NBA vs NCAA

By Benjamin Nguyen, Yamah Karim, Kerwin Hsu, Tanner Owen

Questions:

1. Do NBA players shoot better than NCAA players?

2. What offensive stats correlate with more wins for NCAA and NBA teams?

3. What defensive stats correlate with more wins for the NCAA and NBA teams?

4. Do teams with superstars perform better than teams with only role players (NCAA and NBA)?

Data Cleaning

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

nba_17_18 = pd.read_csv("NBA data per game (team)/17-18 NBA team per.csv")
nba_16_17 = pd.read_csv("NBA data per game (team)/16-17 NBA team per.csv")
nba_15_16 = pd.read_csv("NBA data per game (team)/15-16 NBA team per.csv")
nba_14_15 = pd.read_csv("NBA data per game (team)/14-15 NBA team per.csv")
nba_13_14 = pd.read_csv("NBA data per game (team)/13-14 NBA team per.csv")

nba_standings_17_18 = pd.read_csv('NBA data (standings)/17-18 NBA standings.csv')
nba_standings_16_17 = pd.read_csv('NBA data (standings)/16-17 NBA standings.csv')
nba_standings_15_16 = pd.read_csv('NBA data (standings)/15-16 NBA standings.csv')
nba_standings_14_15 = pd.read_csv('NBA data (standings)/14-15 NBA standings.csv')
nba_standings_13_14 = pd.read_csv('NBA data (standings)/13-14 NBA standings.csv')

nba_13_14.head()
```

	Rk	Team	G	MP	FG	FGA	FG%	3P	3PA	3P%	...	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
0	1.0	Los Angeles Clippers*	82	240.9	39.1	82.5	0.474	8.5	24.0	0.352	...	0.730	10.5	32.5	43.0	24.6	8.6	4.8	13.9	21.5	107.9
1	2.0	Houston Rockets*	82	241.8	38.0	80.5	0.472	9.5	26.6	0.358	...	0.712	11.2	34.1	45.3	21.4	7.6	5.6	16.1	20.4	107.7
2	3.0	Minnesota Timberwolves	82	242.1	38.9	87.5	0.444	7.3	21.4	0.341	...	0.778	12.5	32.2	44.7	23.9	8.8	3.6	13.9	18.3	106.9
3	4.0	Portland Trail Blazers*	82	242.1	39.1	87.0	0.450	9.4	25.3	0.372	...	0.815	12.5	34.0	46.4	23.2	5.5	4.7	13.7	19.2	106.7
4	5.0	Oklahoma City Thunder*	82	241.5	39.0	82.7	0.471	8.1	22.4	0.361	...	0.806	10.8	33.9	44.7	21.9	8.3	6.1	15.3	22.7	106.2

```python
nba_standings_13_14.head()
```

	Rk	Team	Age	W	L	PW	PL	MOV	SOS	SRS	...	TOV%	ORB%	FT/FGA	eFG%.1	TOV%.1	DRB%	FT/FGA.1	Arena	A
0	1.0	San Antonio Spurs*	28.9	62.0	20.0	61	21	7.72	0.28	8.00	...	13.5	22.7	0.188	0.482	12.8	76.4	0.184	AT&T Center	7!
1	2.0	Los Angeles Clippers*	28.1	57.0	25.0	59	23	6.98	0.30	7.27	...	12.7	25.0	0.258	0.484	13.8	72.5	0.222	STAPLES Center	7!
2	3.0	Oklahoma City Thunder*	26.2	59.0	23.0	58	24	6.34	0.32	6.66	...	14.0	26.5	0.244	0.488	13.9	75.6	0.221	Chesapeake Energy Arena	7·
3	4.0	Golden State Warriors*	26.3	51.0	31.0	54	28	4.80	0.34	5.15	...	13.8	25.1	0.186	0.477	13.6	76.3	0.224	Oracle Arena	8(
4	5.0	Houston Rockets*	25.4	54.0	28.0	53	29	4.56	0.50	5.06	...	14.6	27.4	0.275	0.489	12.5	74.1	0.193	Toyota Center	7·

5 rows × 28 columns

```python
standings_13_14 = pd.DataFrame(nba_standings_13_14[['Team', 'W', 'L']])

standings_13_14 = standings_13_14.sort_values(by='W', ascending=False)
standings_13_14_combined = standings_13_14.merge(nba_13_14, on='Team')
standings_13_14_combined.head()
```

	Team	W	L	Rk	G	MP	FG	FGA	FG%	3P	...	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
0	San Antonio Spurs*	62.0	20.0	6.0	82	240.9	40.6	83.5	0.486	8.5	...	0.785	9.3	34.0	43.3	25.2	7.4	5.1	14.4	18.2	105.4
1	Oklahoma City Thunder*	59.0	23.0	5.0	82	241.5	39.0	82.7	0.471	8.1	...	0.806	10.8	33.9	44.7	21.9	8.3	6.1	15.3	22.7	106.2
2	Los Angeles Clippers*	57.0	25.0	1.0	82	240.9	39.1	82.5	0.474	8.5	...	0.730	10.5	32.5	43.0	24.6	8.6	4.8	13.9	21.5	107.9
3	Indiana Pacers*	56.0	26.0	24.0	82	241.2	36.0	80.2	0.449	6.7	...	0.779	10.2	34.5	44.7	20.1	6.7	5.4	15.1	20.4	96.7
4	Houston Rockets*	54.0	28.0	2.0	82	241.8	38.0	80.5	0.472	9.5	...	0.712	11.2	34.1	45.3	21.4	7.6	5.6	16.1	20.4	107.7

Question 1:

Do NBA Players Shoot Better Than NCAA Players?

Do NBA Players Shoot Better Than NCAA Players?

Field Goal Comparision Between NCAA And NBA
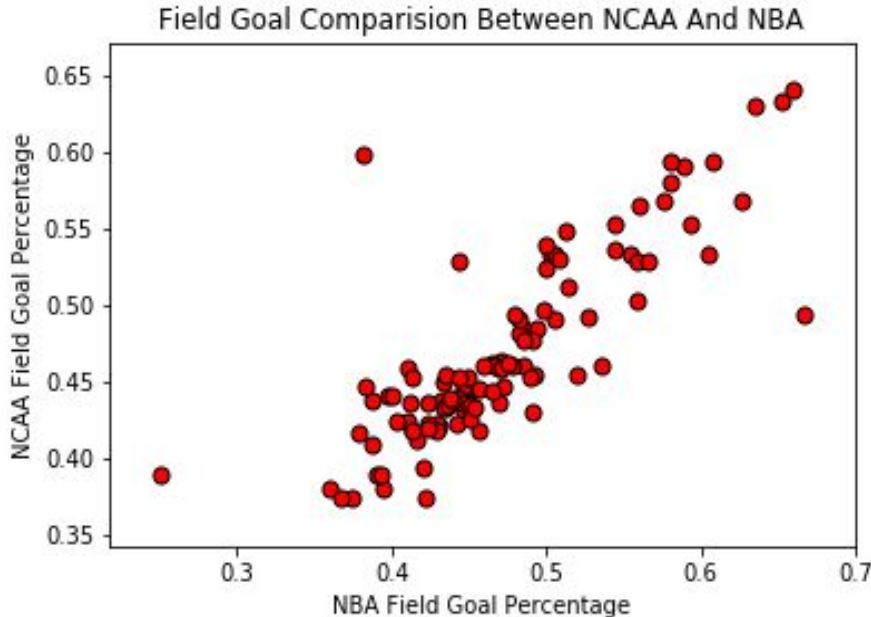
In general, NCAA players have a higher shooting percentage than NBA players. However, there are some players who shoot better in the NBA. This can be due to many factors such as the pace of game, coaching, and defensive strategies.
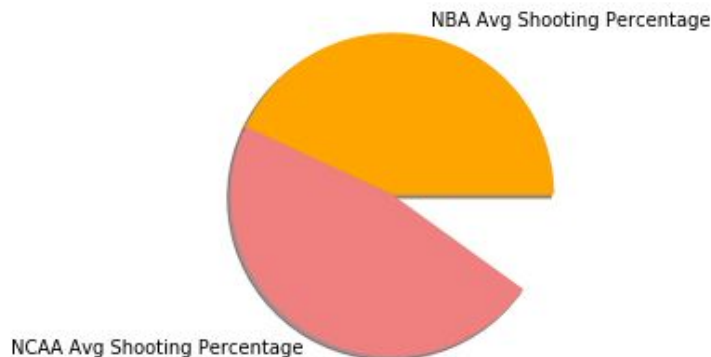
```
In [8]:  basketball_data=pd.merge(nba_13, ncaa_basketball,how="inner", on=["Player"])
         basketball_data=pd.merge(nba_14, ncaa_basketball, how="inner", on=["Player"])
         basketball_data=pd.merge(nba_15, ncaa_basketball, how="inner", on=["Player"])
         basketball_data=pd.merge(nba_16, ncaa_basketball, how="inner", on=["Player"])
         basketball_data=pd.merge(nba_17,ncaa_basketball, how="inner", on=["Player"])
         basketball_data.drop_duplicates(subset='Player',keep='first')
         basketball_data.dropna()
```

Player	Age	Tm	G	GS	MP	FG	FGA	FG%_x	...	steals	blocks	assists_turnover_ratio	personal_fouls	tech_fouls	flagrant_fouls	Year	FG%_y	FT%_y	Rd
Jarrett Allen	19	BRK	72	31	20.0	3.3	5.5	0.589	...	19.0	51.0	12.08	68.0	0.0	0.0	2017.0	0.590	0.732	1.0
Justin Anderson	24	PHI	38	0	13.7	2.3	5.3	0.431	...	30.0	42.0	57.90	88.0	1.0	0.0	2015.0	0.421	0.773	1.0
Kyle Anderson	24	SAS	74	67	26.7	3.1	5.9	0.527	...	83.0	35.0	135.76	122.0	0.0	0.0	2014.0	0.492	0.706	1.0
OG Anunoby	20	TOR	74	62	20.0	2.2	4.7	0.471	...	48.0	47.0	16.40	86.0	0.0	0.0	2017.0	0.463	0.598	1.0
Lonzo Ball	20	LAL	52	50	34.2	3.9	10.8	0.360	...	66.0	28.0	152.11	65.0	1.0	0.0	2017.0	0.380	0.437	1.0
Malik Beasley	21	DEN	62	0	9.4	1.2	2.9	0.410	...	32.0	6.0	32.50	75.0	1.0	0.0	2016.0	0.459	0.806	1.0
Jordan Bell	23	GSW	57	13	14.2	2.0	3.2	0.627	...	119.0	246.0	90.49	281.0	3.0	0.0	2017.0	0.568	0.642	2.0
Devin Booker	21	PHO	54	54	34.5	8.4	19.5	0.432	...	17.0	2.0	18.58	58.0	0.0	0.0	2015.0	0.437	0.854	1.0
Malcolm B...	25	MIL	48	20	29.9	5.1	10.5	0.485	...	103.0	28.0	159.26	212.0	1.0	0.0	2016.0	0.484	0.895	2.0

Do NBA Players Have a Higher Overall Shooting Average?

NBA players generally have a lower overall shooting average than NCAA players. One of the major reasons NBA players have a lower shooting average is the distance they have to shoot from is further.

NBA Avg Shooting Percentage

NCAA Avg Shooting Percentage

```python
avg_basketball_data=pd.merge(nba_13, ncaa_basketball,how="inner", on=["Player"])
avg_basketball_data=pd.merge(nba_14, ncaa_basketball, how="inner", on=["Player"])
avg_basketball_data=pd.merge(nba_15, ncaa_basketball, how="inner", on=["Player"])
avg_basketball_data=pd.merge(nba_16, ncaa_basketball, how="inner", on=["Player"])
avg_basketball_data=pd.merge(nba_17,ncaa_basketball, how="inner", on=["Player"])
nba_avg=avg_basketball_data['FG%_x'].mean()
ncaa_avg=avg_basketball_data['FG%_y'].mean()
avg_basketball_data.drop_duplicates(subset='Player',keep='first')
avg_basketball_data.dropna()
```

Question 2:

What offensive stats correlate with more wins for NCAA and NBA teams?

Data Analysis

```python
#Bot 5 averages
#drop unwanted rows
standings_17_18_bot5 = standings_17_18_combined.drop(standings_17_18_combined.index[:25]
                                                     ).drop(standings_17_18_combined.index[30])
standings_16_17_bot5 = standings_16_17_combined.drop(standings_16_17_combined.index[:25]
                                                     ).drop(standings_16_17_combined.index[30])
standings_15_16_bot5 = standings_15_16_combined.drop(standings_15_16_combined.index[:25]
                                                     ).drop(standings_15_16_combined.index[30])
standings_14_15_bot5 = standings_14_15_combined.drop(standings_14_15_combined.index[:25]
                                                     ).drop(standings_14_15_combined.index[30])
standings_13_14_bot5 = standings_13_14_combined.drop(standings_13_14_combined.index[:25]
                                                     ).drop(standings_13_14_combined.index[30])
#Pull columns needed
nba_17_18_x_bot5 = pd.DataFrame(standings_17_18_bot5[['3P', '3PA', '2P', '2PA']])
nba_16_17_x_bot5 = pd.DataFrame(standings_16_17_bot5[['3P', '3PA', '2P', '2PA']])
nba_15_16_x_bot5 = pd.DataFrame(standings_15_16_bot5[['3P', '3PA', '2P', '2PA']])
nba_14_15_x_bot5 = pd.DataFrame(standings_14_15_bot5[['3P', '3PA', '2P', '2PA']])
nba_13_14_x_bot5 = pd.DataFrame(standings_13_14_bot5[['3P', '3PA', '2P', '2PA']])
#Add year to correct data
nba_13_14_x_bot5['Year'] = '2014'
nba_14_15_x_bot5['Year'] = '2015'
nba_15_16_x_bot5['Year'] = '2016'
nba_16_17_x_bot5['Year'] = '2017'
nba_17_18_x_bot5['Year'] = '2018'
#calculate the avg
nba_13_14_x_bot5 = nba_13_14_x_bot5.groupby(['Year'])
nba_13_14_x_bot5 = (nba_13_14_x_bot5.sum()/5)
nba_14_15_x_bot5 = nba_14_15_x_bot5.groupby(['Year'])
nba_14_15_x_bot5 = nba_14_15_x_bot5.sum()/5
nba_15_16_x_bot5 = nba_15_16_x_bot5.groupby(['Year'])
nba_15_16_x_bot5 = nba_15_16_x_bot5.sum()/5
nba_16_17_x_bot5 = nba_16_17_x_bot5.groupby(['Year'])
nba_16_17_x_bot5 = nba_16_17_x_bot5.sum()/5
nba_17_18_x_bot5 = nba_17_18_x_bot5.groupby(['Year'])
nba_17_18_x_bot5 = nba_17_18_x_bot5.sum()/5
```

```python
combined_2and3tot_bot5 = pd.merge(nba_13_14_x_bot5, nba_14_15_x_bot5, how='outer',
                                  on=['Year', '3P', '3PA', '2P', '2PA'])
combined_2and3tot_bot5 = pd.merge(combined_2and3tot_bot5, nba_15_16_x_bot5, how='outer',
                                  on=['Year', '3P', '3PA', '2P', '2PA'])
combined_2and3tot_bot5 = pd.merge(combined_2and3tot_bot5, nba_16_17_x_bot5, how='outer',
                                  on=['Year', '3P', '3PA', '2P', '2PA'])
combined_2and3tot_bot5 = pd.merge(combined_2and3tot_bot5, nba_17_18_x_bot5, how='outer',
                                  on=['Year', '3P', '3PA', '2P', '2PA'])
combined_2and3tot_bot5
```

```python
plt.figure(figsize=(12,9))
plt.errorbar(combined_2and3tot_top5.index, combined_2and3tot_top5['3PA'],
             marker='p', linestyle='--', label='3PA Top 5', color='green')
plt.errorbar(combined_2and3tot.index, combined_2and3tot['3PA'],
             marker='o', linestyle='--', label='3PA League Avg')
plt.errorbar(combined_2and3tot_bot5.index, combined_2and3tot_bot5['3PA'],
             marker='v', linestyle='--', label='3PA Bot 5', color='red')

plt.grid()
plt.legend(loc='best', fancybox=True)
plt.title("3PT Shots Over the Years (NBA)")
plt.xlabel("Year")
plt.ylabel("3PT Shots Attempted (per game)")
plt.savefig("Charts/3pt_nba")
# Save the Figure
plt.show()
```

Year	2P	2PA	3P	3PA
2014	18.164612	37.152041	6.302776	18.228471
2015	17.177578	35.726903	6.400627	18.575993
2016	18.356929	37.438970	7.106534	20.477794
2017	18.288412	36.846901	7.420760	21.113553
2018	18.351471	36.461522	7.678220	21.832399

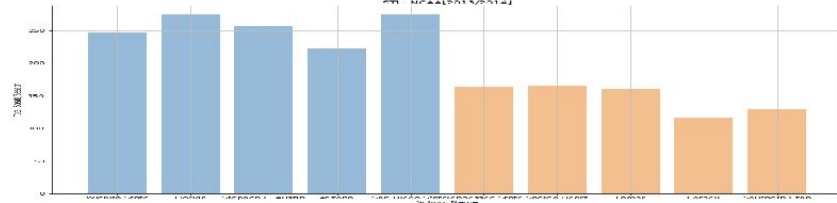# Offensive Stat: 3PT
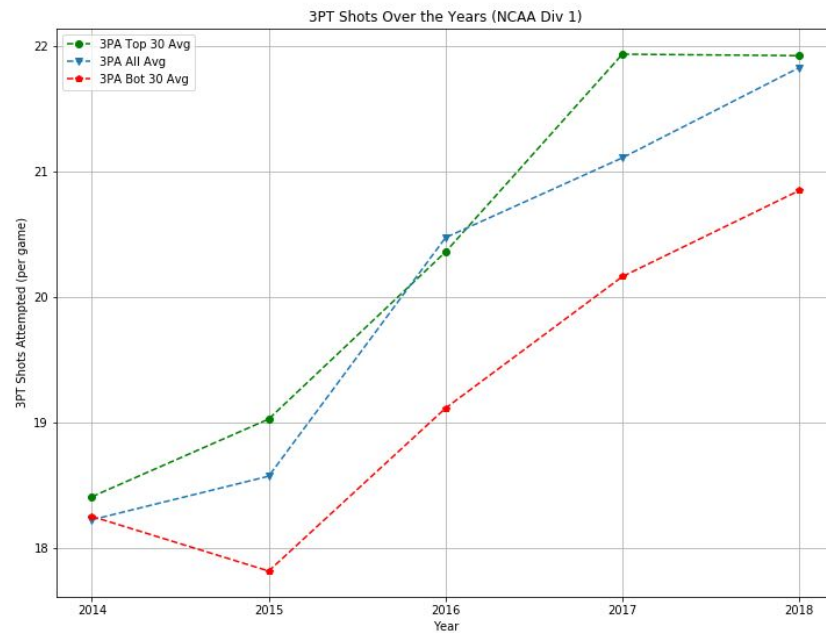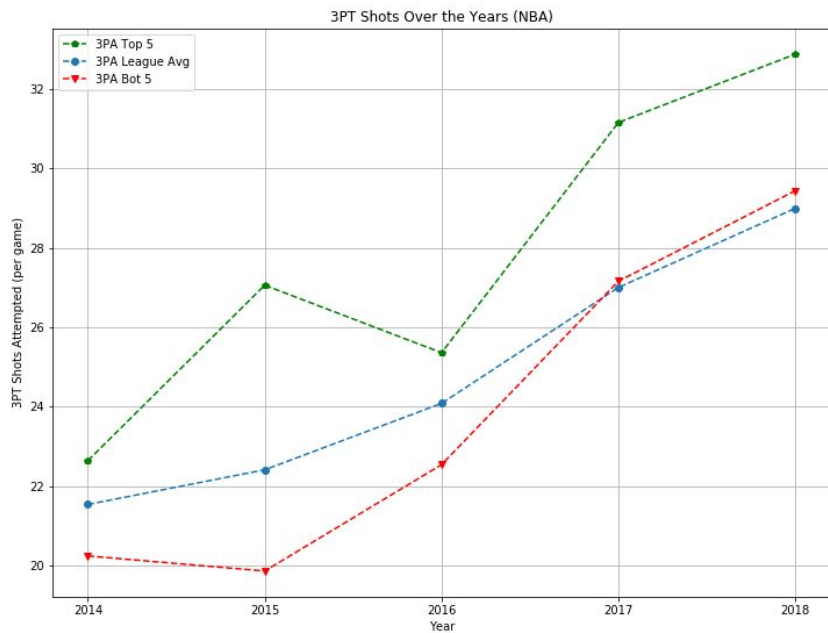
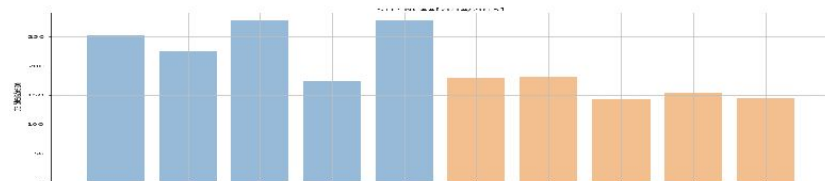# Offensive Stat: 2PT
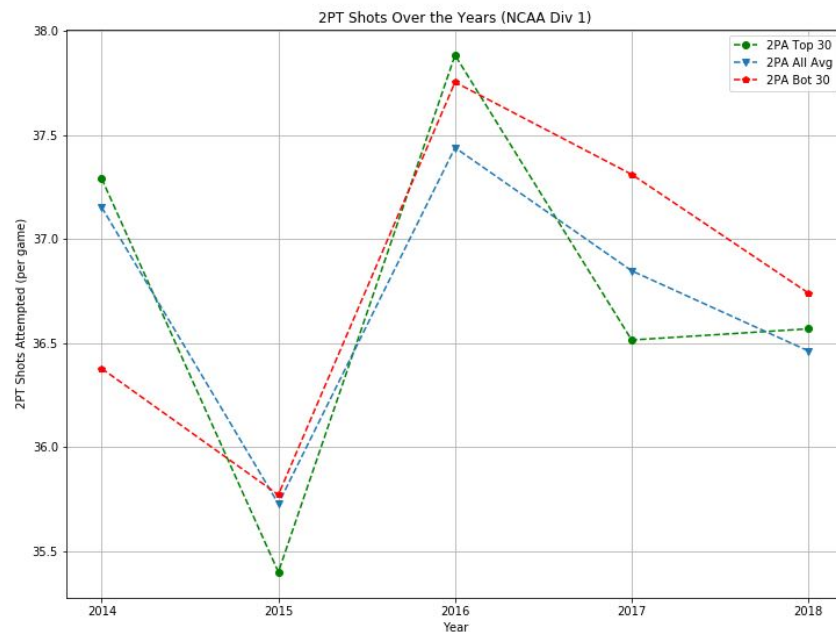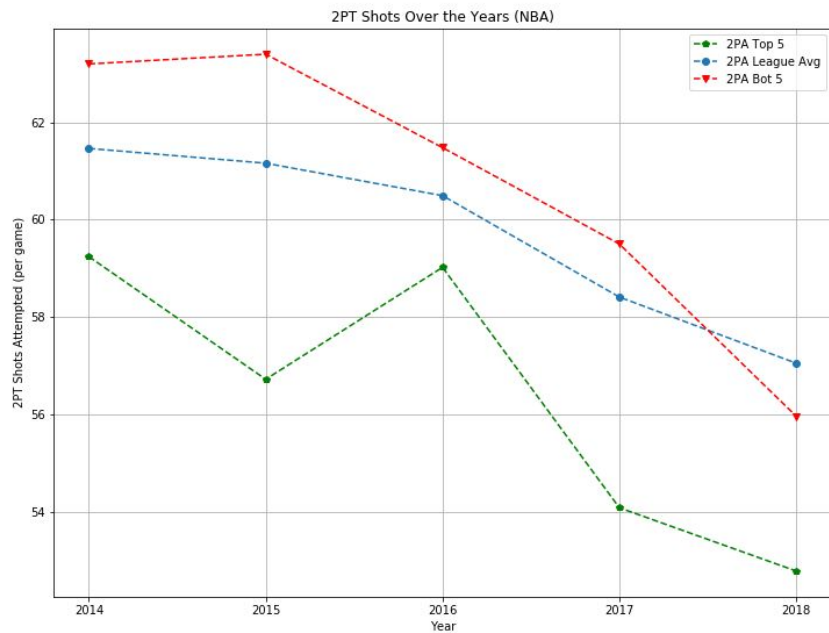
Question 3:

What defensive stats correlate with more wins for NCAA and NBA teams?

Is there a difference in the way that steals, blocks, rebounds affect the success of a team between college and professional basketball?
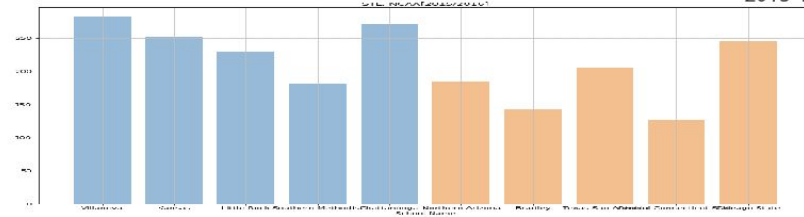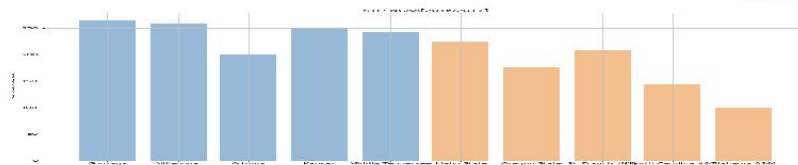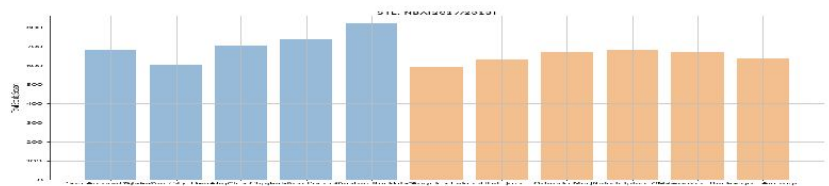
STEALS COMPARISON
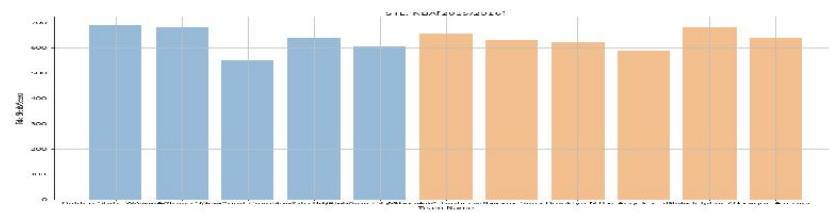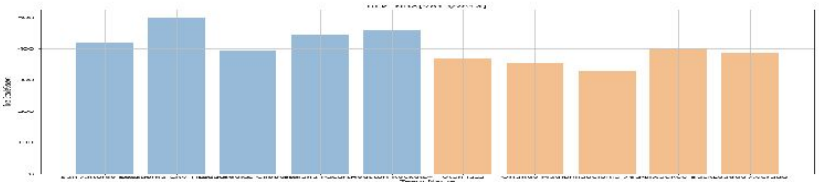
NCAA

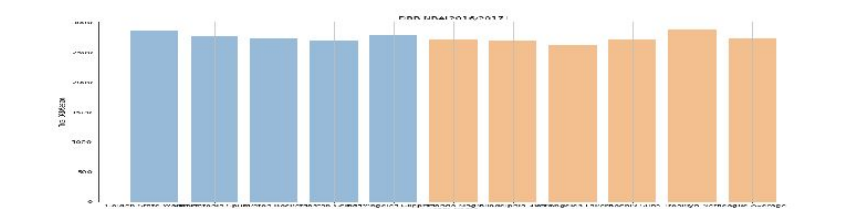NBA

2013-14
2014-15
2015-16
2016-17

BLOCKS COMPARISON

NCAA

NBA

2013-14

2014-15

2015-16

2016-17
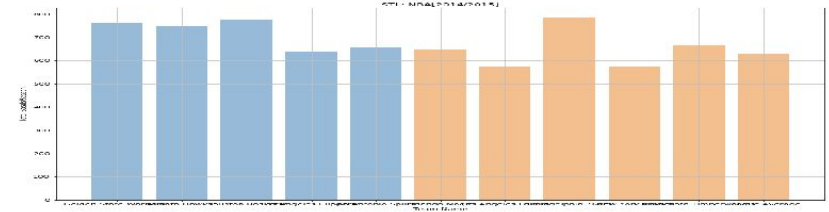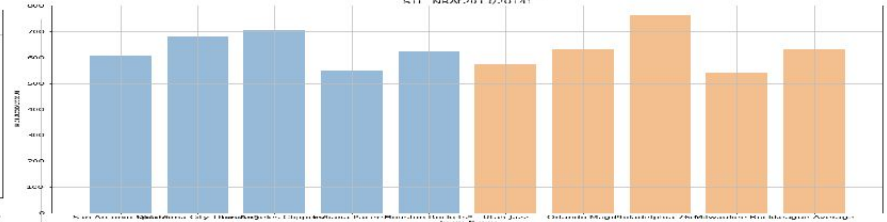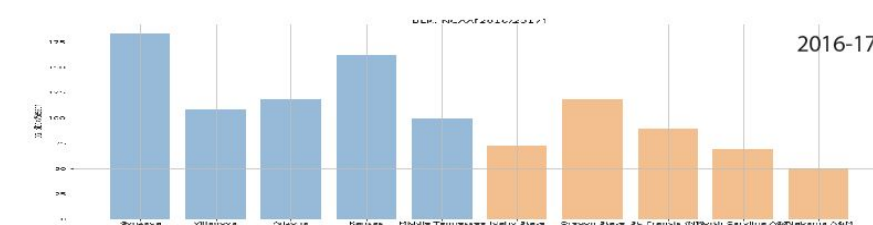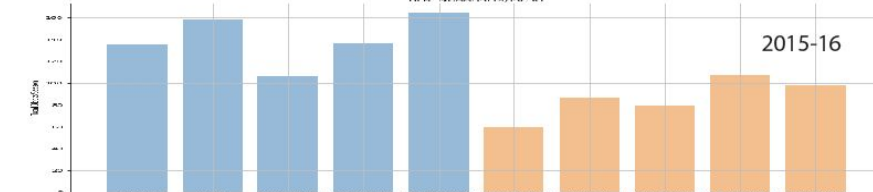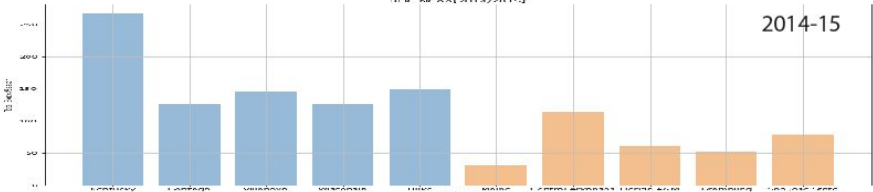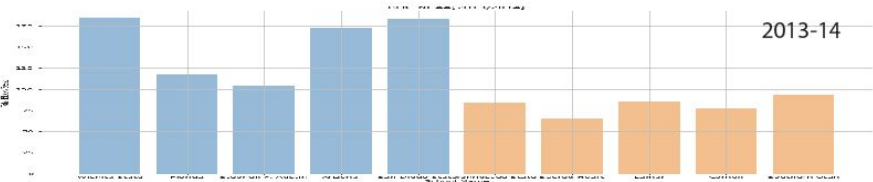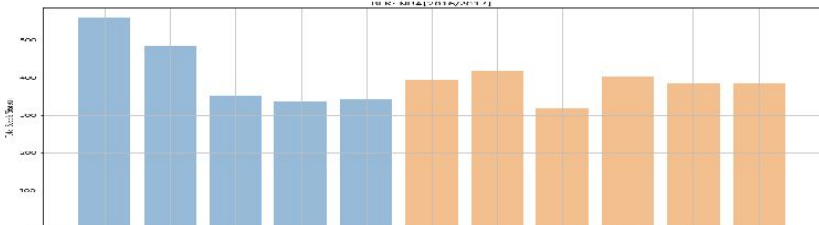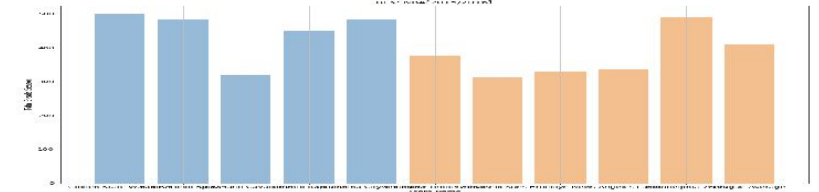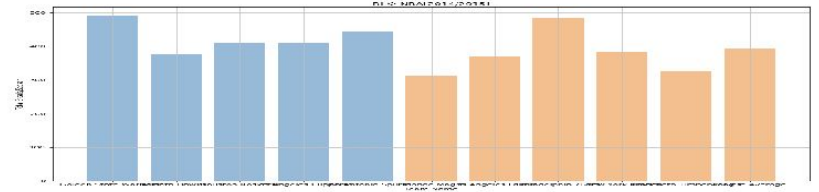
DEFENSIVE REBOUND COMPARISON

NCAA

NBA

2013-14

2014-15

2015-16

2016-17
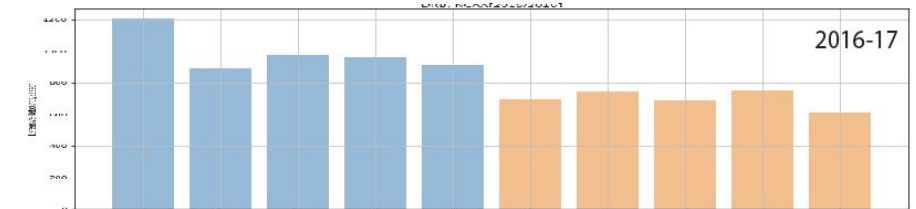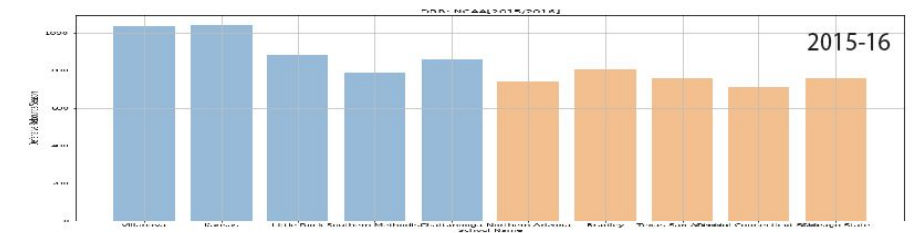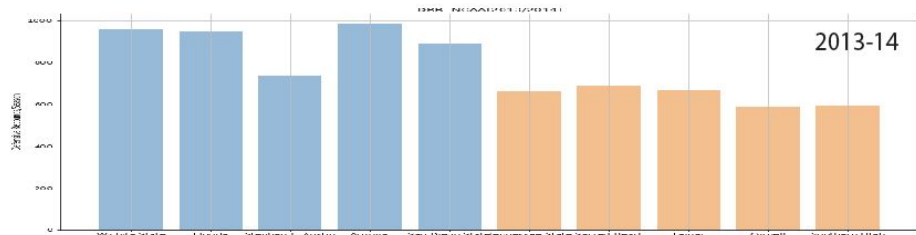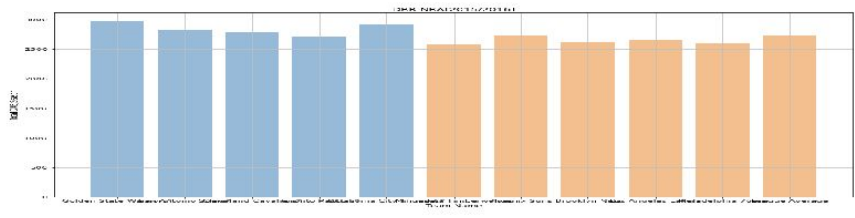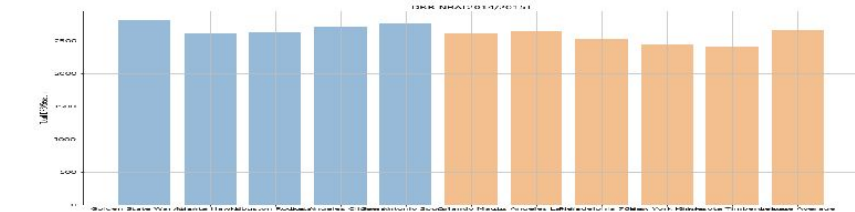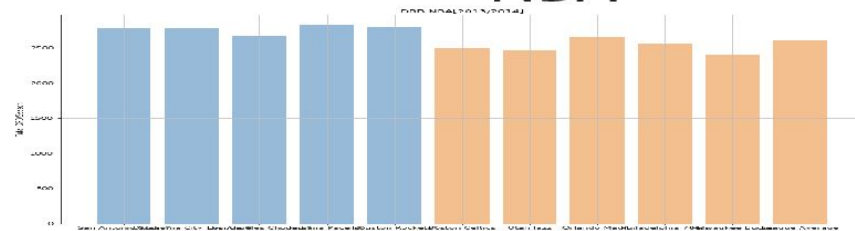
Question 4:

Do teams with superstars perform better than teams with only role players (NCAA and NBA)?

Meet the Metrics

```python
#player_contribution
working_boxscores = working_boxscores.assign(
    player_contribution = working_boxscores['playPTS']/working_boxscores['teamPTS']*100)
working_boxscores.head()
```

	game_id	teamAbbr	teamRslt	teamAST	playDispNm	playPTS	teamPTS	player_contribution
0	0	WAS	Loss	26	A.J. Price	7	84	8.333333
1	0	WAS	Loss	26	Trevor Ariza	9	84	10.714286
2	0	WAS	Loss	26	Emeka Okafor	10	84	11.904762
3	0	WAS	Loss	26	Bradley Beal	8	84	9.523810
4	0	WAS	Loss	26	Trevor Booker	4	84	4.761905

```python
#star_power = top1/avg(top5-top1)
star_calcs = []
for game in working_boxscores['game_id']:
    top_five = working_boxscores.loc[working_boxscores.game_id == game].nlargest(5, 'player_contribution')
    ['player_contribution'].tolist()
    star_calcs.append(top_five[0]/(np.mean(top_five[1:5])))
working_boxscores = working_boxscores.assign(star_power = star_calcs)
```
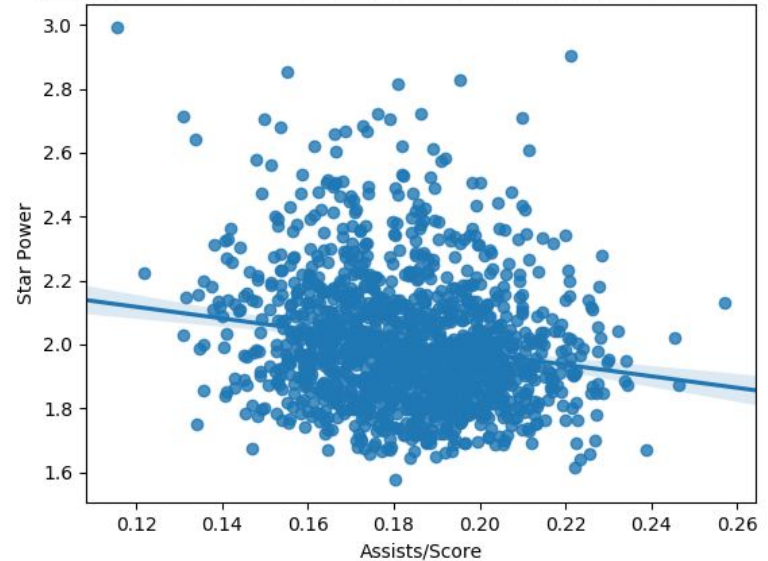
Assists/Score vs Star Power

Assists/Score vs Star Power by Team per Season (NBA: 2013-2017)
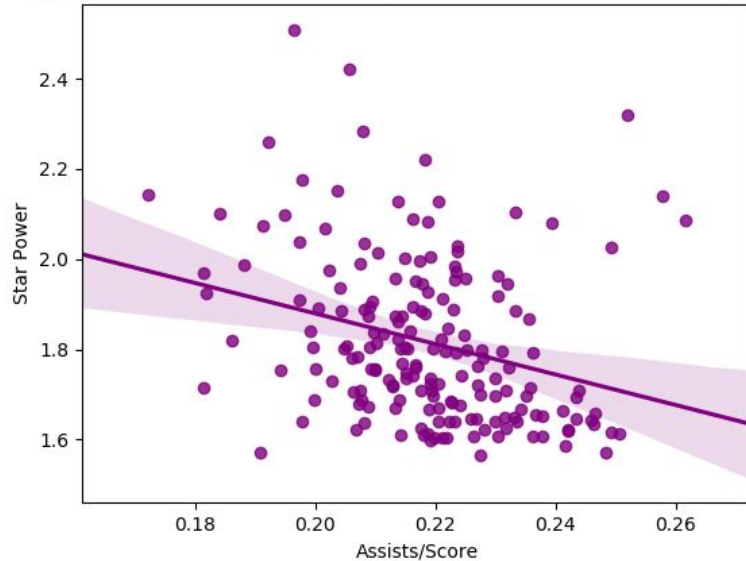
Star Power = 3.0 + -3.0 * Assists/Score
r = -0.29
p value = 7.25e-05
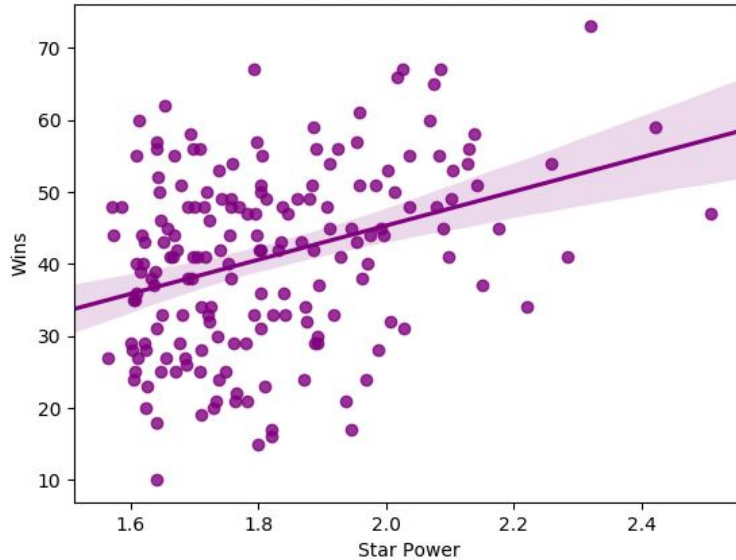

Assists/Score vs Star Power by Team per Season (NCAA: 2013-2017)

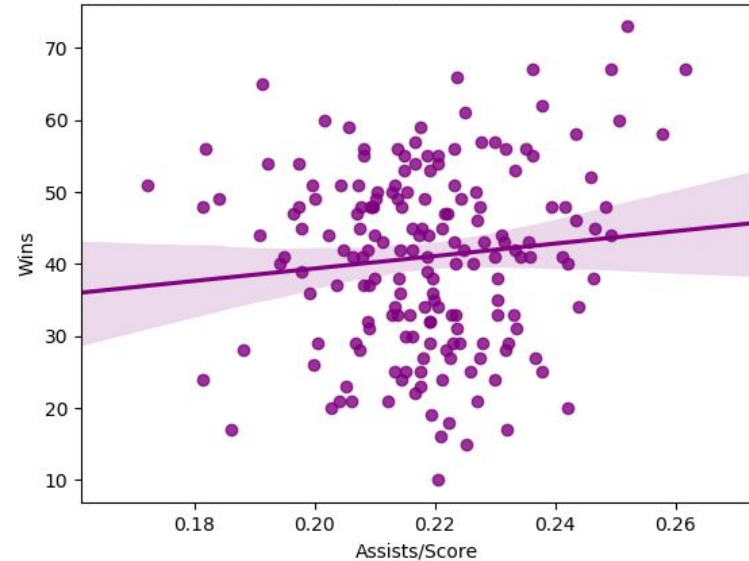Star Power = 2.0 + -2.0 * Assists/Score
r = -0.18
p value = 2.75e-11

NBA Team Performance

Star Power vs Wins by Team per Season (NBA: 2013-2017)

Wins = -2.0 + 24.0 * Star Power
r = 0.34
p value = 2.81e-06
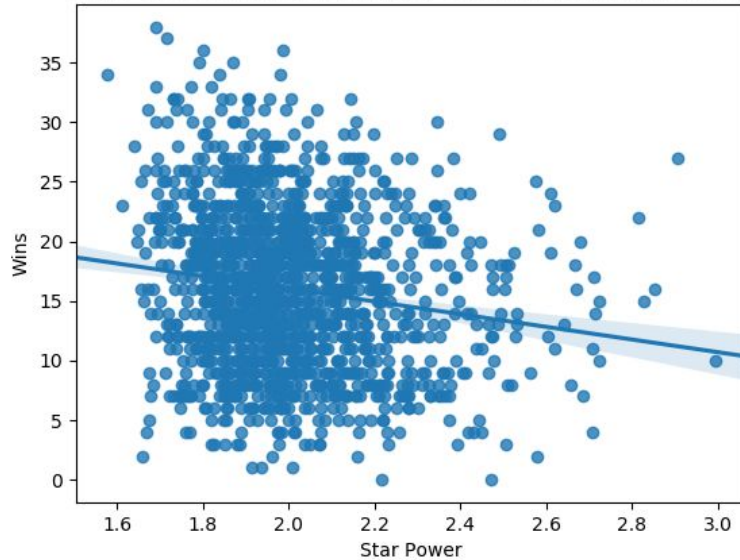
Assists/Score vs Wins by Team per Season (NBA: 2013-2017)
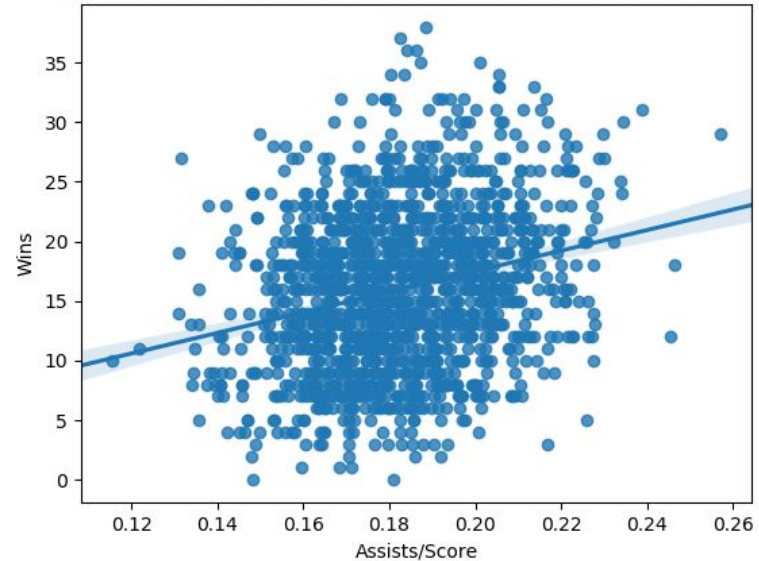
Wins = 22.0 + 87.0 * Assists/Score
r = 0.11
p value = 0.15

NCAA Team Performance

Star Power vs Wins by Team per Season (NCAA: 2013-2017)

Wins = 27.0 + -5.0 * Star Power
r = -0.16
p value = 1.96e-09

Assists/Score vs Wins by Team per Season (NCAA: 2013-2017)

Wins = 0.0 + 86.0 * Assists/Score
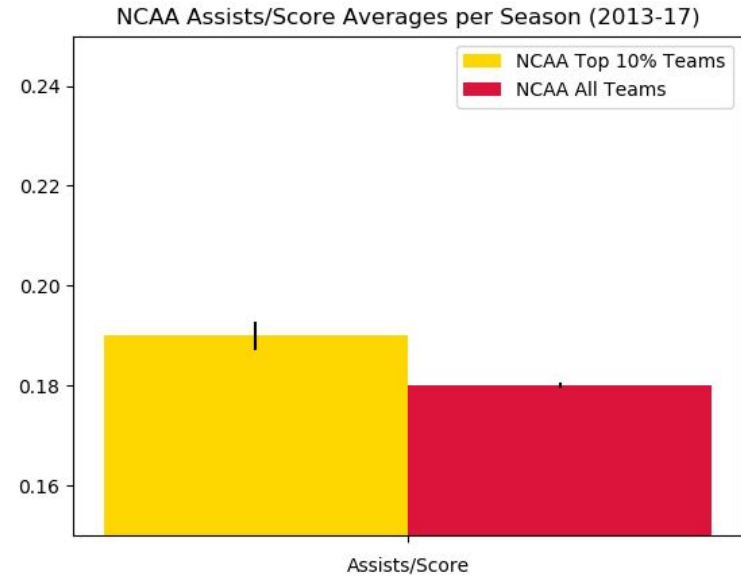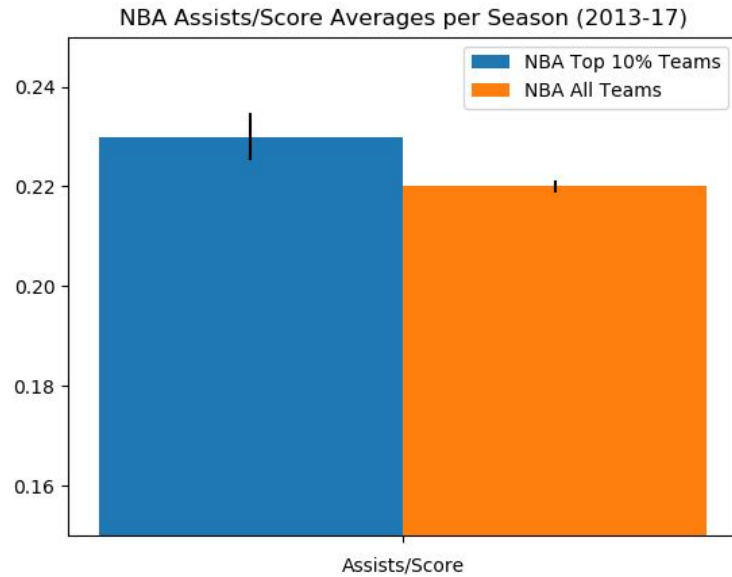r = 0.25
p value = 8.81e-22

Top 10%

```python
#avg team_performance by year for top 10 teams
season_avg = []
for season in final_df['season'].unique():
    season_avg.append(final_df.loc[final_df.season == season].nlargest(3, 'win_count'))
top_ten_per_season = pd.concat(season_avg)
top_ten_per_season.head()
```

	season	teamAbbr	star_power	assists/score	win_count	w/l_ratio
15	2012	MIA	2.016983	0.223727	66.0	0.804878
20	2012	OKC	2.068063	0.201693	60.0	0.731707
25	2012	SA	1.693733	0.243450	58.0	0.707317
55	2013	SA	1.653307	0.237808	62.0	0.756098
50	2013	OKC	2.421223	0.205716	59.0	0.719512

p value = 3.07e-4, 2.5e-6

NBA Assists/Score Averages per Season (2013-17)

- NBA Top 10% Teams
- NBA All Teams

NCAA Assists/Score Averages per Season (2013-17)

- NCAA Top 10% Teams
- NCAA All Teams

Assists/Score

p value = 6.78e-3, 5.5e-8