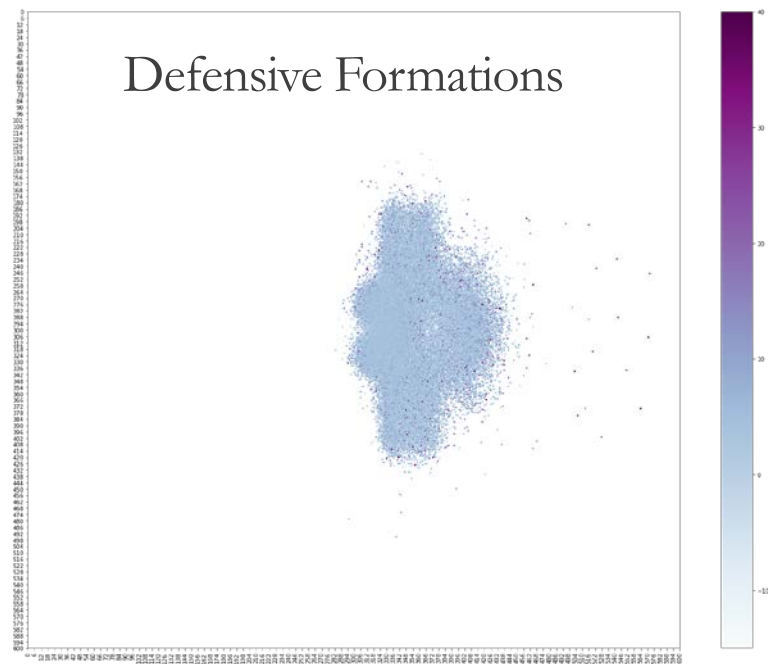
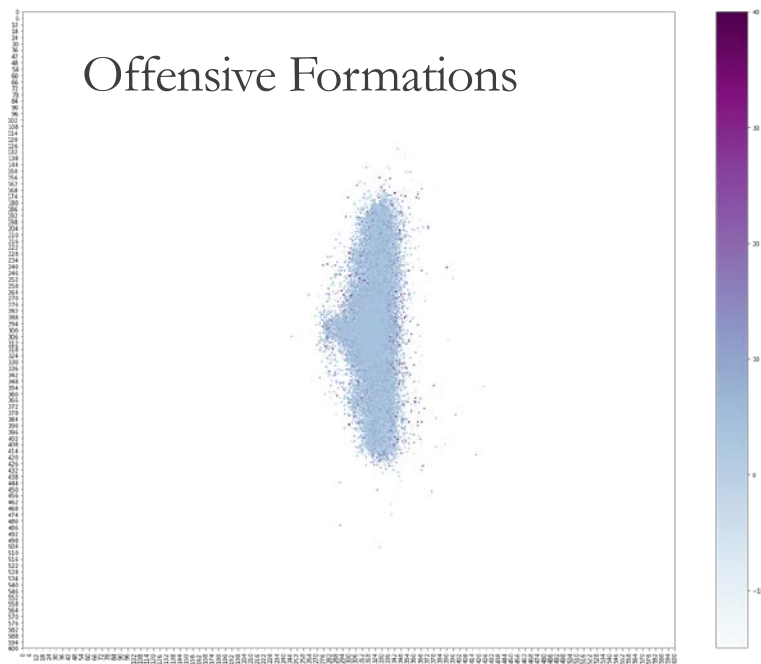
Data Analysis



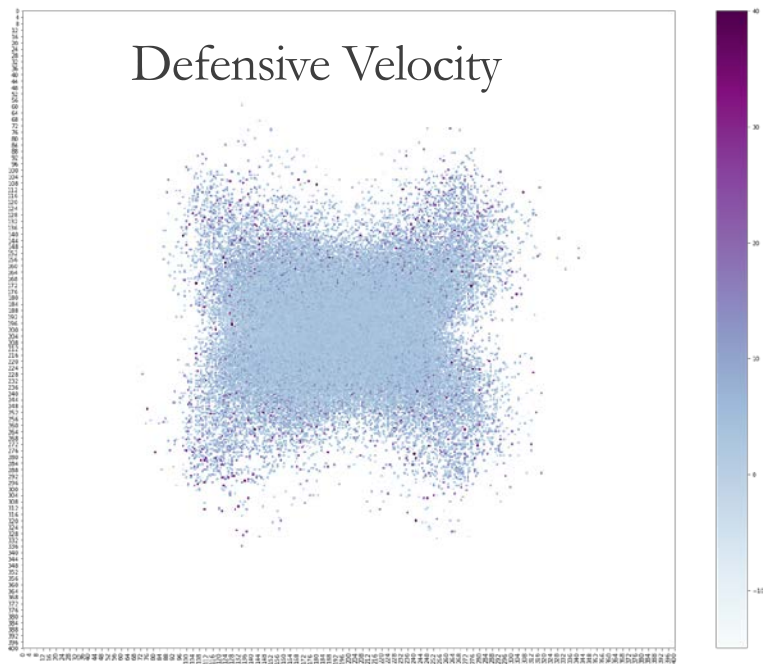
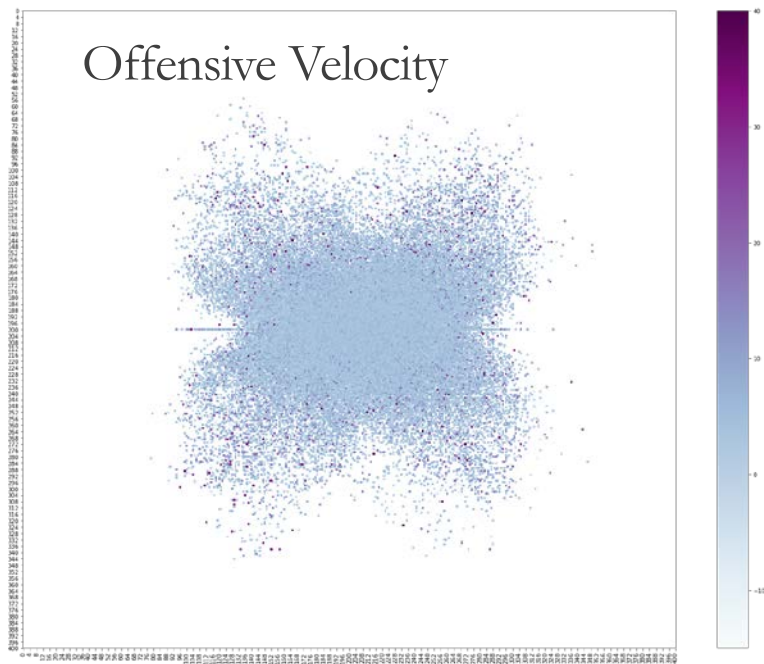
Principal Component Analysis



Data Characteristics

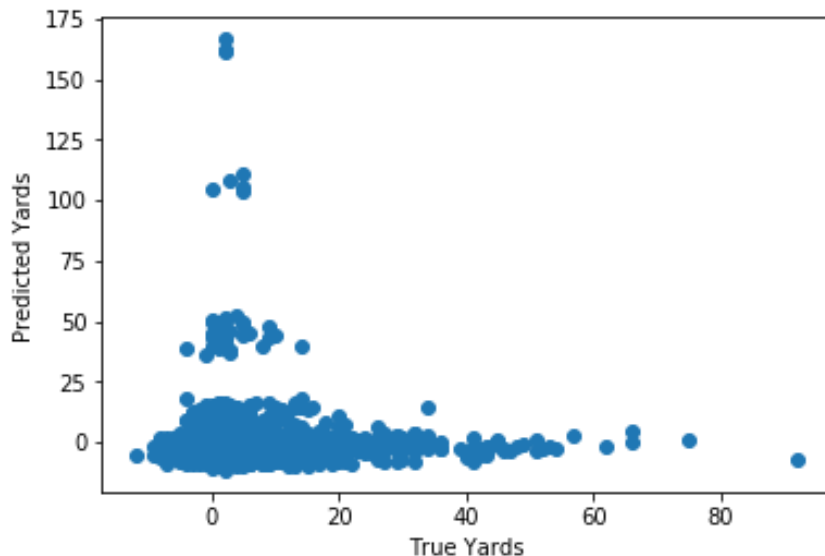


Data Characteristics



Model Comparison

Model:	Average Value	Linear Regression	Regression Tree	Neural Network
MSE:	38.01	37.24	38.73	47.88



Most important features

- Offensive formation
- Defensive backs

Conclusions

- Linear regression slightly improved from the zero-r model
- Next steps:
 - Robust feature extraction
 - Physics model