Is the NBA Combine a quality predictor of NBA Success?

Agenda

- Overview
- Research Questions
- High Level Findings
- Draft Position vs. Player Statistics
- Combine vs. NBA Statistics
- Summary
- Methodology
- Future Items
- Appendix

Overview

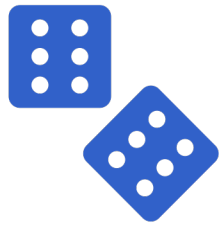
- The NBA combine is falling out of favor, with some superstars calling it a “waste of time”[1]
- Some prospects are opting out all together
- “Draft busts” seem as common as “draft gold”
- Goal of this analysis is to determine if there is a quantifiable relationship between combine statistics and in game performance
- Will use Combine data from 2009 – 2017 & NBA regular season statistics from the ‘09 – ’10 through ’18 - ‘19 seasons

ABC News (ABC News Network), accessed August 9, 2020, https://abcnews.go.com/Sports/kevin-durant-calls-nba-combine-waste-time-top/story?id=47338234

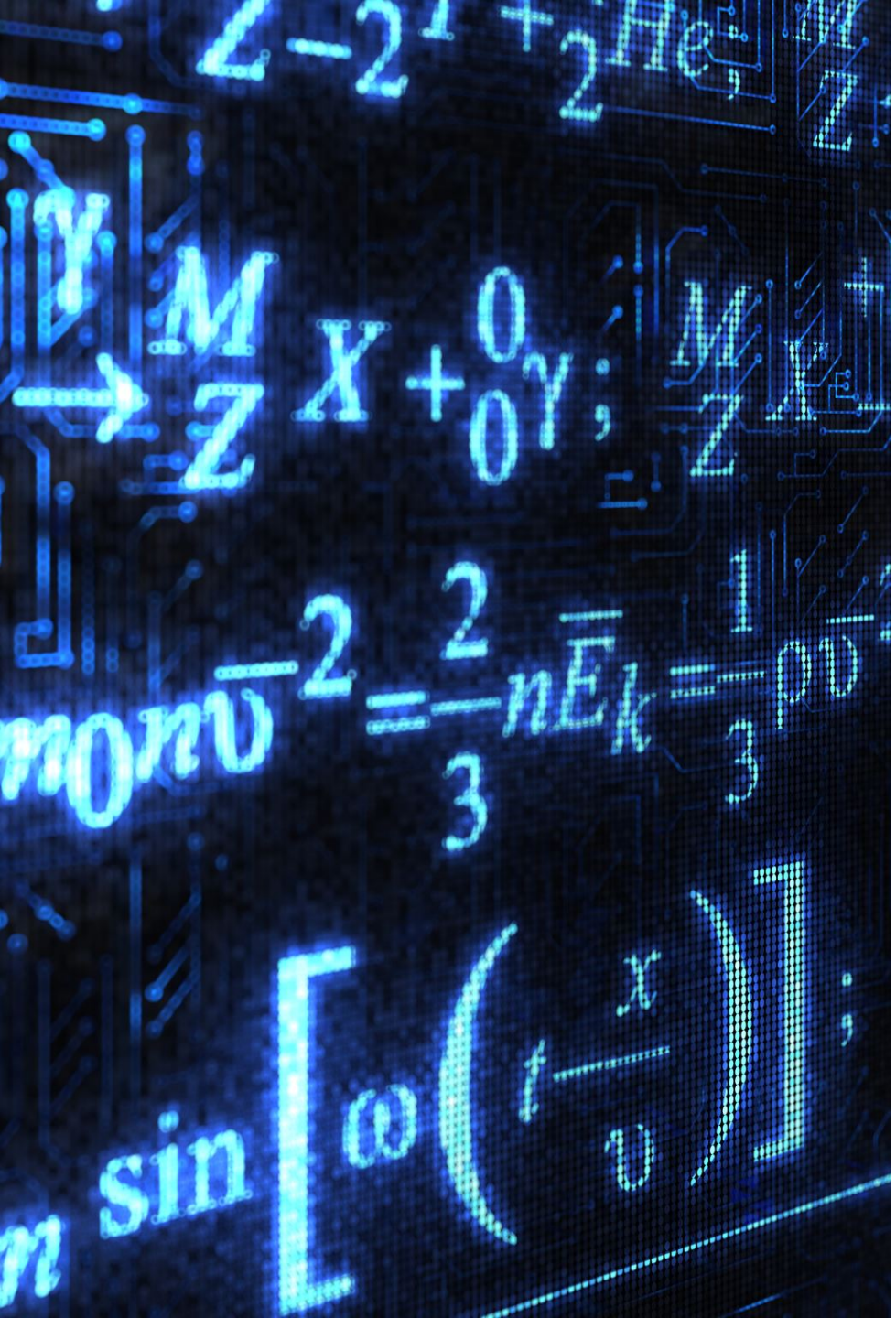
Research Questions

How good are the NBA's evaluation methods? E.g. is there a strong relationship between draft position and in-game performance?

Is there any relationship between combine performance and draft position? E.g. are players hurt by not going to the combine? Do combine performances hurt or help their draft stock?



Are there any strong relationships between combine measures and in-game statistics like points per game and win shares?



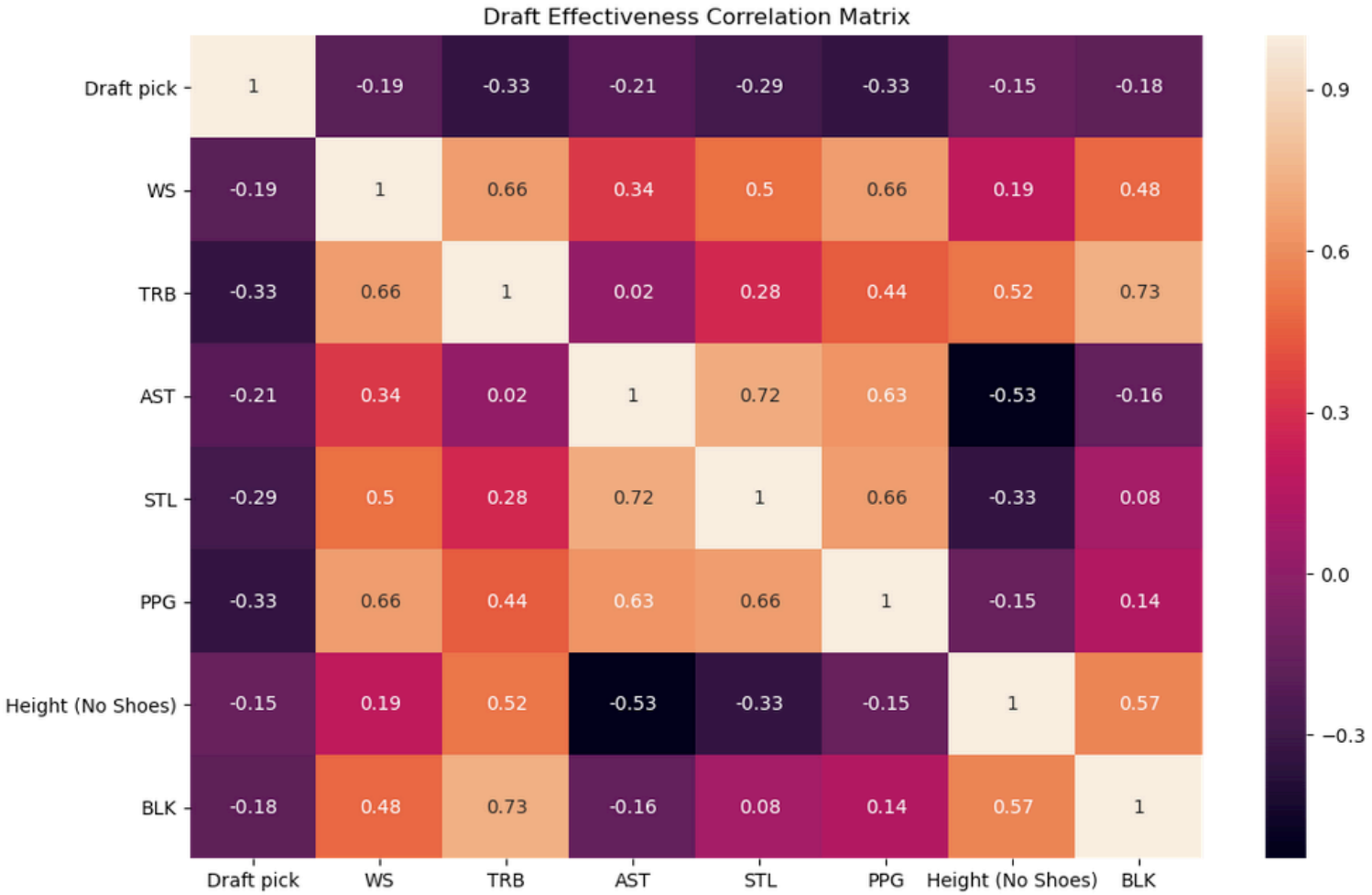
About Win Shares

- Win Shares Attempts to quantify or assign responsibility of certain number of a team's wins to an individual player
- It's a composite statistic that allows the evaluation of players on equal footing despite different playing styles
- It was chosen over other composite statistics like player efficiency ratio because it it has fewer anomalies or issues rewarding players for missed shots, if their overall shooting percentage is above a certain threshold
- This analysis used win shares data as reported by Basketball Encyclopedia

High Level Findings

- No significant relationship between combine measurements and win shares or points per game (PPG)
- No relationship between combine stats and draft position
- Top players were nearly always high draft picks, but the chance of a high draft being being above average for PPG or WS is 50/50 at best.
- On average only 1/3 of players taken in the first 15 picks were above average in terms of win share
- NBA's evaluation methods work on aggregate to identify cohorts where a top athlete will likely be, but tend to miss the mark for individual athletes
- Over the years studied, only 22.47% of the athletes taken with the first 15 picks were in the 75th percentile for win shares
- 29.21% of the athletes taken in the top 15 picks of the draft were in the 75th percentile for PPG

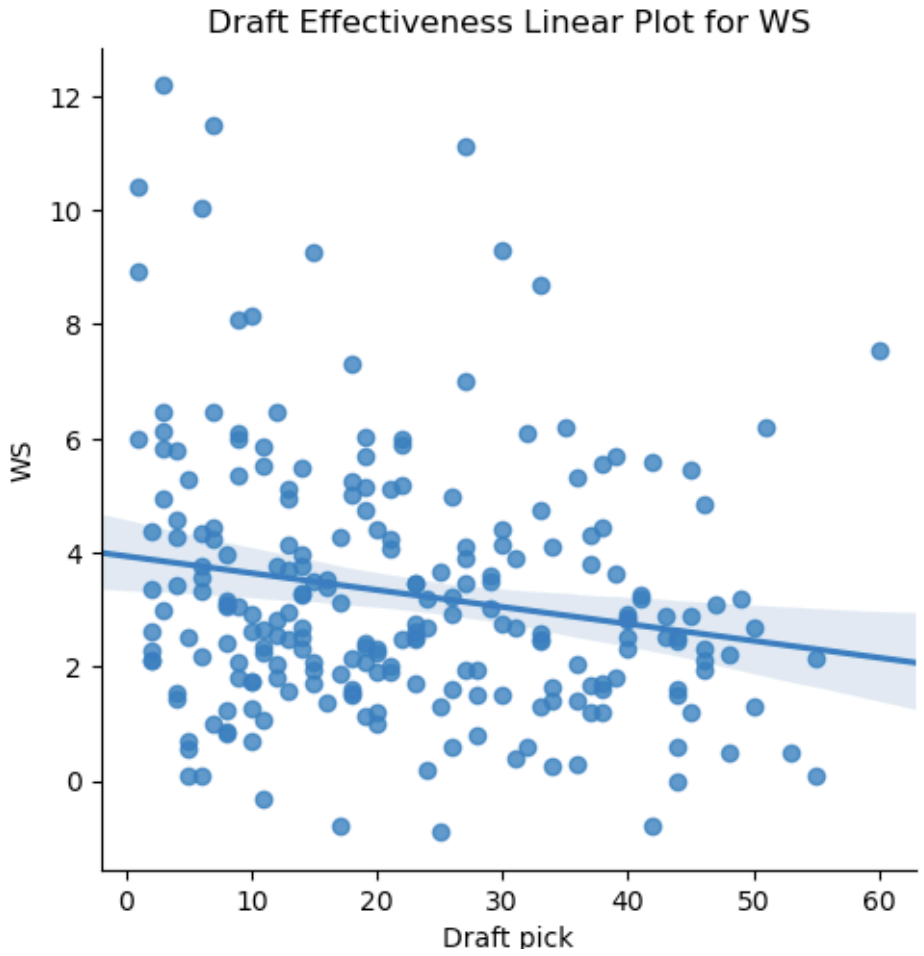
Draft Position vs. In Game Performance



- No significant relationship between draft position and in game performance
- As future slides will show there is a strong correlation for top players, but the correlation is weak overall

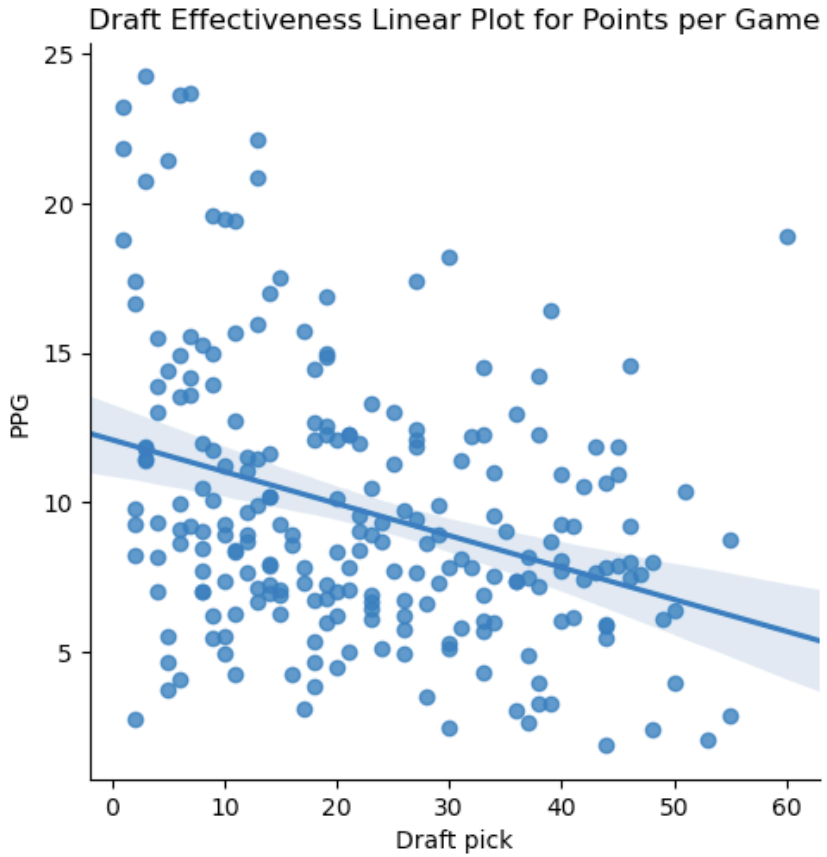
- Correlation coefficients indicate the strength of the relationship between two variables
- Positive = move in the same direction, negative = the opposite

Draft Position and Player Statistics - WS



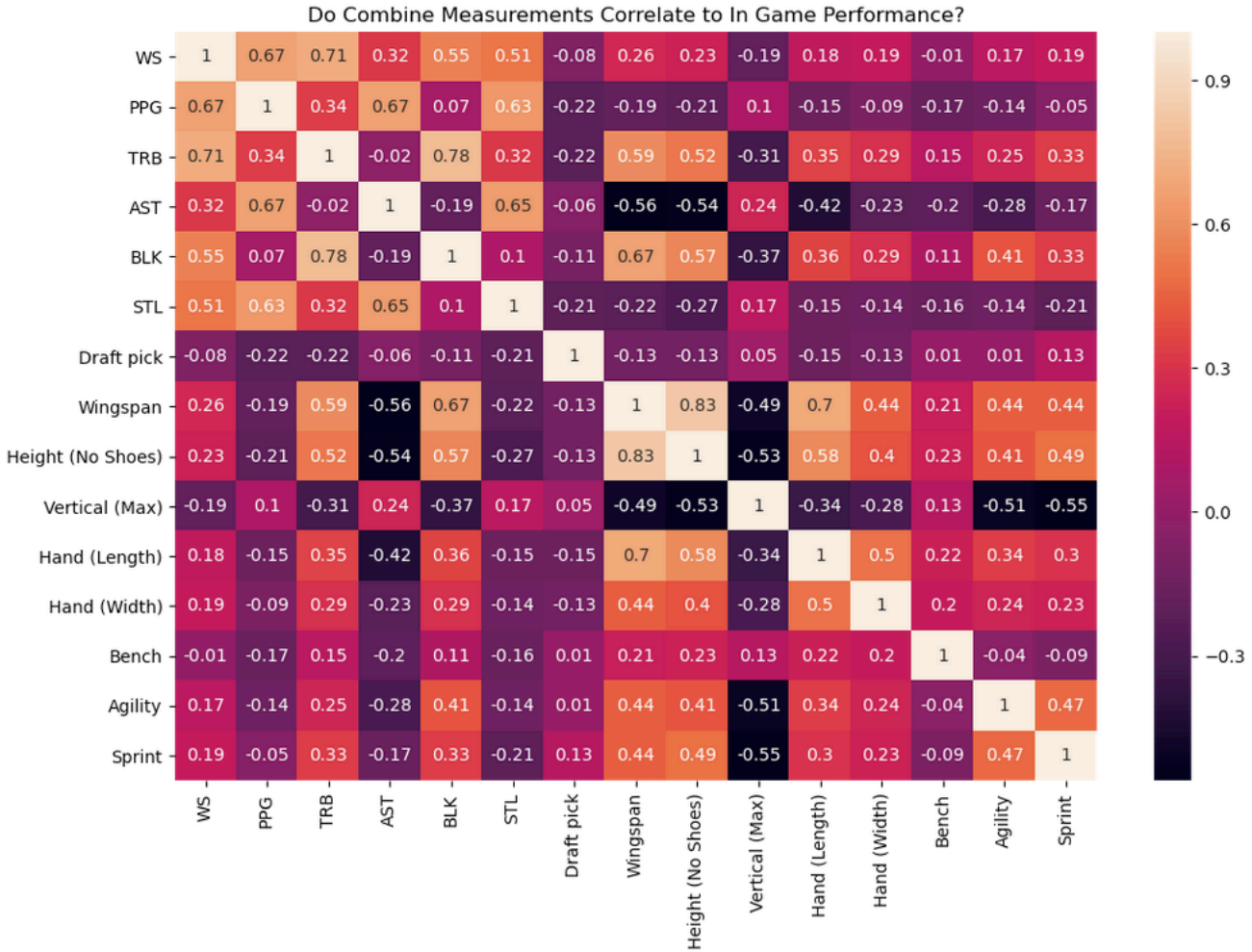
- The athletes with the highest average win shares from the ‘09 – ’10 through ‘18 – ’19 seasons were all selected at the top of the draft
- The overwhelming majority of high draft picks were below average for win shares (~77%)
- The NBA’s evaluation methods work on aggregate as far as identifying cohorts where stars will be, but tend to fail on an individual basis

Draft Position and Player Statistics - PPG



- Athletes with scoring averages over 20 were all taken with the first 15 picks
- Despite the above, 50/50 chance of a top 15 draft pick being above average for PPG
- Barely 1/3 of athletes taken in the top 15 of the draft will perform in the 75th percentile
- A possible implication is that the skill level to be in the top 25% is rare/isn't in most drafts
- NBA evaluation methods can identify cohorts over several drafts where the stars will be, but lack efficacy at the individual or draft level

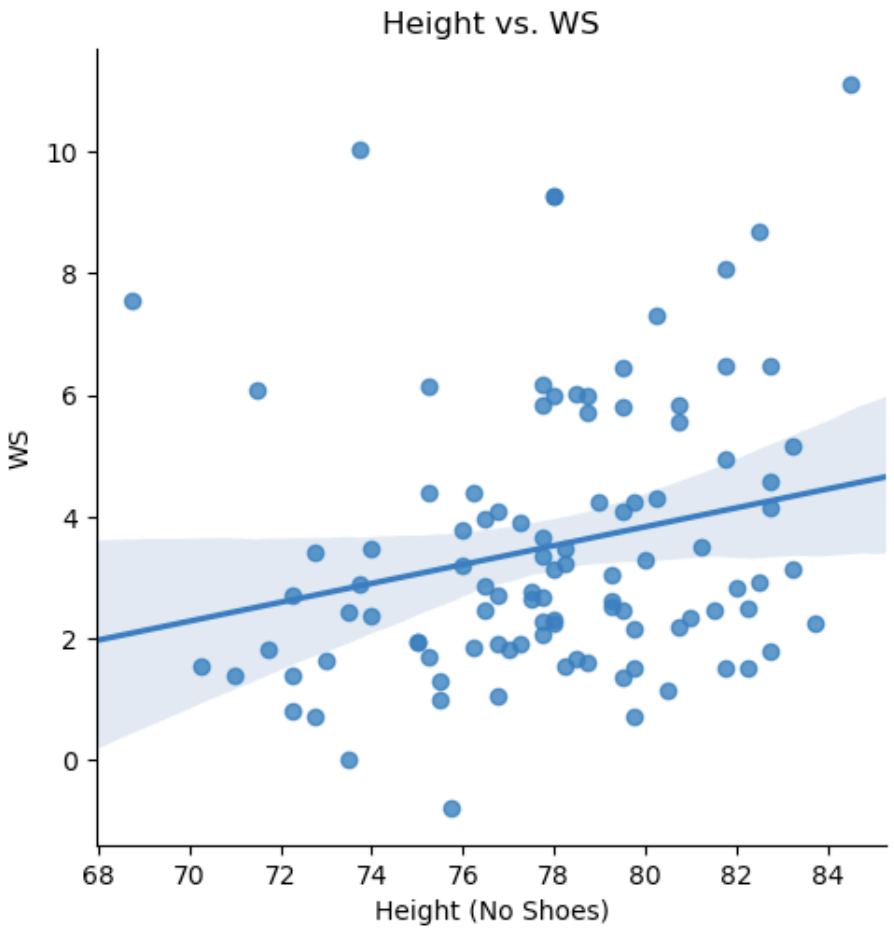
Combine and NBA Statistics – Correlations?



- The correlations with combine measurements and PPG and WS were weak at best
- No relationship between draft pick and combine performance
- Some interesting relationships were identified, E.g. despite the importance placed on an athlete's jumping ability the relationship was extremely weak with WS and PPG.
- None of the coefficients are strong enough to justify building a machine learning or regression model
- Only wingspan and rebounds have a "good" but not necessarily important relationship

- Correlation coefficients indicate the strength of the relationship between two variables
- Positive = move in the same direction, negative = the opposite

Win shares and height had the strongest relationship @ 0.23



- Rudy Goebert is 7’ 1” and has the highest WS in this cohort (11.1)
- 6’5” James Harden has a highest WS for players drafted during this time frame, but he doesn’t show up in the data to the left as he didn’t fully participate in the combine (12.18)
- 6’3” Steph Curry has the 2nd highest (11.48)
- Looking at athletes with WS ~2x the league average, heights range from 5’9 - 7’1
- 0.23 is too weak to be significant or useful for modeling
- Pursuing Machine linear and/or regression models isn’t supported by the math

Summary

- There is no significant relationship between measures at the NBA combine and in-game performance
- Even seemingly intuitive measures like agility, speed, height and jumping ability were extremely weak predictors of in game performance
- NBA evaluation methods are best at identifying the draft cohorts will superstars will be over the course of several drafts
- In a given draft most teams will fail at identifying players that will be above average, identifying athletes in the 75th or higher percentile is even more difficult
- The correlation coefficients that were the ”strongest” relatively speaking, were still too weak to be of use in a predictive model
- Only a minority % of athletes who attend the combine will participate in all events and many others are opting out all together

About The Data - Methodology

- '09 – ’10 → ‘18 – ’19 seasons[1]
- ‘09 → ‘17 combines[2]
- Took the average of the athlete’s 58 game seasons[3]
- There are typically only 360 active NBA players at any given time
- Used correlation matrices to identify relationships between metrics, where 0.70 is the correlation coefficient threshold to qualify as strong
- Used linear model plots to further illustrate relationships
- Ideally there would be strong relationships with enough variables to build a machine learning model
- ~247 total athletes in the combine cohort, but only ~100 participated in every studied combine event
- 692 total athletes in the NBA dataset, which reflects players drafted before 2009, undrafted athletes and athletes who either didn’t fully participate in the combine or skipped it entirely

1. Source: Basketball Encyclopedia
2. Source: Kaggle – web scraped data
3. Per the NBA an individual athlete must play in 58 games for their stats to be “significant”, e.g. scoring titles, rankings for points, assists and rebounds, etc.

Caveats

- Combine data was web scraped from NBA.com, so there is a chance that some data was missed
- Given that not all athletes attend or fully participate in the combine, it's "possible" that higher participation rates could change the data
- The draft position vs. in game statistics analysis could be significantly improved by having draft positions for all 692 athletes in the NBA player statistics data set as opposed to just the 247 we had draft data for.
- No adjustments were made for longevity or scenarios statistical averages were significantly hurt by poor seasons in the beginning and end of the athlete's career and/or from injuries

Future Items

- Use the NBA API to gather a dataset spanning 20 → 30 years for in game stats + combine performances.
- Add draft positions for all players, redo the draft position vs. performance analysis
- Repeat the analysis for the WNBA, does the women's league show the same patterns?

Appendix

Statistical Averages: 2009 → 2019 seasons

NBA Overall

Win shares

Mean: 4.19
Standard Deviation: 2.93
Highest: 19.3

PPG

Mean: 11.01
Standard Deviation: 5.53
Highest: 36.12

Combine Cohort

Win shares

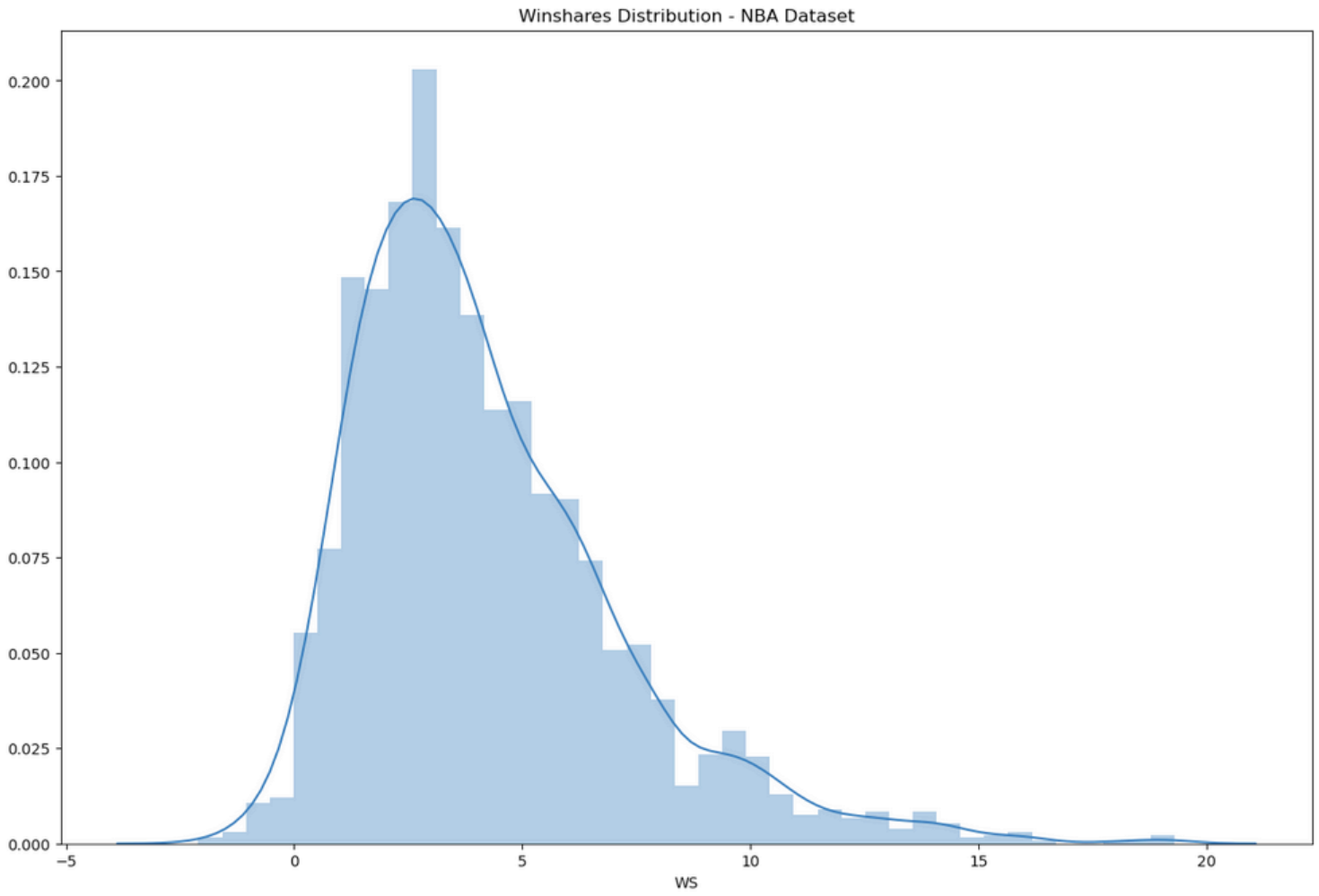
Mean: 3.22
Standard Deviation: 2.21
Highest: 12.18

PPG

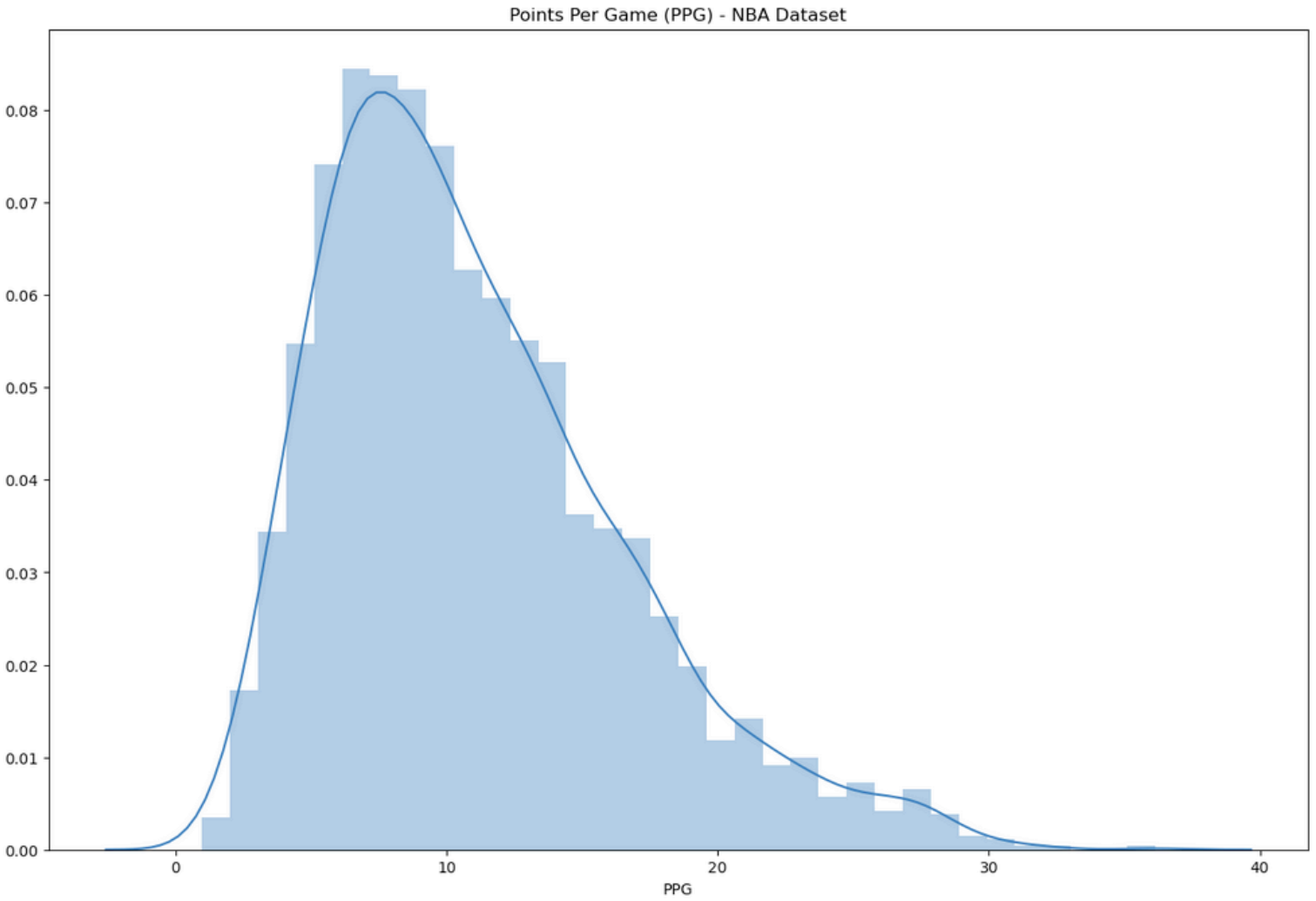
Mean: 9.58
Standard Deviation: 4.51
Highest: 24.22

- The NBA overall outperforming the combine cohort suggests the best players don't always go to the combine
- A more comprehensive dataset with draft pick numbers for all athletes playing in the NBA over that time period would improve the draft pick analysis

Win Shares Distribution – NBA Cohort



PPG Distribution – NBA Cohort



Detailed Summary Statistics - NBA Data

	WS	G	MP	TRB	AST	STL	BLK	PTS	TOV	PPG
count	2529.000000	2529.000000	2529.000000	2529.000000	2529.000000	2529.000000	2529.00000	2529.000000	2529.000000	2529.000000
mean	4.193199	72.222222	1850.041914	330.522341	172.988533	58.776592	37.51720	802.836299	104.901542	11.007937
std	2.938548	7.396039	586.837982	189.924047	142.893163	31.092214	35.49262	423.972868	58.351803	5.529088
min	-2.100000	58.000000	267.000000	22.000000	8.000000	6.000000	0.00000	62.000000	11.000000	0.968750
25%	2.100000	66.000000	1402.000000	192.000000	74.000000	36.000000	14.00000	484.000000	62.000000	6.870968
50%	3.600000	73.000000	1859.000000	284.000000	128.000000	53.000000	26.00000	724.000000	93.000000	9.920635
75%	5.700000	79.000000	2293.000000	419.000000	226.000000	76.000000	49.00000	1039.000000	134.000000	14.082192
max	19.300000	83.000000	3239.000000	1247.000000	907.000000	191.000000	269.00000	2818.000000	464.000000	36.128205

* Averages are higher than the combine cohort

Detailed Summary Statistics – Combine Cohort (All)

	WS	G	MP	ORB	DRB	TRB	AST	STL	BLK	PTS	TOV	PPG
count	247.000000	247.000000	247.000000	247.000000	247.000000	247.000000	247.000000	247.000000	247.000000	247.000000	247.000000	247.000000
mean	3.219196	71.596580	1666.202175	72.560825	219.791270	292.352095	145.027386	54.804547	33.213741	695.833171	91.141654	9.586418
std	2.217590	5.242624	483.893094	55.885412	103.443289	150.791097	111.933981	26.628137	29.874199	343.388525	50.941738	4.509504
min	-0.900000	58.000000	441.000000	6.000000	24.000000	32.000000	11.000000	9.000000	2.000000	111.000000	12.000000	1.881356
25%	1.775000	68.550000	1336.800000	35.250000	148.500000	190.250000	67.375000	34.416667	13.750000	475.000000	56.800000	6.679919
50%	2.750000	72.166667	1644.600000	55.600000	208.000000	261.000000	106.500000	51.000000	24.800000	626.000000	78.750000	8.644622
75%	4.268750	75.366667	2018.450000	92.666667	264.400000	359.500000	190.000000	68.850000	41.861111	879.033333	110.650000	11.973208
max	12.180000	82.000000	2843.857143	372.571429	688.000000	1060.571429	706.166667	135.125000	188.666667	1862.700000	292.500000	24.220056

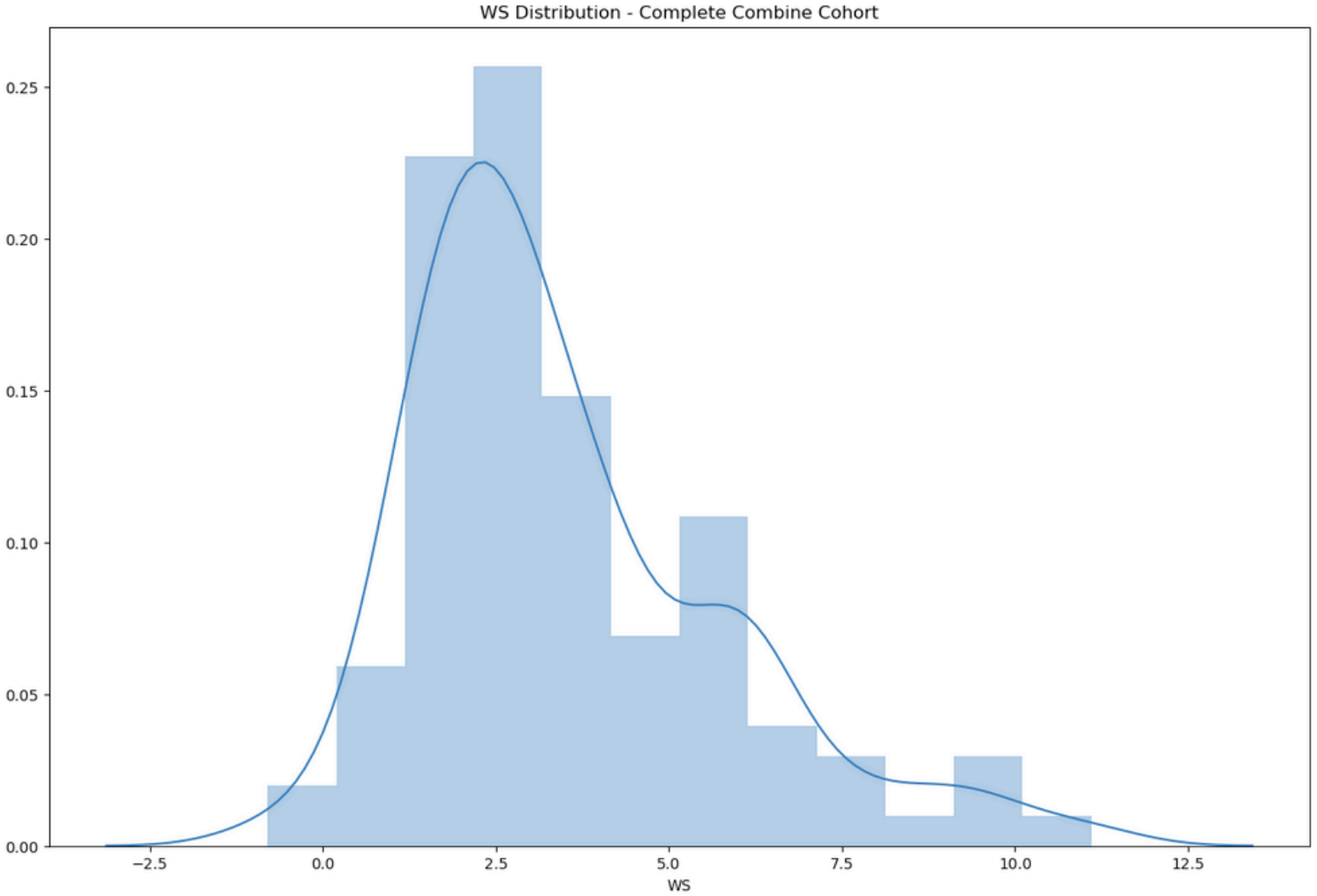
** All athletes who participated in the combine, even if they didn't participate in all combine events*

Detailed Summary Statistics – Complete Combine Cohort

	WS	G	MP	ORB	DRB	TRB	AST	STL	BLK	PTS	TOV	PPG
count	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000	102.000000
mean	3.507382	71.688994	1706.222771	80.681182	231.321771	312.002953	134.717678	55.003221	35.156435	707.429369	87.623125	9.712010
std	2.268007	5.246972	501.985141	67.842633	113.729105	173.164928	99.099288	25.554404	33.046846	340.869286	45.253674	4.440690
min	-0.800000	58.000000	441.000000	8.000000	24.000000	32.000000	22.000000	11.250000	2.000000	111.000000	20.000000	1.881356
25%	1.900000	68.083333	1340.700000	40.466667	152.016667	199.783333	64.462500	37.083333	15.000000	495.843750	60.500000	6.743794
50%	2.841667	72.633333	1719.071429	56.107143	210.187500	268.375000	101.866667	52.600000	25.750000	650.773810	74.683333	8.980595
75%	4.395000	75.906250	2085.375000	96.083333	280.250000	377.642857	176.178571	69.916667	42.041667	875.112500	108.383929	11.875959
max	11.100000	80.000000	2843.857143	372.571429	688.000000	1060.571429	497.000000	124.000000	188.666667	1844.142857	250.000000	23.635619

- *Small but not necessarily meaningful difference between WS and PPG averages, with the complete cohort being slightly better*
- *Complete refers to the cohort that participated in all combine events*

Win Shares Distribution – Complete Combine Cohort



** Complete refers to the cohort that participated in all combine events*