An analysis of college statistics vs. draft and actual NBA success across different eras

By Manan Khattar

Introduction

Most players in the NBA come from college basketball, and it is of great importance to NBA franchises to model how various college statistics translate to success in the NBA to determine how to draft college players. As a moderately enthusiastic basketball fan, I am interested in this relationship myself and as such, have kept track of historical trends in draft pick selection and how they change over time. From the mainstream media, I had always heard that wins in college was a surefire indicator of a player's draft position, and therefore success in the NBA—if a player couldn't make his team win in college, against much weaker opposition, then how was he going to lead his team against stronger, smarter, better athletes? However, with Ben Simmons going number 1 overall in the 2016 draft and Markelle Fultz number 1 overall in 2017 draft, with both players' college teams not performing well (Markelle Fultz's University of Washington team even had a below 50% winrate), I was surprised at this seeming departure from the conventional wisdom in recent years.

Based on this admittedly small sample size, I hypothesized that there might be some relationship between the current "era" of the NBA and the influence of a player's college winrate on his draft position. In the modern era, focusing on analytics, three-point shooting, and efficiency, perhaps wins in college were not as important a metric as various other advanced stats, but in previous years, with an emphasis on hand-check defense and post-up play, winrate could've

been a more important metric. I decided to test this hypothesis by building my own model.

Methodology

All of my data was scraped from basketball-reference.com and sports-reference.com/cbb using the BeautifulSoup package in Python and Rvest in R. To classify NBA eras, I first compiled a dataframe of NBA season averages of various statistical measures from the 1973-1974 season (the year the 3-point shot first started being tracked) upto the present day:

Filter

	FG	FGA	FT	FTA	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	FG%	FT%	Pace	eFG%	TOV%	ORB%	FT/FGA	ORtg
1	39.6	86.1	16.6	21.7	9.7	33.8	43.5	23.2	7.7	4.8	14.3	19.9	106.3	0.460	0.767	97.3	0.521	13.0	22.3	0.193	108.6
2	39.0	85.4	17.8	23.1	10.1	33.4	43.5	22.6	7.7	4.7	14.0	19.9	105.6	0.457	0.772	96.4	0.514	12.7	23.3	0.209	108.8
3	38.2	84.6	17.7	23.4	10.4	33.3	43.8	22.3	7.8	5.0	14.4	20.3	102.7	0.452	0.757	95.8	0.502	13.2	23.8	0.209	106.4
4	37.5	83.6	17.1	22.8	10.9	32.4	43.3	22.0	7.7	4.8	14.4	20.2	100.0	0.449	0.750	93.9	0.496	13.3	25.1	0.205	105.6
5	37.7	83.0	17.8	23.6	10.9	31.8	42.7	22.0	7.7	4.7	14.6	20.7	101.0	0.454	0.756	93.9	0.501	13.6	25.5	0.215	106.6
6	37.1	82.0	16.7	22.2	11.2	31.0	42.1	22.1	7.8	5.1	14.6	19.8	98.1	0.453	0.753	92.0	0.496	13.7	26.5	0.204	105.8
7	36.5	81.4	16.9	22.5	11.4	30.8	42.2	21.0	7.7	5.1	14.6	19.6	96.3	0.448	0.752	91.3	0.487	13.8	27.0	0.208	104.6
8	37.2	81.2	18.6	24.4	10.9	30.5	41.4	21.5	7.3	4.9	14.3	20.7	99.6	0.459	0.763	92.1	0.498	13.4	26.4	0.229	107.3
9	37.7	81.7	18.6	24.5	11.0	30.8	41.7	21.2	7.2	4.9	14.2	20.9	100.4	0.461	0.759	92.7	0.501	13.3	26.3	0.228	107.6
10	37.1	80.9	19.1	24.7	11.0	30.3	41.3	21.0	7.3	4.8	14.0	21.0	100.0	0.459	0.771	91.7	0.500	13.3	26.7	0.236	108.3
11	37.3	81.5	18.8	24.9	11.2	30.8	42.0	21.8	7.3	4.7	14.1	21.0	99.9	0.457	0.755	92.4	0.497	13.2	26.7	0.231	107.5
12	36.5	79.7	19.6	26.1	11.1	29.9	41.1	21.3	7.2	4.6	15.1	22.2	98.7	0.458	0.752	91.9	0.496	14.2	27.1	0.246	106.5
13	35.8	79.0	19.6	26.3	11.2	29.8	41.0	20.6	7.2	4.7	14.4	22.8	97.0	0.454	0.745	90.5	0.490	13.7	27.3	0.248	106.2
14	35.9	80.3	19.7	26.1	12.0	29.8	41.9	21.3	7.5	4.9	14.5	22.6	97.2	0.447	0.756	90.9	0.482	13.6	28.7	0.245	106.1
15	35.0	79.8	18.2	24.2	12.1	30.1	42.2	21.3	7.9	5.1	15.0	21.4	93.4	0.439	0.752	90.1	0.471	14.2	28.6	0.228	102.9
16	35.7	80.8	18.5	24.4	12.0	30.3	42.3	21.5	7.9	5.0	14.9	21.8	95.1	0.442	0.758	91.0	0.474	14.0	28.5	0.229	103.6

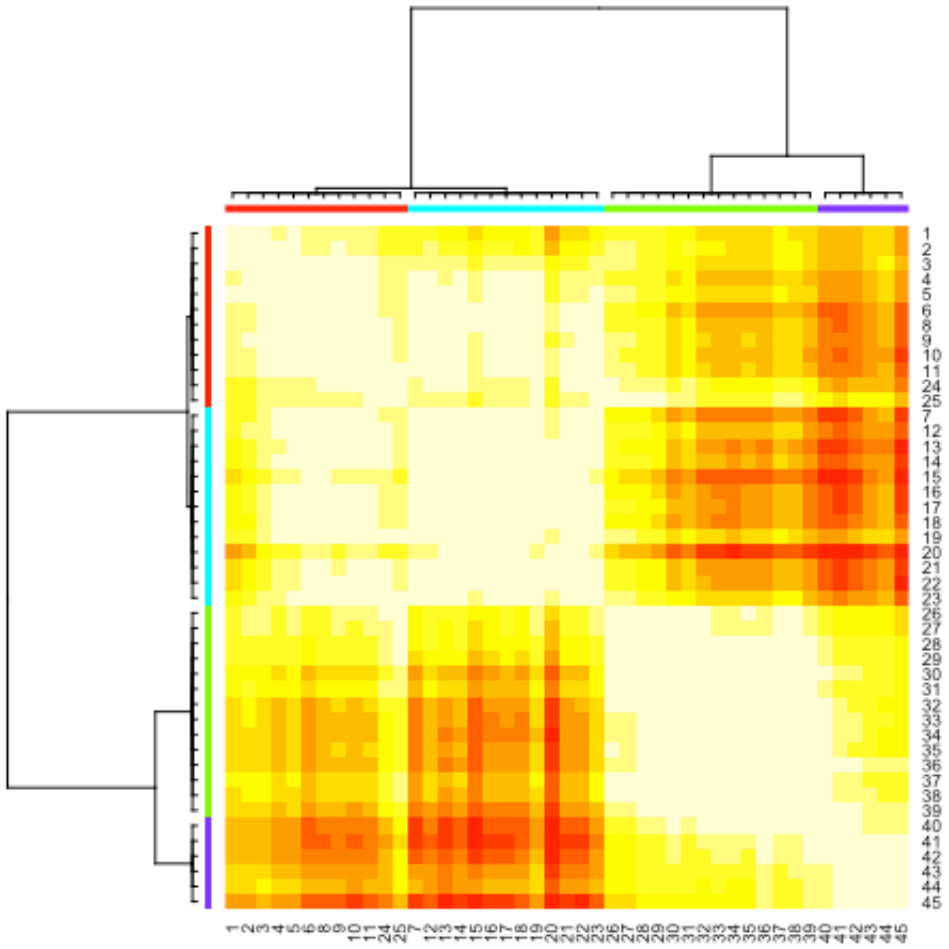
Showing 1 to 16 of 45 entries

I then used the apcluster package in R that performs examplar-based affinity propogation, an algorithm that clusters data without being explicitly provided the number of clusters, unlike K-means clustering. Based on the data provided, the algorithm sorted NBA seasons into the following four clusters, to which I've added my own descriptions:

Era	Seasons in Era
1 (Dark Ages, Cocaine)	1978, 1977, 1976, 1975, 1974, 1973
2 (High Pace Era)	1992, 1991, 1990, 1989, 1988, 1987, 1986, 1985, 1984, 1983, 1982, 1981, 1980, 1979
3 (Athletic	2011, 2006, 2005, 2004, 2003, 2002, 2001,

Revolution, Low Pace)	2000, 1999, 1998, 1997, 1996, 1995
4 (Offensive Evolution)	2017, 2016, 2015, 2014, 2013, 2012, 2010, 2009, 2008, 2007, 1994, 1993

And produced the following heatmap:



It's interesting to note that most years placed in different eras were in chronological order, with the exception of the 1993, 1994, and 2011 seasons. This makes contextual sense, as 1993-1994 were the seasons with a shorter 3-point line and thus were quite similar to the modern run-and-gun offense, while the 2011-2012 season was a lockout-shortened and injury-ravaged season that had a very slow, defensive style of play, atypical for its era.

Once I obtained rough categorizations for NBA eras, I then scraped player draft data for each season in each era, only looking at the first round draft picks (top 30 picks). For each drafted player who went to college, I scraped their college statistics page on sports-reference.com/cbb to obtain counting stats for their final season in college, and also introduced two additional metrics of my own—their winrate in their final season and the number of seasons they attended college. I then scraped basketball-reference.com to obtain the average winshares per 48 minutes for each of those players during their NBA careers, as a career-length-unbiased estimate of their productivity in the NBA. I further cleaned the data, and then ran a series of 8 regressions using Statsmodels.api in Python—one set for trying to predict the draft position of a player given a set of college parameters for each of the four eras, and the other set for trying to predict a player's winshares per 48 minutes in the NBA given college parameters for each of the four eras. The goal is to observe the relationship between draft position and college statistics (specifically college winrate), see if it varies across eras, and see if there is a similar relationship between college statistics and wins contributed in the NBA.

Results & Analysis

For regression between draft position and college statistics, here are the four regressions:

- **Era 1 (Dark Ages, Cocaine):**

OLS Regression Results

Dep. Variable:	draft_pick	**R-squared:**	0.239
Model:	OLS	**Adj. R-squared:**	0.137
Method:	Least Squares	**F-statistic:**	2.348
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	0.00579
Time:	01:37:29	**Log-Likelihood:**	-439.13
No. Observations:	128	**AIC:**	910.3
Df Residuals:	112	**BIC:**	955.9
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-369.9608	1119.312	-0.331	0.742	-2587.734	1847.812
Unnamed: 0	-0.0402	0.019	-2.138	0.035	-0.077	-0.003
FG	-12.2987	21.466	-0.573	0.568	-54.832	30.234
FG%	43.1528	61.172	0.705	0.482	-78.052	164.357
FGA	2.0401	2.053	0.994	0.322	-2.028	6.108
FT	-7.9646	11.223	-0.710	0.479	-30.201	14.272
FT%	11.5705	26.784	0.432	0.667	-41.498	64.639
FTA	2.3107	4.015	0.576	0.566	-5.644	10.265
G	-0.1931	0.243	-0.795	0.428	-0.674	0.288
PF	1.5567	1.543	1.009	0.315	-1.501	4.614
PTS	3.8427	11.179	0.344	0.732	-18.307	25.993
SOS	-0.8258	0.215	-3.832	0.000	-1.253	-0.399
Season	0.1860	0.568	0.327	0.744	-0.939	1.311
TRB	-0.5122	0.261	-1.966	0.052	-1.028	0.004
winrate	-2.7715	4.767	-0.581	0.562	-12.217	6.674
years_in_college	2.0667	1.261	1.639	0.104	-0.432	4.565

- **Era 2 (High Pace Era)**

Out[631]:

OLS Regression Results

Dep. Variable:	draft_pick	**R-squared:**	0.242
Model:	OLS	**Adj. R-squared:**	0.203
Method:	Least Squares	**F-statistic:**	6.202
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	7.94e-14
Time:	01:37:28	**Log-Likelihood:**	-1337.8
No. Observations:	389	**AIC:**	2716.
Df Residuals:	369	**BIC:**	2795.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1425.1768	345.442	-4.126	0.000	-2104.458	-745.896
Unnamed: 0	0.0095	0.005	1.829	0.068	-0.001	0.020
AST	-1.3124	0.413	-3.177	0.002	-2.125	-0.500
BLK	-2.0107	0.482	-4.173	0.000	-2.958	-1.063
FG	4.3821	2.422	1.809	0.071	-0.380	9.145
FG%	-2.4411	32.470	-0.075	0.940	-66.291	61.409
FGA	-0.4500	1.400	-0.321	0.748	-3.203	2.303
FT	3.4144	3.214	1.062	0.289	-2.906	9.734
FT%	3.9118	17.379	0.225	0.822	-30.263	38.086
FTA	-1.3320	2.337	-0.570	0.569	-5.928	3.264
G	-0.1262	0.140	-0.901	0.368	-0.402	0.149
MP	0.0685	0.196	0.349	0.728	-0.318	0.455
PF	-1.2099	0.959	-1.262	0.208	-3.095	0.675
PTS	-2.3319	0.814	-2.866	0.004	-3.932	-0.732
SOS	-0.6270	0.126	-4.986	0.000	-0.874	-0.380
STL	1.7142	0.873	1.964	0.050	-0.002	3.431
Season	0.7394	0.172	4.292	0.000	0.401	1.078
TRB	-0.5138	0.247	-2.081	0.038	-0.999	-0.028
winrate	-4.3762	3.940	-1.111	0.267	-12.124	3.371
years_in_college	-0.1041	0.629	-0.166	0.869	-1.340	1.132

Omnibus:	12.226	**Durbin-Watson:**	0.639
Prob(Omnibus):	0.002	**Jarque-Bera (JB):**	8.612
Skew:	0.243	**Prob(JB):**	0.0135
Kurtosis:	2.457	**Cond. No.**	1.75e+06

- **Era 3 (Athletic Revolution, Low Pace)**

Out[630]:

OLS Regression Results

Dep. Variable:	draft_pick	**R-squared:**	0.447
Model:	OLS	**Adj. R-squared:**	0.394
Method:	Least Squares	**F-statistic:**	8.339
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	6.03e-20
Time:	01:36:55	**Log-Likelihood:**	-859.51
No. Observations:	261	**AIC:**	1767.
Df Residuals:	237	**BIC:**	1853.
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	172.6671	884.901	0.195	0.845	-1570.609	1915.943
Unnamed: 0	0.0136	0.019	0.709	0.479	-0.024	0.052
2P	-19.7576	9.164	-2.156	0.032	-37.811	-1.704
2P%	45.8864	39.239	1.169	0.243	-31.416	123.189
2PA	1.2389	9.021	0.137	0.891	-16.533	19.010
3P	-33.8034	12.890	-2.622	0.009	-59.198	-8.409
3P%	-3.1471	3.897	-0.807	0.420	-10.825	4.531
3PA	-0.0562	9.101	-0.006	0.995	-17.985	17.872
AST	-0.7977	0.387	-2.060	0.040	-1.561	-0.035
BLK	-1.9595	0.662	-2.959	0.003	-3.264	-0.655
FG	-13.6457	13.710	-0.995	0.321	-40.655	13.363
FG%	-24.3210	41.676	-0.584	0.560	-106.424	57.782
FGA	1.3836	8.936	0.155	0.877	-16.221	18.988
FT	-19.4305	6.843	-2.839	0.005	-32.912	-5.949
FT%	25.1589	14.545	1.730	0.085	-3.495	53.813
FTA	3.7804	1.968	1.921	0.056	-0.097	7.657
G	-0.0643	0.124	-0.518	0.605	-0.309	0.180
PTS	13.7496	6.505	2.114	0.036	0.934	26.565
SOS	-0.0066	0.138	-0.047	0.962	-0.279	0.266
STL	0.4712	0.896	0.526	0.599	-1.294	2.236
Season	-0.0870	0.445	-0.196	0.845	-0.963	0.789
TRB	-0.2505	0.307	-0.815	0.416	-0.856	0.355
winrate	-12.9895	4.014	-3.236	0.001	-20.897	-5.082
years_in_college	3.6088	0.467	7.725	0.000	2.688	4.529

Omnibus:	4.815	**Durbin-Watson:**	1.037
Prob(Omnibus):	0.090	**Jarque-Bera (JB):**	6.400
Skew:	0.060	**Prob(JB):**	0.0408
Kurtosis:	3.758	**Cond. No.**	4.21e+06

- **Era 4: (Offensive Evolution)**

Out[629]:

OLS Regression Results

Dep. Variable:	draft_pick	**R-squared:**	0.344
Model:	OLS	**Adj. R-squared:**	0.278
Method:	Least Squares	**F-statistic:**	5.212
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	1.59e-12
Time:	01:36:41	**Log-Likelihood:**	-917.36
No. Observations:	274	**AIC:**	1887.
Df Residuals:	248	**BIC:**	1981.
Df Model:	25		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1656.5938	513.004	3.229	0.001	646.193	2666.994
Unnamed: 0	0.0411	0.011	3.713	0.000	0.019	0.063
2P	8.8897	9.751	0.912	0.363	-10.316	28.095
2P%	-31.8829	31.136	-1.024	0.307	-93.208	29.442
2PA	-1.2758	10.044	-0.127	0.899	-21.059	18.508
3P	7.8157	13.142	0.595	0.553	-18.069	33.700
3PA	-0.1735	9.845	-0.018	0.986	-19.565	19.218
AST	-1.0885	0.462	-2.357	0.019	-1.998	-0.179
BLK	-2.3593	0.791	-2.984	0.003	-3.917	-0.802
FG	-2.8186	14.246	-0.198	0.843	-30.877	25.240
FG%	13.1548	35.860	0.367	0.714	-57.475	83.785
FGA	0.7523	9.925	0.076	0.940	-18.796	20.301
FT	6.3369	6.975	0.909	0.364	-7.401	20.075
FT%	-13.7245	14.293	-0.960	0.338	-41.875	14.426
FTA	-3.3098	2.124	-1.558	0.120	-7.493	0.873
G	-0.0364	0.118	-0.309	0.758	-0.268	0.196
MP	0.4664	0.164	2.849	0.005	0.144	0.789
PF	1.6394	1.152	1.423	0.156	-0.629	3.908
PTS	-3.1284	6.434	-0.486	0.627	-15.801	9.544
SOS	-0.2496	0.151	-1.656	0.099	-0.546	0.047
STL	-0.1369	1.133	-0.121	0.904	-2.367	2.094
Season	-0.8110	0.257	-3.160	0.002	-1.317	-0.305
TOV	-0.9314	1.165	-0.800	0.425	-3.225	1.362
TRB	-0.3223	0.314	-1.026	0.306	-0.941	0.296
winrate	-10.3372	4.238	-2.439	0.015	-18.684	-1.990
years_in_college	2.8783	0.470	6.120	0.000	1.952	3.805

Omnibus:	0.695	**Durbin-Watson:**	0.813
Prob(Omnibus):	0.706	**Jarque-Bera (JB):**	0.788
Skew:	0.028	**Prob(JB):**	0.674
Kurtosis:	2.743	**Cond. No.**	2.40e+06

General Observations:

1. In general, the model meets expectations: steals, blocks, points, rebounds, and assists are all negatively related to draft position, meaning the better a player's stats in these categories, the lower (better) his draft position.
2. In terms of regular counting stats, blocks seem to be by far the most definitively correlated with draft position (lowest p-value), at least for the last three eras where blocks data was available.
3. Percentage field goal categories have consistently lower p-values than their counting stats counterparts, suggesting that having gaudy raw stats is less indicative of high draft position than someone who is selective and efficient with their shots.
4. Years in college has as strong a correlation or even higher a correlation than winrate in college for better (lower) draft positions, suggesting that even if a player has not achieved a lot of success in college, if he is young and shows potential, teams are much more likely to pick him than a junior or senior putting up better stats.
5. Contrary to my initial hypothesis it seems that, despite the very recent trend against the importance of college winrate led by Markelle Fultz and Ben Simmons, college winrate is about as important a factor as ever when it comes to determining player draft position. In the first two eras the p-values for college winrate were 0.562 and 0.267, which are both not statistically significant. In the last 25 years or so, however, the p-value between winrate and draft position has skyrocketed (so-to-speak) to 0.001 and 0.015, indicating a much more definite correlation that is statistically significant.

For regression between winshares per 48 minutes in the NBA and college statistics, here are the four regressions:

- **Era 1 (Dark Ages, Cocaine):**

OLS Regression Results

Dep. Variable:	ws_per_48	**R-squared:**	0.233
Model:	OLS	**Adj. R-squared:**	0.130
Method:	Least Squares	**F-statistic:**	2.268
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	0.00784
Time:	14:55:09	**Log-Likelihood:**	239.74
No. Observations:	128	**AIC:**	-447.5
Df Residuals:	112	**BIC:**	-401.8
Df Model:	15		
Covariance Type:	nonrobust		

-

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.5658	5.567	0.641	0.523	-7.464	14.596
Unnamed: 0	0.0002	9.34e-05	1.914	0.058	-6.26e-06	0.000
FG	0.0768	0.107	0.719	0.474	-0.135	0.288
FG%	-0.3246	0.304	-1.067	0.288	-0.927	0.278
FGA	-0.0177	0.010	-1.732	0.086	-0.038	0.003
FT	0.0034	0.056	0.061	0.952	-0.107	0.114
FT%	0.1631	0.133	1.224	0.223	-0.101	0.427
FTA	0.0162	0.020	0.811	0.419	-0.023	0.056
G	0.0004	0.001	0.367	0.714	-0.002	0.003
PF	-0.0068	0.008	-0.892	0.374	-0.022	0.008
PTS	-0.0209	0.056	-0.376	0.708	-0.131	0.089
SOS	0.0030	0.001	2.767	0.007	0.001	0.005
Season	-0.0018	0.003	-0.623	0.534	-0.007	0.004
TRB	0.0029	0.001	2.217	0.029	0.000	0.005
winrate	0.0084	0.024	0.355	0.723	-0.039	0.055
years_in_college	-0.0088	0.006	-1.404	0.163	-0.021	0.004

- **Era 2 (High Pace Era)**:

```
In [69]: get_summary(2, 'ws_per_48')
```

Out[69]:

OLS Regression Results

Dep. Variable:	ws_per_48	**R-squared:**	0.333
Model:	OLS	**Adj. R-squared:**	0.299
Method:	Least Squares	**F-statistic:**	9.692
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	6.63e-23
Time:	14:55:01	**Log-Likelihood:**	564.56
No. Observations:	389	**AIC:**	-1089.
Df Residuals:	369	**BIC:**	-1010.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	14.9896	2.597	5.771	0.000	9.882	20.097
Unnamed: 0	0.0001	3.91e-05	3.729	0.000	6.89e-05	0.000
AST	0.0046	0.003	1.492	0.136	-0.001	0.011
BLK	0.0205	0.004	5.653	0.000	0.013	0.028
FG	-0.0184	0.018	-1.008	0.314	-0.054	0.017
FG%	-0.0254	0.244	-0.104	0.917	-0.506	0.455
FGA	-0.0005	0.011	-0.050	0.960	-0.021	0.020
FT	-0.0346	0.024	-1.432	0.153	-0.082	0.013
FT%	0.0207	0.131	0.158	0.874	-0.236	0.278
FTA	0.0293	0.018	1.670	0.096	-0.005	0.064
G	-0.0001	0.001	-0.103	0.918	-0.002	0.002
MP	-0.0017	0.001	-1.125	0.261	-0.005	0.001
PF	-0.0018	0.007	-0.253	0.801	-0.016	0.012
PTS	0.0090	0.006	1.471	0.142	-0.003	0.021
SOS	0.0036	0.001	3.775	0.000	0.002	0.005
STL	0.0167	0.007	2.539	0.012	0.004	0.030
Season	-0.0076	0.001	-5.860	0.000	-0.010	-0.005
TRB	-0.0009	0.002	-0.499	0.618	-0.005	0.003
winrate	-0.0043	0.030	-0.144	0.885	-0.063	0.054
years_in_college	0.0220	0.005	4.651	0.000	0.013	0.031

- **Era 3 (Athletic Revolution, Low Pace):**

OLS Regression Results

Dep. Variable:	ws_per_48	**R-squared:**	0.415
Model:	OLS	**Adj. R-squared:**	0.358
Method:	Least Squares	**F-statistic:**	7.313
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	2.28e-17
Time:	14:52:32	**Log-Likelihood:**	379.07
No. Observations:	261	**AIC:**	-710.1
Df Residuals:	237	**BIC:**	-624.6
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.9948	7.690	0.519	0.604	-11.155	19.145
Unnamed: 0	8.65e-05	0.000	0.517	0.605	-0.000	0.000
2P	0.0338	0.080	0.425	0.671	-0.123	0.191
2P%	0.1771	0.341	0.519	0.604	-0.495	0.849
2PA	-0.0564	0.078	-0.720	0.472	-0.211	0.098
3P	0.1314	0.112	1.173	0.242	-0.089	0.352
3P%	-0.0002	0.034	-0.006	0.995	-0.067	0.067
3PA	-0.0523	0.079	-0.661	0.509	-0.208	0.103
AST	0.0062	0.003	1.834	0.068	-0.000	0.013
BLK	0.0033	0.006	0.572	0.568	-0.008	0.015
FG	0.1625	0.119	1.364	0.174	-0.072	0.397
FG%	-0.0344	0.362	-0.095	0.924	-0.748	0.679
FGA	0.0489	0.078	0.630	0.529	-0.104	0.202
FT	0.0846	0.059	1.422	0.156	-0.033	0.202
FT%	0.1286	0.126	1.018	0.310	-0.120	0.378
FTA	0.0106	0.017	0.619	0.536	-0.023	0.044
G	9.614e-05	0.001	0.089	0.929	-0.002	0.002
PTS	-0.0967	0.057	-1.710	0.089	-0.208	0.015
SOS	-0.0013	0.001	-1.065	0.288	-0.004	0.001
STL	0.0104	0.008	1.330	0.185	-0.005	0.026
Season	-0.0021	0.004	-0.535	0.593	-0.010	0.006
TRB	0.0087	0.003	3.255	0.001	0.003	0.014
winrate	0.0245	0.035	0.703	0.482	-0.044	0.093
years_in_college	-0.0082	0.004	-2.024	0.044	-0.016	-0.000

- **Era 4: (Offensive Evolution):**

OLS Regression Results

Dep. Variable:	ws_per_48	**R-squared:**	0.275
Model:	OLS	**Adj. R-squared:**	0.201
Method:	Least Squares	**F-statistic:**	3.754
Date:	Thu, 03 May 2018	**Prob (F-statistic):**	3.62e-08
Time:	14:52:25	**Log-Likelihood:**	470.88
No. Observations:	274	**AIC:**	-889.8
Df Residuals:	248	**BIC:**	-795.8
Df Model:	25		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-3.9312	3.234	-1.216	0.225	-10.301	2.438
Unnamed: 0	-0.0001	6.97e-05	-1.496	0.136	-0.000	3.3e-05
2P	0.0616	0.061	1.003	0.317	-0.059	0.183
2P%	-0.1514	0.196	-0.771	0.441	-0.538	0.235
2PA	-0.1128	0.063	-1.781	0.076	-0.237	0.012
3P	-0.0202	0.083	-0.243	0.808	-0.183	0.143
3PA	-0.0895	0.062	-1.441	0.151	-0.212	0.033
AST	-0.0041	0.003	-1.402	0.162	-0.010	0.002
BLK	0.0060	0.005	1.201	0.231	-0.004	0.016
FG	-0.0996	0.090	-1.110	0.268	-0.277	0.077
FG%	0.2502	0.226	1.107	0.270	-0.195	0.695
FGA	0.0946	0.063	1.512	0.132	-0.029	0.218
FT	0.0055	0.044	0.124	0.901	-0.081	0.092
FT%	-0.1389	0.090	-1.542	0.124	-0.316	0.039
FTA	-0.0268	0.013	-2.005	0.046	-0.053	-0.000
G	-0.0005	0.001	-0.713	0.477	-0.002	0.001
MP	0.0012	0.001	1.180	0.239	-0.001	0.003
PF	-0.0177	0.007	-2.439	0.015	-0.032	-0.003
PTS	0.0350	0.041	0.863	0.389	-0.045	0.115
SOS	-0.0015	0.001	-1.610	0.109	-0.003	0.000
STL	0.0276	0.007	3.871	0.000	0.014	0.042
Season	0.0020	0.002	1.261	0.208	-0.001	0.005
TOV	-0.0058	0.007	-0.789	0.431	-0.020	0.009
TRB	0.0012	0.002	0.610	0.543	-0.003	0.005
winrate	0.0326	0.027	1.220	0.224	-0.020	0.085
years_in_college	-0.0056	0.003	-1.874	0.062	-0.011	0.000

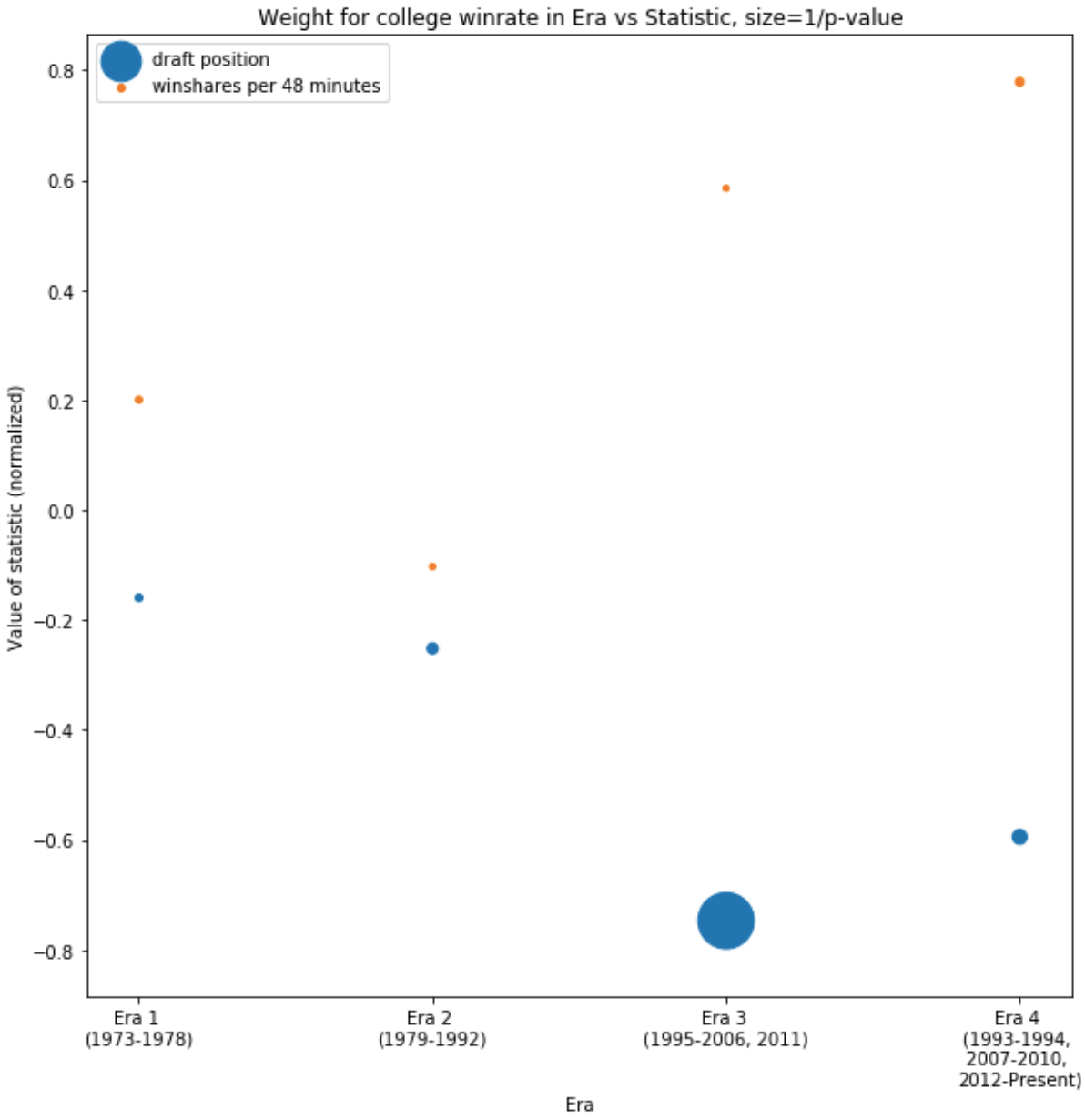
-

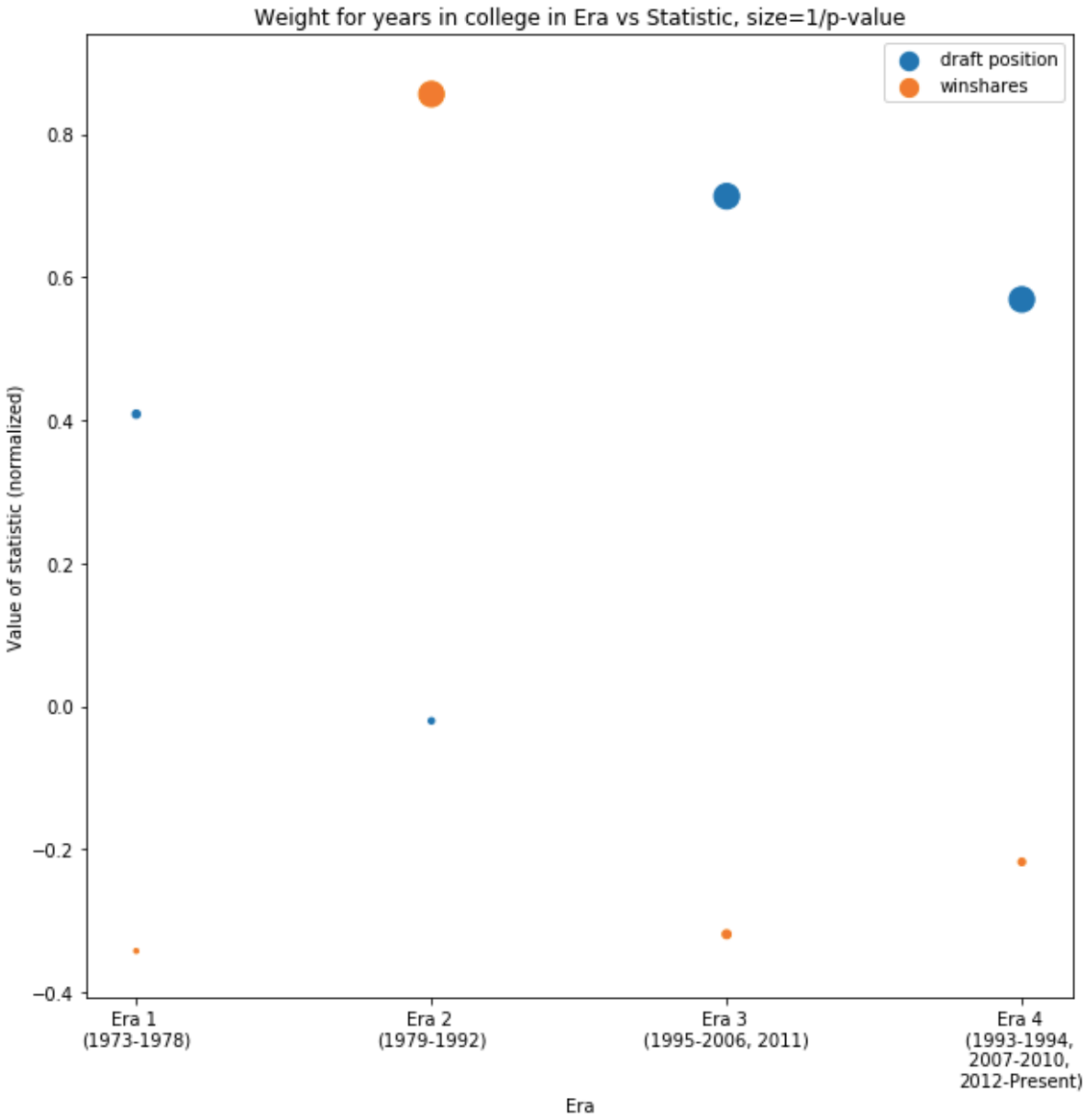
General Observations:

1. Compared to draft position, winshares per 48 minutes has significantly worse p-values across the board, for most statistical measures. This shows that as expected, college success can correlate to draft positions relatively

well but it is much harder to predict success in the NBA from college statistics alone.
2. For earlier eras, rebounds and blocks had significant correlations to NBA win shares, which qualitatively makes sense because with the high pace, poor shooting and emphasis on defense, both these stats were very relavant to the success of a team at the NBA level.
3. For winshares, winrate in college has consistently less significant correlation than years in college, suggesting that this is one aspect that NBA execs get right during the draft—younger players with potential do better in the NBA than older players on better teams.
4. In the modern era (2007-Present), only steals (very strong positive correlation) and personal fouls (strong negative correlation) have coefficients in the regression that are statistically significant. This indicates that it is getting harder to determine NBA success through college statistics, especially with the massive difference between modern NBA training and coaching and college coaching. Qualitatively, there seem to be a high number of lottery pick busts in recent times, and conversely many players (such as Kyle Kuzma and Donovan Mitchell from this year's draft) whose college stats didn't stand out and got picked late in the draft but have been tearing it up at the NBA level. NBA general managers have a seriously difficult task on their hands.

College Winrate vs Years in College


Weight for college winrate in Era vs Statistic, size=1/p-value
draft position
winshares per 48 minutes
Value of statistic (normalized)
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
Era 1
(1973-1978)
Era 2
(1979-1992)
Era 3
(1995-2006, 2011)
Era 4
(1993-1994,
2007-2010,
2012-Present)
Era



These two graphs summarize the data for the two custom metrics (college winrate and years in college) and their correlation to draft position and wins in the NBA. The y-value for each point reprsents the (normalized) coefficient for the variable in the regression, so a larger y-value represents a comparatively larger weight for that variable in that era compared to another value. The size of point, on the other hand, represents the inverse of the p-value—in other words, the larger the data point, the more statistically significant the correlation is for that variableOne can see that, with time, the weight for college winrates seem to have increased in the regression for draft position and winshares—that is, a better winrate in college correlates with a relatively better draft position and more wins in the NBA than it did in the past. However, apart from the weight for

draft position in Era 3, none of these values have statistically significant p-values, showing that there is high degree of variance in the relationship between college winrate and each of the dependent parameters (draft position and winshares per 48 minutes in the NBA).

In comparison, the correlation values for years in college are decreasing in magnitude over eras, showing that years in college doesn't have as strong a correlation to draft position and wins in the NBA as years past. However, the p-values in recent eras are statistically significant, showing a more definite relationship than college winrate to each of our dependent parameters.

Concluding Thoughts

My analysis of era-based correlation between college statistics and draft position/NBA winrate proved quite fruitful. Some of the most interesting results I found were that, contrary to my initial hypothesis, the strength of correlation between college winrate and draft position increased over time, not decreased as I originally thought. In addition, even though years spent in college does not measure a player's performance or skills in the NBA in any way, it has a more statistically significant correlation to draft position and winshares in the NBA than college winrate does, across most eras. This sort of era-based approach to handling NBA data can be a very useful tool for NBA executives and general managers to see the general drafting patterns that teams have had over time, how those change with trends in eras, and how they can use those trends to better find the players who will perform on their teams at a high level.