

Home Court Advantage In the NBA



Questions to Explore

1. How big of an effect is Home Court Advantage (referred to as HCA from here on)?
2. Any noticeable trends? Is it getting stronger or weaker?
3. What game statistics are most/least affected by HCA?
4. Why might those statistics be affected by HCA?
5. Does attendance, average travel distance, or elevation have an effect?
6. If an NBA team wanted to maximize HCA and/or minimize Away Court Disadvantage what strategies might they employ?

Initial Hypotheses

1. HCA is a well known phenomenon, so I imagine it will be noticeable and have a fairly significant effect.
2. Perhaps HCA is declining, because teams may be trying to account for it, travel to away games sooner to acclimate, nicer planes for more comfortable travel.
3. Some statistics it may affect more than others: perhaps the number of fouls, people often complain about 'home cooking', perceived bad calls in favor of the home team.
4. Attendance may have an effect, louder fuller stadiums perhaps give the home team a bigger boost. Average travel distance could certainly have an effect on a travelling team. Elevation is certainly known to have effects on performance so it too would likely have an effect

Data Sources

ESPN Standings (Home and Away records): 2002-2019

ESPN Attendance Stats: 2001-2019

ESPN Pace Statistics: 2002-2019

NBA - Game Statistics: 1996 - 2019

How I got the Data

Scraping ESPN was fairly straightforward using Pandas read HTML

Once I had a pandas dataframe, I had some minimal cleaning to do, occasionally needing to recast data types

Then I created any extra columns I needed (e.g. Home Win Percentage, Away Win Percentage, Season)

#ESPN Standings data retrieval

```
def getWinPctDiff(year, AllCols = False):
```

```
    if year == 2019:
```

```
        HTML = pd.read_html('https://www.espn.com/nba/standings/_/group/league')
```

```
    else:
```

```
        HTML = pd.read_html(f'https://www.espn.com/nba/standings/_/season/{year}/group/league')
```

```
    if year <=2015:
```

```
        result = pd.DataFrame(HTML[0]['W'][2:32]).reset_index(drop = True)
```

```
        result['0'] = [x.strip('z --y --x --') for x in result['W']]
```

```
        result = pd.DataFrame(result['0']).join(HTML[3])
```

```
    else:
```

```
        result = pd.DataFrame(HTML[0]['W'][0].split('--')[1:31])
```

```
        result = result.join(HTML[3])
```

```
    result['HOME_LIST'] = result['HOME'].str.split('-')
```

```
    result.loc[:, 'HOME_WIN_PCT'] = result.HOME_LIST.map(lambda x:int(x[0])/(int(x[0])+int(x[1])))
```

```
    result['AWAY_LIST'] = result['AWAY'].str.split('-')
```

```
    result.loc[:, 'AWAY_WIN_PCT'] = result.AWAY_LIST.map(lambda x:int(x[0])/(int(x[0])+int(x[1])))
```

```
    result['win_pct_diff'] = result['HOME_WIN_PCT']-result['AWAY_WIN_PCT']
```

```
    result = result.rename(columns={0:'Team', '0':'Team'})
```

```
    result['year'] = year
```

```
    if AllCols:
```

```
        return result
```

```
    else:
```

```
        return result[['Team', 'win_pct_diff', 'year']]
```

```
def GetAllWinPctDiffSince(FirstYear):
```

```
    seasons = []
```

```
    for year in range(FirstYear, 2020):
```

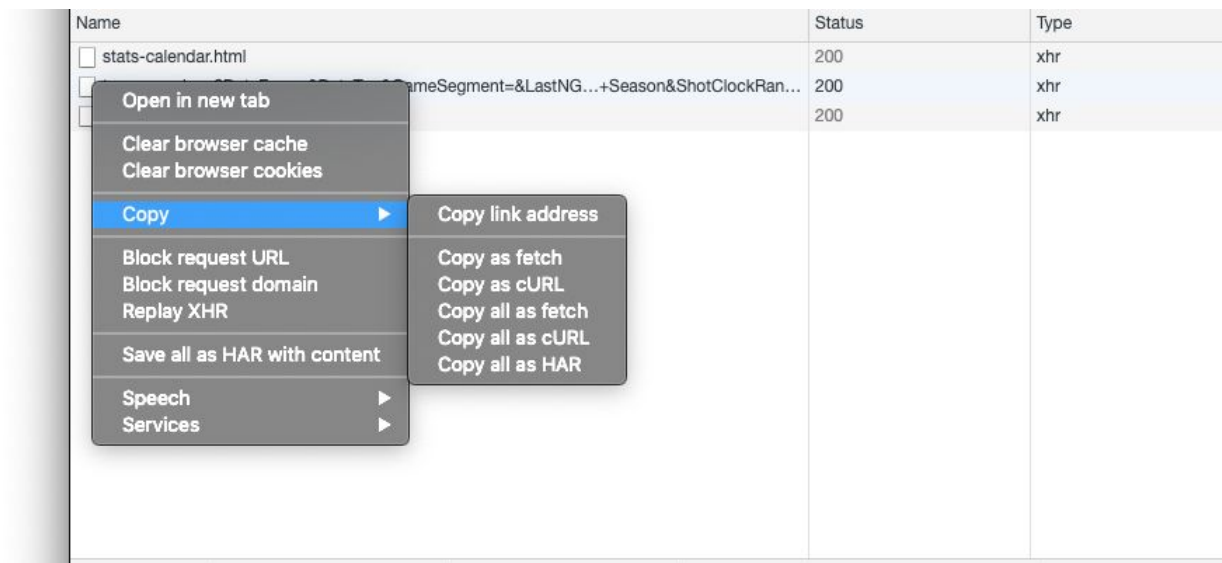
```
        seasons.append(getWinPctDiff(year))
```

```
    return pd.concat(seasons)
```

NBA data was a little trickier

The tables with the data were done using javascript and as such the data was not present in the site's html

So instead I inspected the page and selected XHR under the network tab and looked for a listing that might be the table. Then chose to copy as cURL



Translate Curl to Python

Convert curl syntax to Python, Ansible URI, Node.js, R, PHP, Strest, Go, Dart, JSON, Rust

GitHub



2,253

curl command

```
curl 'https://stats.nba.com/stats/teamgamelogs?
DateFrom=&DateTo=&GameSegment=&LastNGames=0&LeagueID=00&Location=&Meas
ureType=Base&Month=0&OpponentTeamID=0&Outcome=&PORound=0&PaceAdj
ust=N&PerMode=Totals&Period=0&PlusMinus=N&Rank=N&Season=2018-
19&SeasonSegment=&SeasonType=Regular+Season&ShotClockRange=&VsConfere
nce=&VsDivision=' -H 'Sec-Fetch-Mode: cors' -H 'X-NewRelic-ID:
VQECWF5UChAHUINtBwgBVw==' -H 'Accept-Encoding: gzip, deflate, br' -H
'Accept-Language: en-US,en;q=0.9' -H 'User-Agent: Mozilla/5.0 (Macintosh; Intel
Mac OS X 10_14_6) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/76.0.3809.132 Safari/537.36' -H 'Sec-Fetch-Site: same-origin' -H 'Accept:
```

Examples: [GET](#) - [POST](#) - [Basic Auth](#)

Python requests

```
'Accept': 'application/json, text/plain, */',
'x-nba-stats-token': 'true',
'Referer': 'https://stats.nba.com/teams/boxscores-traditional/?Season=2018-
19&SeasonType=Regular%20Season',
'Connection': 'keep-alive',
'x-nba-stats-origin': 'stats',
}

params = (
    ('DateFrom', ''),
```

Language Python




```

def GetSeason(year, less=True):
    import requests

    cookies = {

    }

    headers = {

    }

    params = (

    )

    response = requests.get('https://stats.nba.com/stats/teamgamelogs', headers=headers, params=params, cookies=cookies)

    df = pd.DataFrame(response.json().get('resultSets')[0]['rowSet'])

    df.columns = response.json().get('resultSets')[0]['headers']

    df['Home'] = [x.count('vs') for x in df['MATCHUP']]

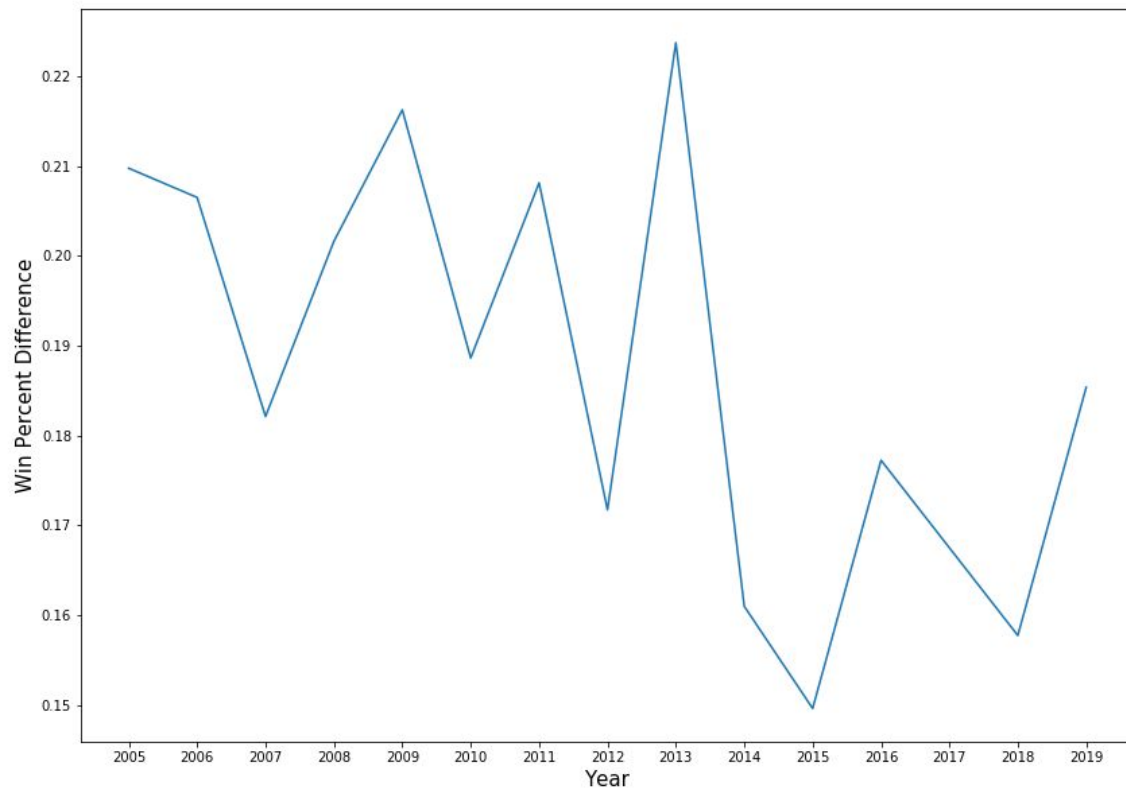
    df_less = df[['SEASON_YEAR', 'TEAM_ID', 'TEAM_ABBREVIATION', 'TEAM_NAME', 'GAME_ID',
                  'GAME_DATE', 'MATCHUP', 'WL', 'MIN', 'FGM', 'FGA', 'FG_PCT', 'FG3M',
                  'FG3A', 'FG3_PCT', 'FTM', 'FTA', 'FT_PCT', 'OREB', 'DREB', 'REB', 'AST',
                  'TOV', 'STL', 'BLK', 'BLKA', 'PF', 'PFD', 'PTS', 'PLUS_MINUS', 'Home']]
    if less:
        return df_less
    else:
        return df

```

Findings

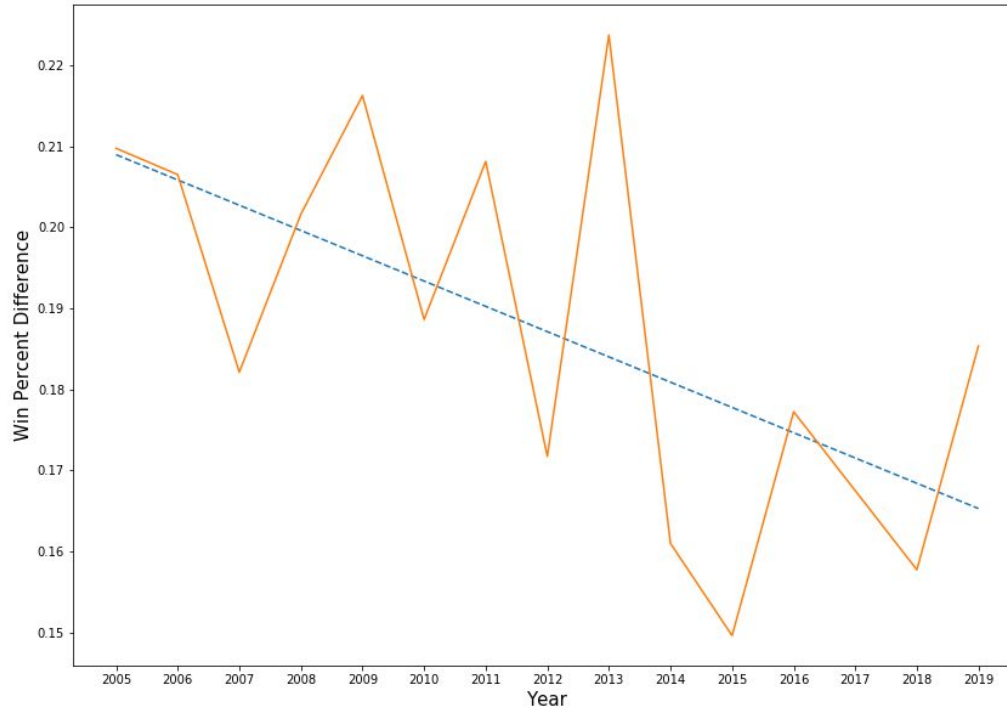


Avg Win Percent Difference (Home-Away) per year in the NBA



HCA seems to be declining

Avg Win Percent Difference (Home-Away) per year in the NBA



What statistics were affected by HCA this season?

```
CalculateZScores2(df2018, .05)
```

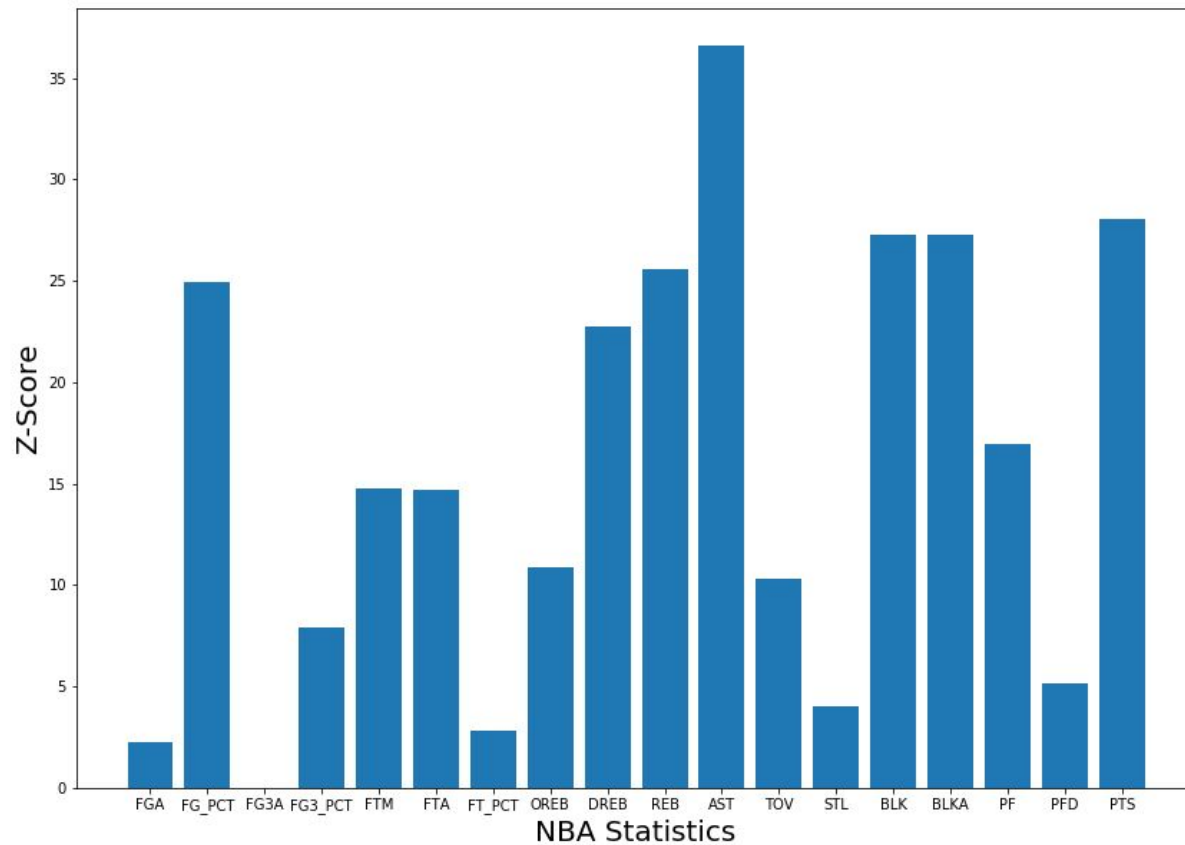
```
['FGA z test: 0.10995907168996052 p-value: 0.912441834082345 significant for FGA: False',  
'FG_PCT z test: 4.893758425957782 p-value: 9.892826163739013e-07 significant for FG_PCT: True',  
'FG3A z test: 0.7060086524221787 p-value: 0.4801827567062702 significant for FG3A: False',  
'FG3_PCT z test: 2.25703835398375 p-value: 0.02400567829222485 significant for FG3_PCT: True',  
'FTM z test: 2.167691159673808 p-value: 0.030182192540978876 significant for FTM: True',  
'FTA z test: 2.16187565730358 p-value: 0.030627761797353346 significant for FTA: True',  
'FT_PCT z test: -0.1526454230070415 p-value: 0.8786779023433369 significant for FT_PCT: False',  
'OREB z test: 1.7386946924478939 p-value: 0.0820884801231451 significant for OREB: False',  
'DREB z test: 4.728845873706575 p-value: 2.257997067441542e-06 significant for DREB: True',  
'REB z test: 4.845146326431501 p-value: 1.2651839865862716e-06 significant for REB: True',  
'AST z test: 5.374359289378431 p-value: 7.685547339886798e-08 significant for AST: True',  
'TOV z test: -0.48252518893351853 p-value: 0.6294329090854964 significant for TOV: False',  
'STL z test: -0.47186783559273343 p-value: 0.6370211275718917 significant for STL: False',  
'BLK z test: 2.395084819305112 p-value: 0.01661652087248278 significant for BLK: True',  
'BLKA z test: -2.395084819305112 p-value: 0.01661652087248278 significant for BLKA: True',  
'PF z test: -2.208926399600807 p-value: 0.027179760625472203 significant for PF: True',  
'PFD z test: 2.2134596621892078 p-value: 0.026865971292953744 significant for PFD: True',  
'PTS z test: 5.370483086720261 p-value: 7.852599834439992e-08 significant for PTS: True']
```


How about over many seasons?

```
CalculateZScores2(AllSeasons, .05)
```

```
['FGA z test: 2.2705712698945653 p-value: 0.02317294432698745 significant for FGA: True',  
'FG_PCT z test: 24.908081815540086 p-value: 6.081788501836621e-137 significant for FG_PCT: True',  
'FG3A z test: 0.021834310119632508 p-value: 0.9825801251875218 significant for FG3A: False',  
'FG3_PCT z test: 7.884219428669565 p-value: 3.1650769201153965e-15 significant for FG3_PCT: True',  
'FTM z test: 14.769618076393762 p-value: 2.30012647986333e-49 significant for FTM: True',  
'FTA z test: 14.717412921186614 p-value: 4.9836405897377915e-49 significant for FTA: True',  
'FT_PCT z test: 2.8011488926404793 p-value: 0.005092101879859729 significant for FT_PCT: True',  
'OREB z test: 10.861967629406294 p-value: 1.7493861513887586e-27 significant for OREB: True',  
'DREB z test: 22.727266288940676 p-value: 2.4087325193316984e-114 significant for DREB: True',  
'REB z test: 25.560921286820232 p-value: 4.1517258472596554e-144 significant for REB: True',  
'AST z test: 36.57545366095745 p-value: 7.025962155611744e-293 significant for AST: True',  
'TOV z test: -10.32016473028055 p-value: 5.712506386462209e-25 significant for TOV: True',  
'STL z test: 4.055540173967259 p-value: 5.001857420712703e-05 significant for STL: True',  
'BLK z test: 27.296538717209557 p-value: 4.66250750558592e-164 significant for BLK: True',  
'BLKA z test: -27.296538717209557 p-value: 4.66250750558592e-164 significant for BLKA: True',  
'PF z test: -16.983335266722257 p-value: 1.0910561480493431e-64 significant for PF: True',  
'PFD z test: 5.145227940381808 p-value: 2.671961749207188e-07 significant for PFD: True',  
'PTS z test: 28.04683471520368 p-value: 4.365872017943844e-173 significant for PTS: True']
```

Z-Score Stat Comparisons



Takeaways

- Assists are most affected - interesting because they are one of the more subjective statistics to keep track of. Perhaps some statistical home cooking is going on
- Free throws attempted and Personal Fouls are modestly affected, could add a little evidence to the home cooking theory, but doesn't seem to be a big enough effect to be the main/only reason for HCA
- Personal Fouls drawn being only modestly affected might be evidence against the home cooking theory. You would think you might draw more personal fouls at home.
- Free Throw Percentage not really affected much, makes sense, not very demanding kind of play, even if players are fatigued from travel or elevation, also not something a referee can affect.

Exploring Why HCA is declining

Increase in 3 Point Field Goals Theory

There is a theory that the rise of three point field goal attempts in the NBA is leading to a decline in home court advantage.

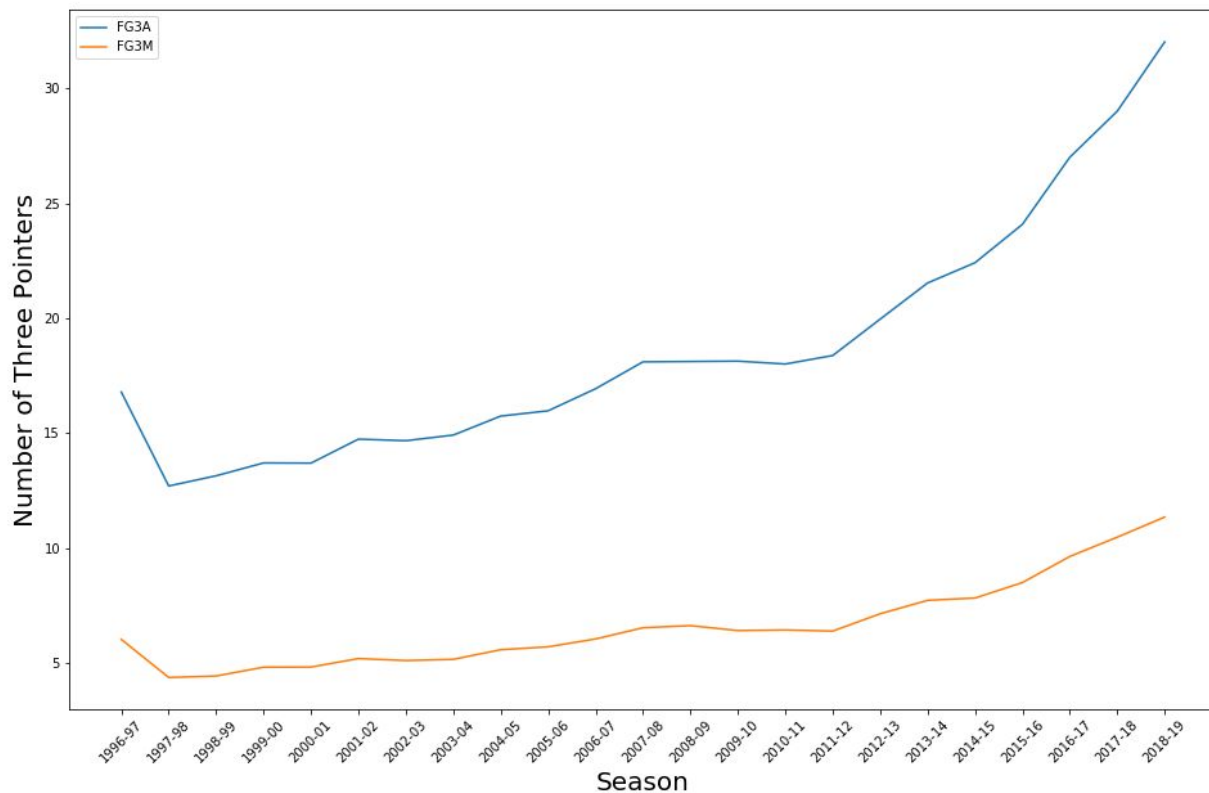
Could be several reasons for this - You're less likely to be fouled on a three point shot, so fewer chances for foul calls or home cooking to have an effect

A three point attempt is less physically demanding than most two point plays so it is perhaps not as affected by fatigue potentially caused by travel or a difference in elevation

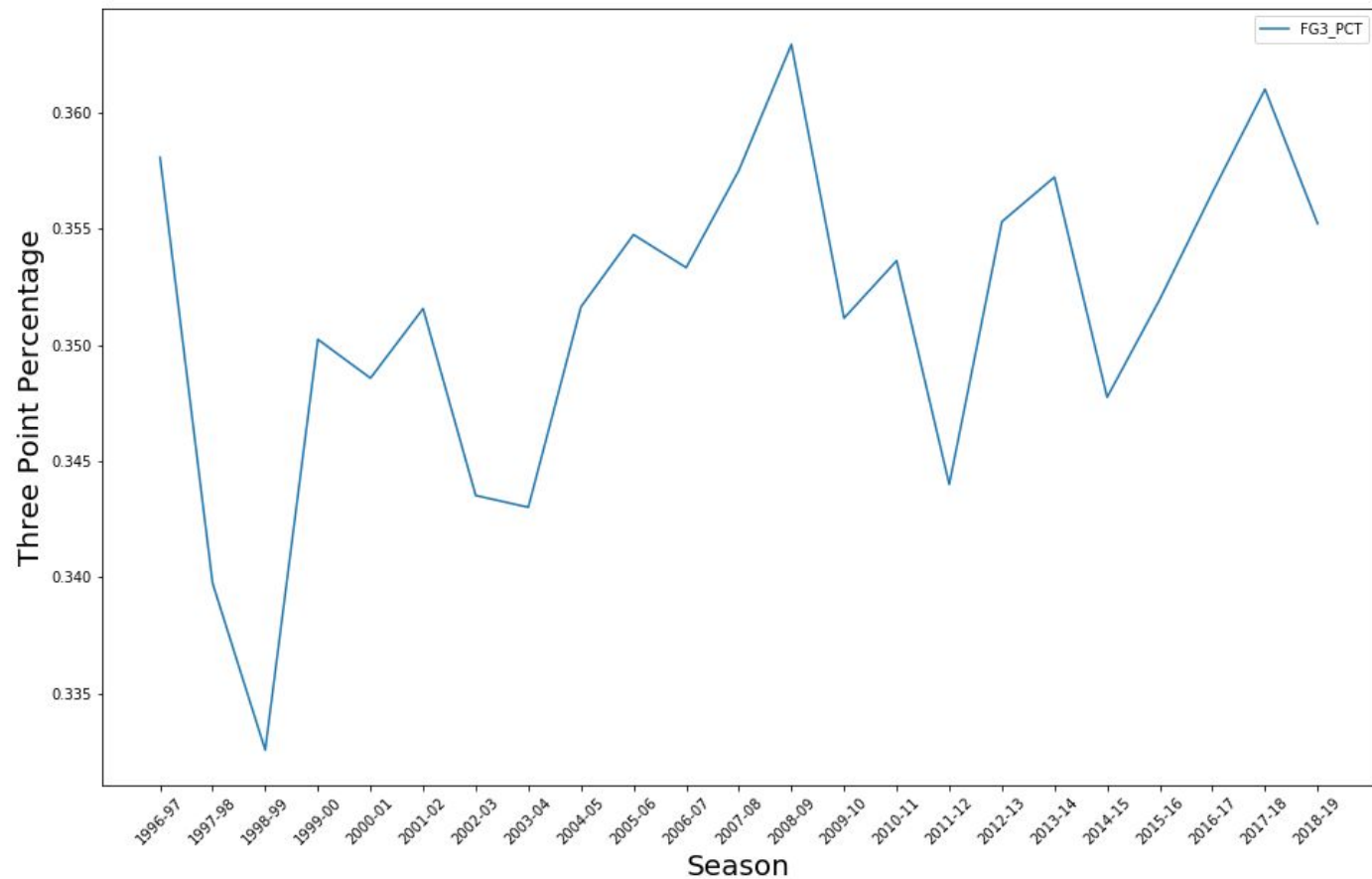
More three point shots might lead to a higher margin of victory thus negating the effects of an extra couple free throws (home cooking theory) or it puts the game out of reach before fatigue can really set in

Three Point Field Goals Attempted by Season

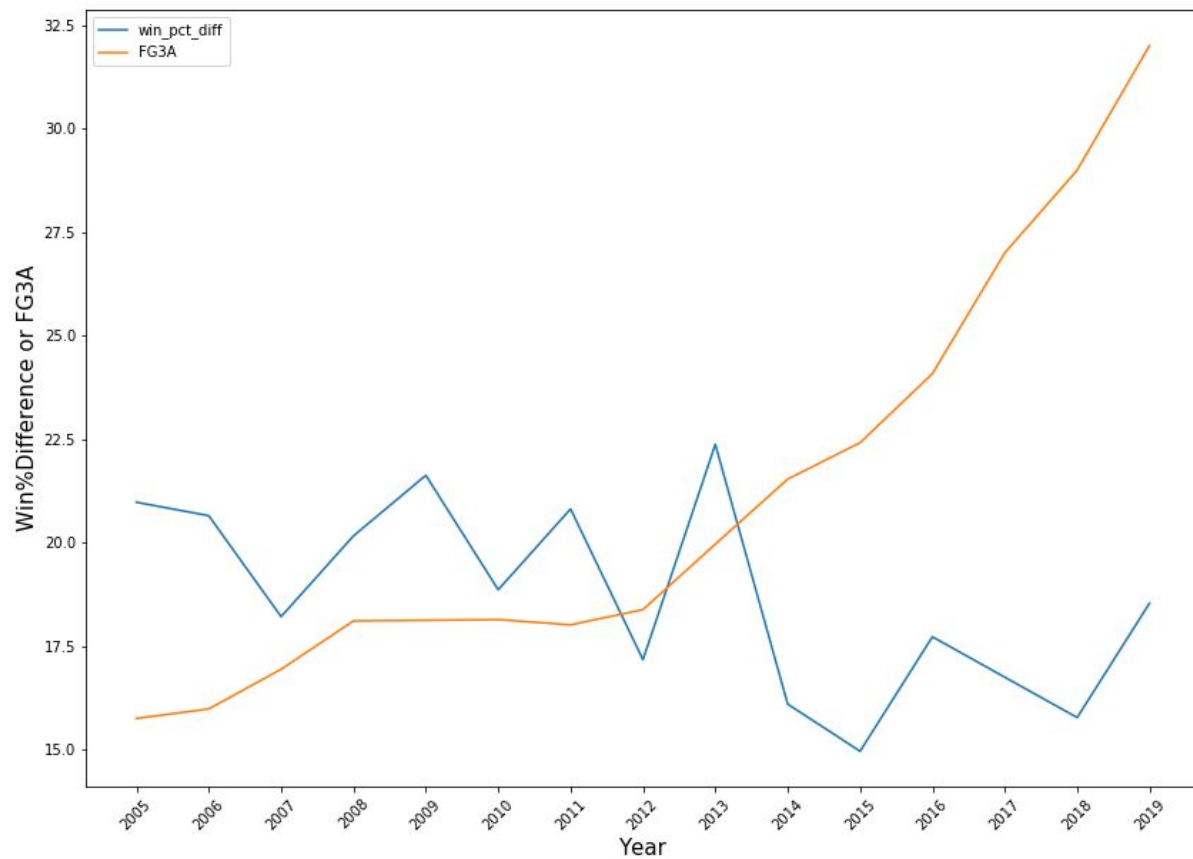
3-Point Field Goals Per NBA Season



3-Point Field Goal Percentage Per NBA Season



FG3A and HCA per season



Correlation

How strong is the correlation between FG3A increase and the decrease in HCA

```
import scipy.stats as stats
stats.pearsonr(AllWinPct2.groupby('year').mean()['win_pct_diff'], AvgStats['FG3A'][8:23])
(-0.5353039944249238, 0.03974744978548436)
```

-0.54 and the probability of an uncorrelated system producing datasets that have a Pearson correlation at least as extreme as this is only 4%, which would be outside of a 95% confidence interval

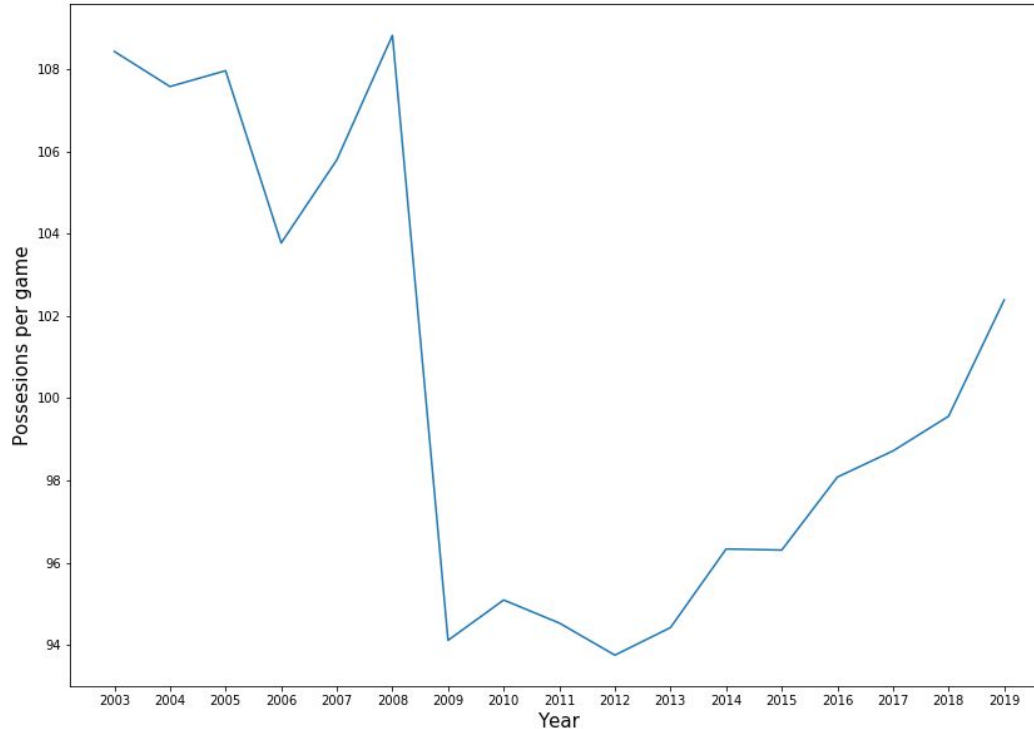
Correlation of all stats

```
for stat in ['FGM', 'FGA', 'FG_PCT', 'FG3M', 'FG3A',  
            'FG3_PCT', 'FTM', 'FTA', 'FT_PCT', 'OREB', 'DREB', 'REB', 'AST', 'TOV',  
            'STL', 'BLK', 'BLKA', 'PF', 'PFD', 'PTS', 'PLUS_MINUS']:  
    correlation, pvalue = stats.pearsonr(AllWinPct2.groupby('year').mean()['win_pct_diff'], AvgStats[stat][8:23])  
    print(str(stat) + ' correlation: ' + str(correlation) + ' p-value: ' + str(pvalue) + ' significant: ' + str(pvalue))
```

```
FGM correlation: -0.42305928633428824 p-value: 0.11613661137621333 significant: False  
FGA correlation: -0.5012372784307941 p-value: 0.05698450777103638 significant: False  
FG_PCT correlation: 0.0694273546202671 p-value: 0.8057911963164703 significant: False  
FG3M correlation: -0.5163934758540198 p-value: 0.04875004355397277 significant: True  
FG3A correlation: -0.5353039944249238 p-value: 0.03974744978548436 significant: True  
FG3_PCT correlation: 0.24515387248071663 p-value: 0.3785039712320628 significant: False  
FTM correlation: 0.4971799067432075 p-value: 0.05935138477398661 significant: False  
FTA correlation: 0.48358047514310304 p-value: 0.06780742456388333 significant: False  
FT_PCT correlation: -0.09030817040809816 p-value: 0.7489127409375236 significant: False  
OREB correlation: 0.5403054410302722 p-value: 0.037588814752301115 significant: True  
DREB correlation: -0.5981169401863129 p-value: 0.018511019337312303 significant: True  
REB correlation: -0.5588721118552907 p-value: 0.030333976038766673 significant: True  
AST correlation: -0.3693910127972028 p-value: 0.17540347677143342 significant: False  
TOV correlation: -0.0793811030471907 p-value: 0.7785468864134885 significant: False  
STL correlation: -0.48690738522190674 p-value: 0.06566300277014432 significant: False  
BLK correlation: 0.21559797681935264 p-value: 0.4402859494764434 significant: False  
BLKA correlation: 0.21559797681935264 p-value: 0.4402859494764434 significant: False  
PF correlation: 0.43218236767496476 p-value: 0.10765727151804479 significant: False  
PFD correlation: -0.2135073366737569 p-value: 0.4448387717579506 significant: False  
PTS correlation: -0.38834477335943346 p-value: 0.1525899498454297 significant: False
```

Pace has been increasing lately, could that explain some of the rise in field goal attempts and rebounds and the decline in HCA?

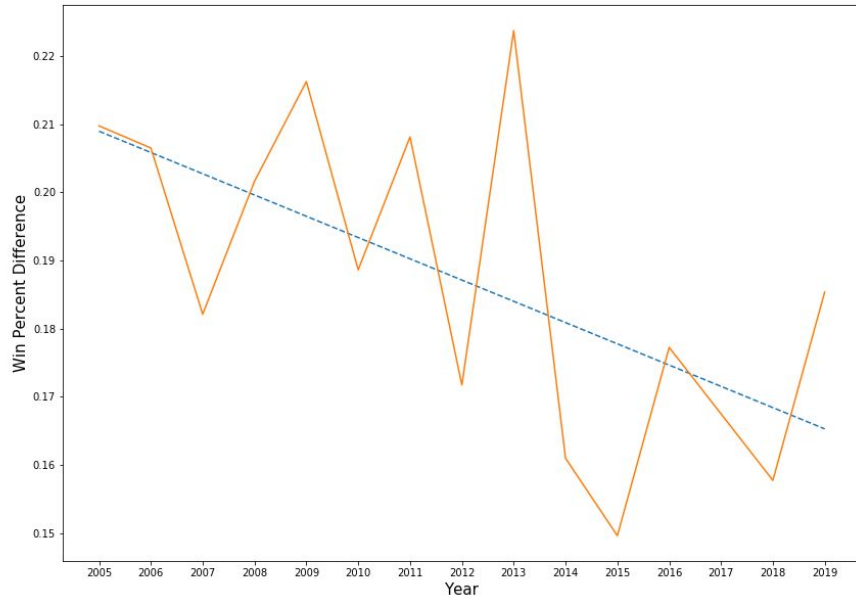
Possessions per game per year in the NBA



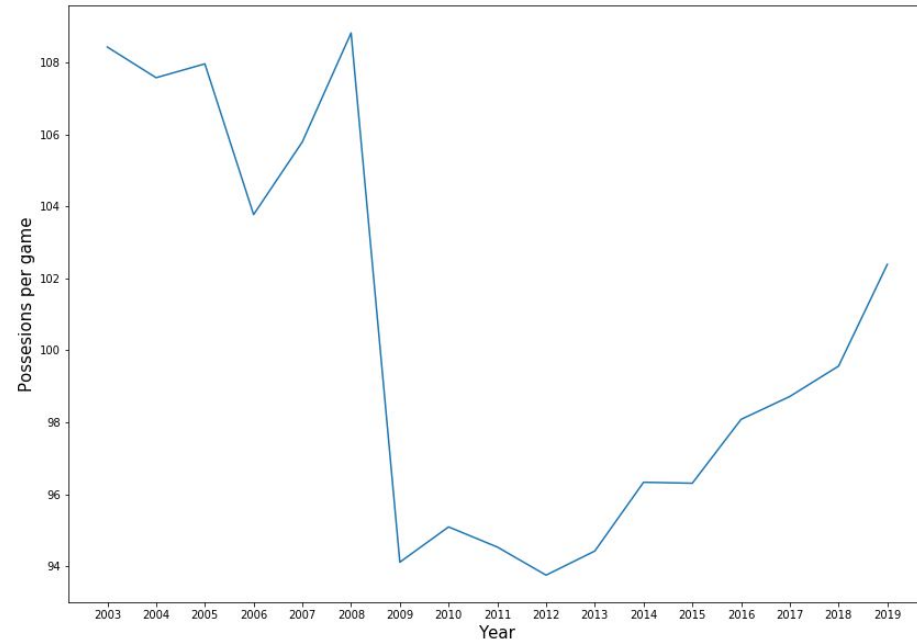
Looks Unlikely

```
import scipy.stats as stats
stats.pearsonr(PaceStats.groupby('year').mean()['PACE'][2:17], AllWinPct2.groupby('year').mean()['win_pct_diff'])
(0.14168747336567394, 0.6144719839663642)
```

Avg Win Percent Difference (Home-Away) per year in the NBA

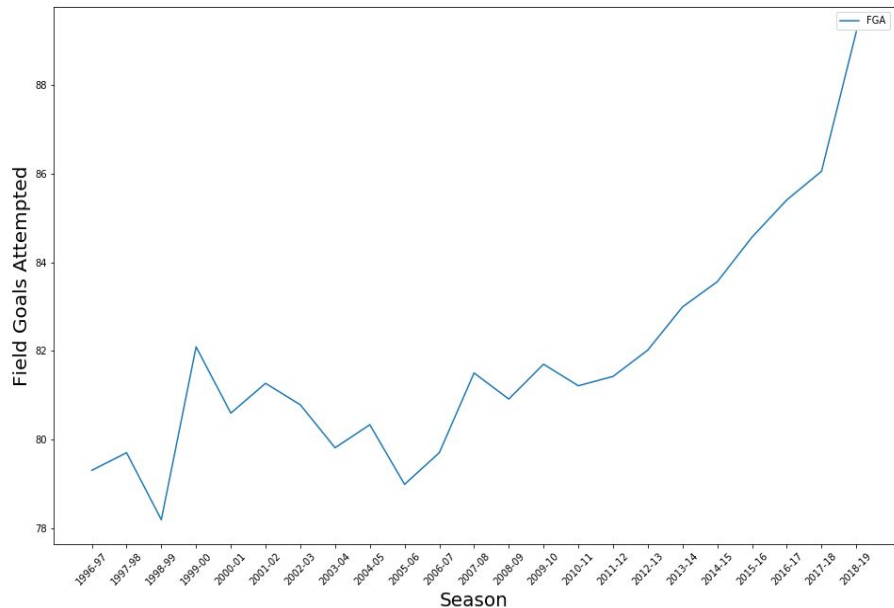


Possessions per game per year in the NBA

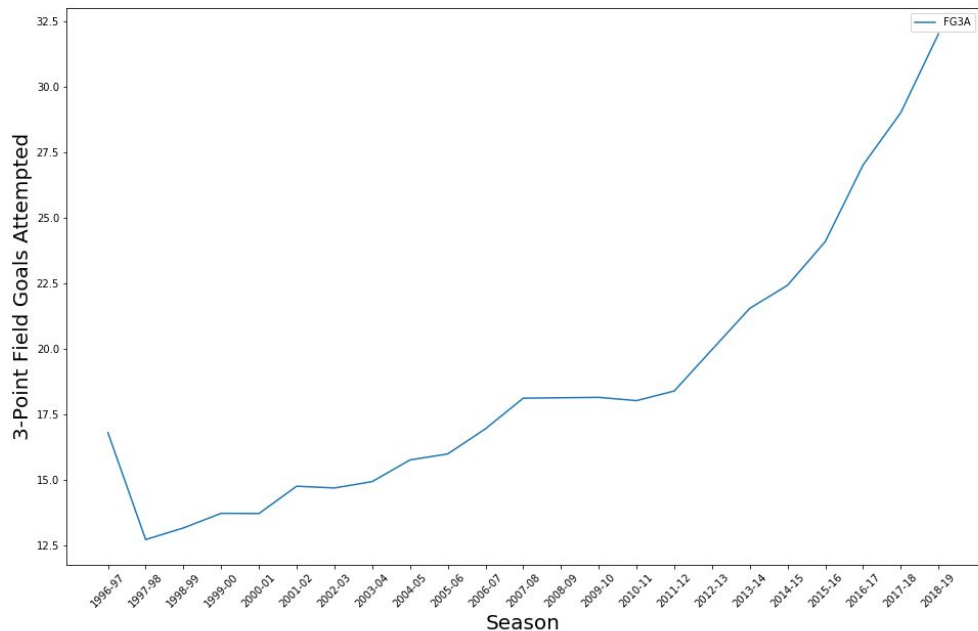


Increased Pace means all shots have been going up, have three-pointers grown significantly faster than overall shots?

Field Goals Attempted Per Game

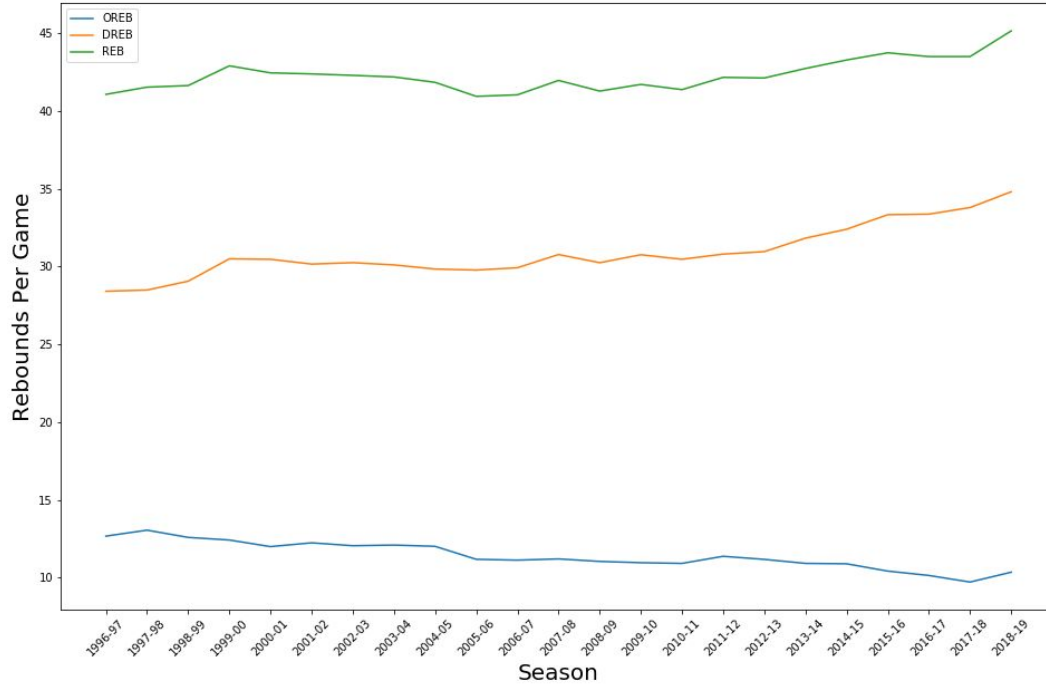


3-Point Field Goals Attempted Per Game



Rebounds

Average Rebounds Per Game



Overall Rebounds and Defensive Rebounds are rising, but Offensive Rebounds are declining

Again Rise in Three Pointers Might Explain The Trends of Rebounds

Three Point Shots are more likely to be missed, leading to a rebound. Three Point Field Goals are also less likely to be offensively rebounded

Rebounding 2's vs. 3's

Ken Pomeroy | 02.26.04

Here's what I have come up for offensive rebounding percentages (OR%) for each type of missed shot. This is based on all games involving at least one D-1 team played between January 1 and February 21.

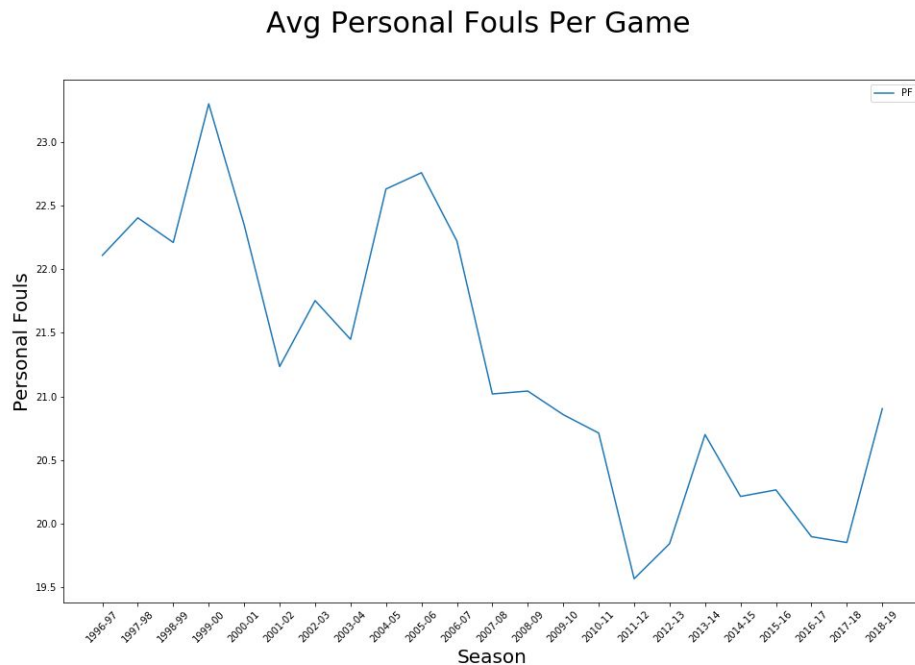
3 point shot: 21.0% (21% of missed 3 pointers are rebounded by the offense)

2 point shot: 41.1%

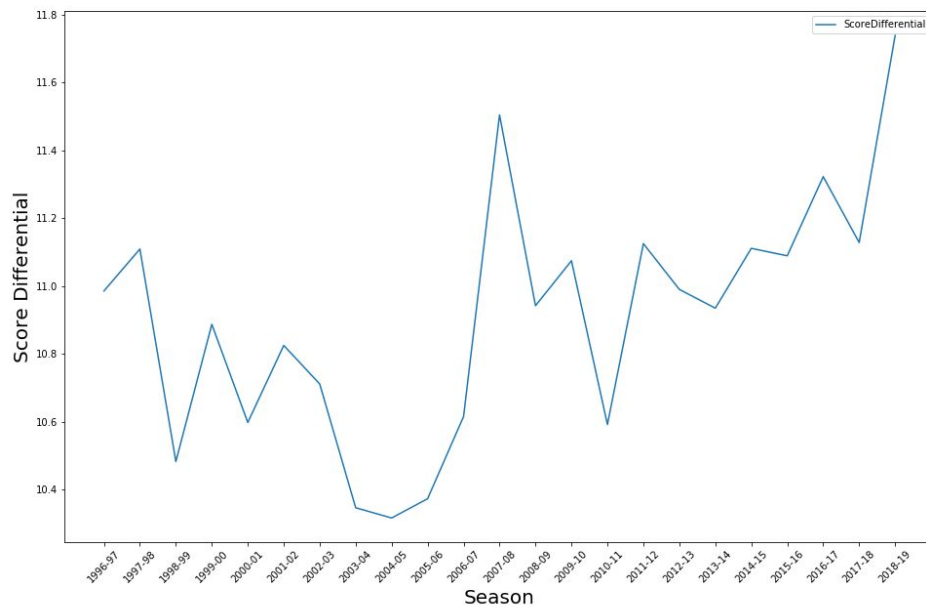
Free throw: 20.3%

PF are also going down in NBA

- Less likely to be fouled on a three point shot
- Perhaps NBA officiating has gotten better over the years
- Video Review
- This might decrease the effects of 'home cooking'



Avg Score Differential



HCA and Score Differential Correlation

```
stats.pearsonr(AllWinPct2.groupby('year').mean()['win_pct_diff'], AvgStats['ScoreDifferential'][8:23])  
(-0.3609461015693904, 0.18625061996660663)
```

FG3A and Score Differential Correlation

```
stats.pearsonr(AvgStats['FG3A'], AvgStats['ScoreDifferential'])
```

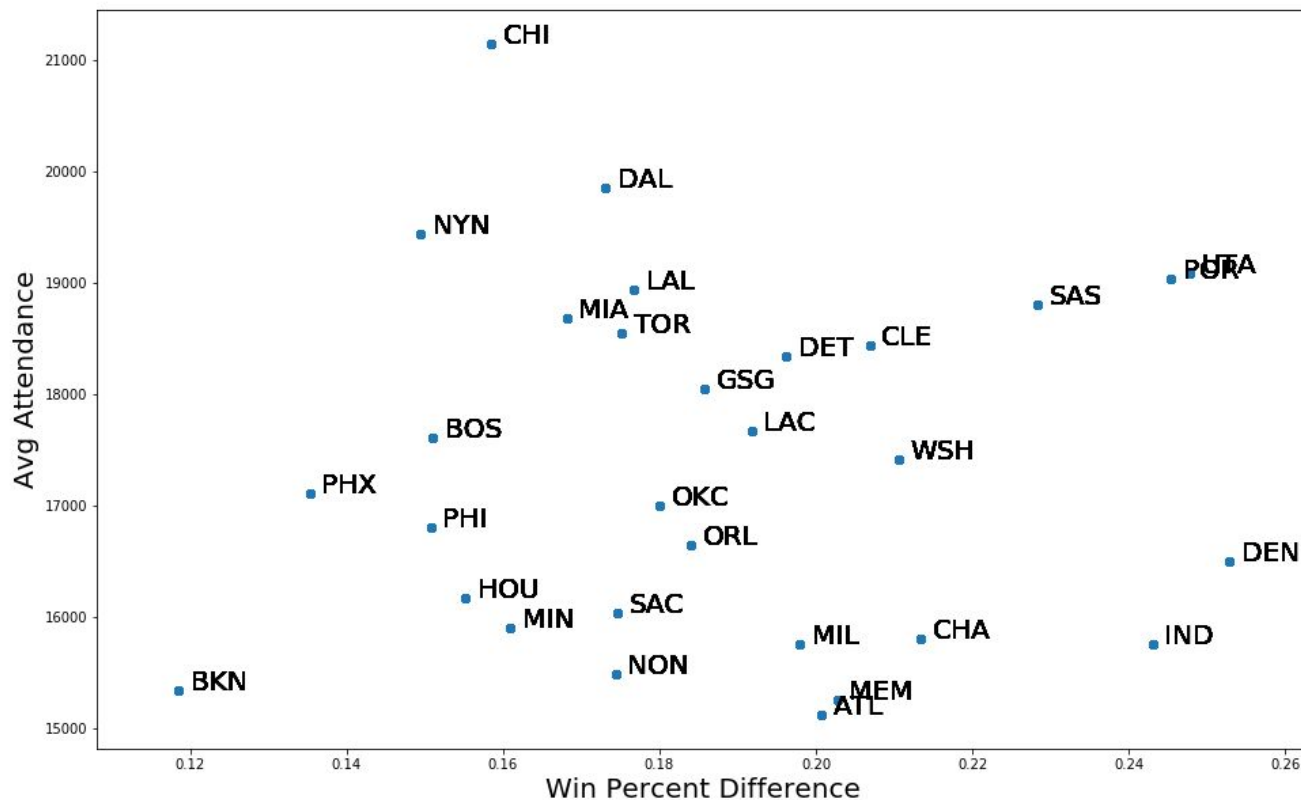
```
(0.6718624820782506, 0.00044656888118660854)
```

Attendance and HCA

AVG_Attend	
Team	
Bulls	21148.722222
Mavericks	19855.277778
NY Knicks	19440.000000
Jazz	19087.444444
Trail Blazers	19035.000000
Lakers	18942.166667
Spurs	18805.333333
Heat	18680.277778
Raptors	18556.444444
Cavaliers	18440.444444
Pistons	18338.500000
Warriors	18047.388889

win_pct_diff	
Team_Abbreviation	
DEN	0.252821
UTA	0.247795
POR	0.245381
IND	0.243081
SAS	0.228381
CHA	0.213402
WSH	0.210544
CLE	0.206898
MEM	0.202710
ATL	0.200739
MIL	0.197931
DET	0.196206

Win Percent Difference vs Avg Attendance

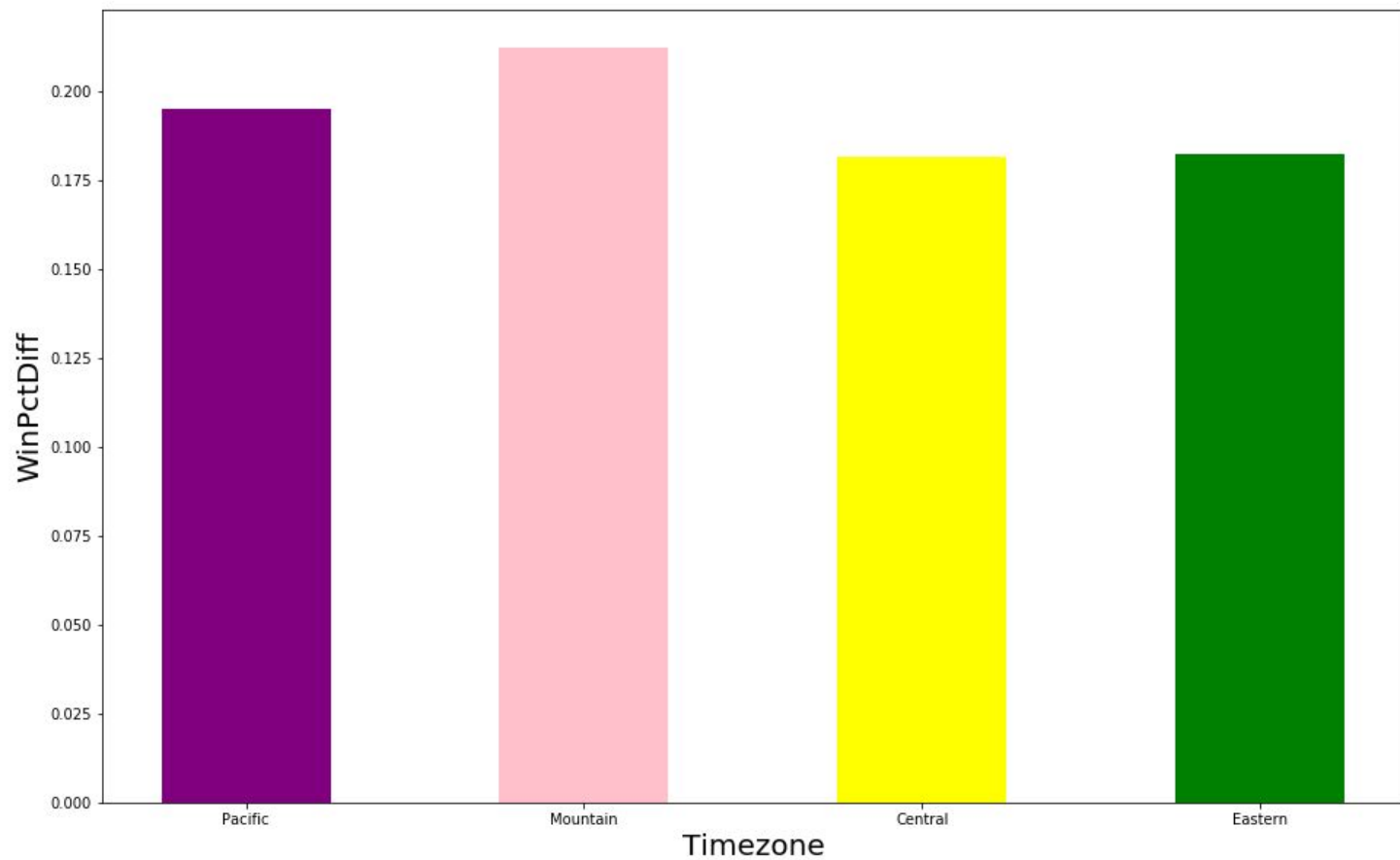


Timezone vs HCA





Avg Win Percent Difference by timezone



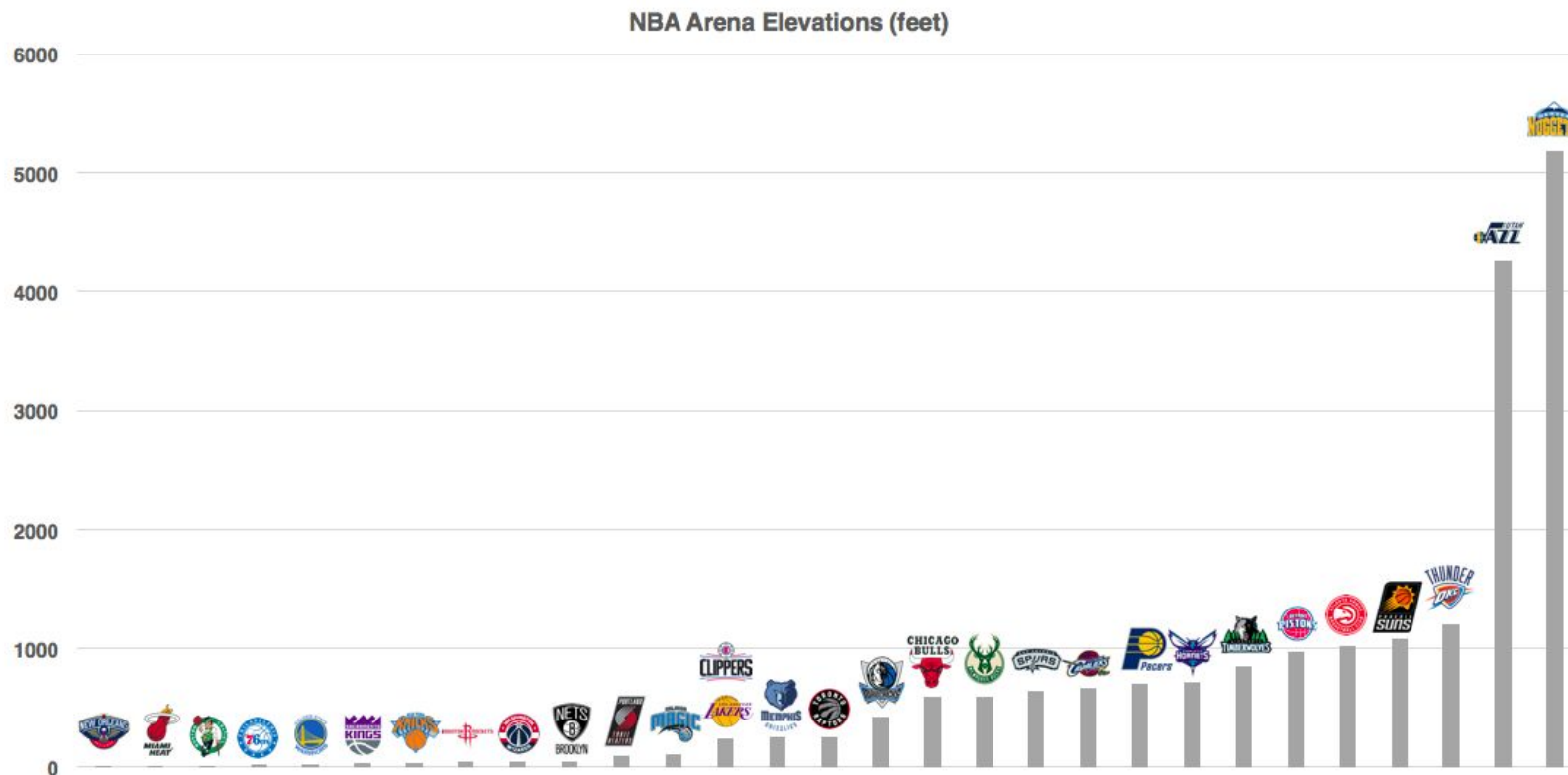
Slight advantage for Western Teams?

There is a theory that teams on the west coast may have a slight advantage over east coast teams when travelling.

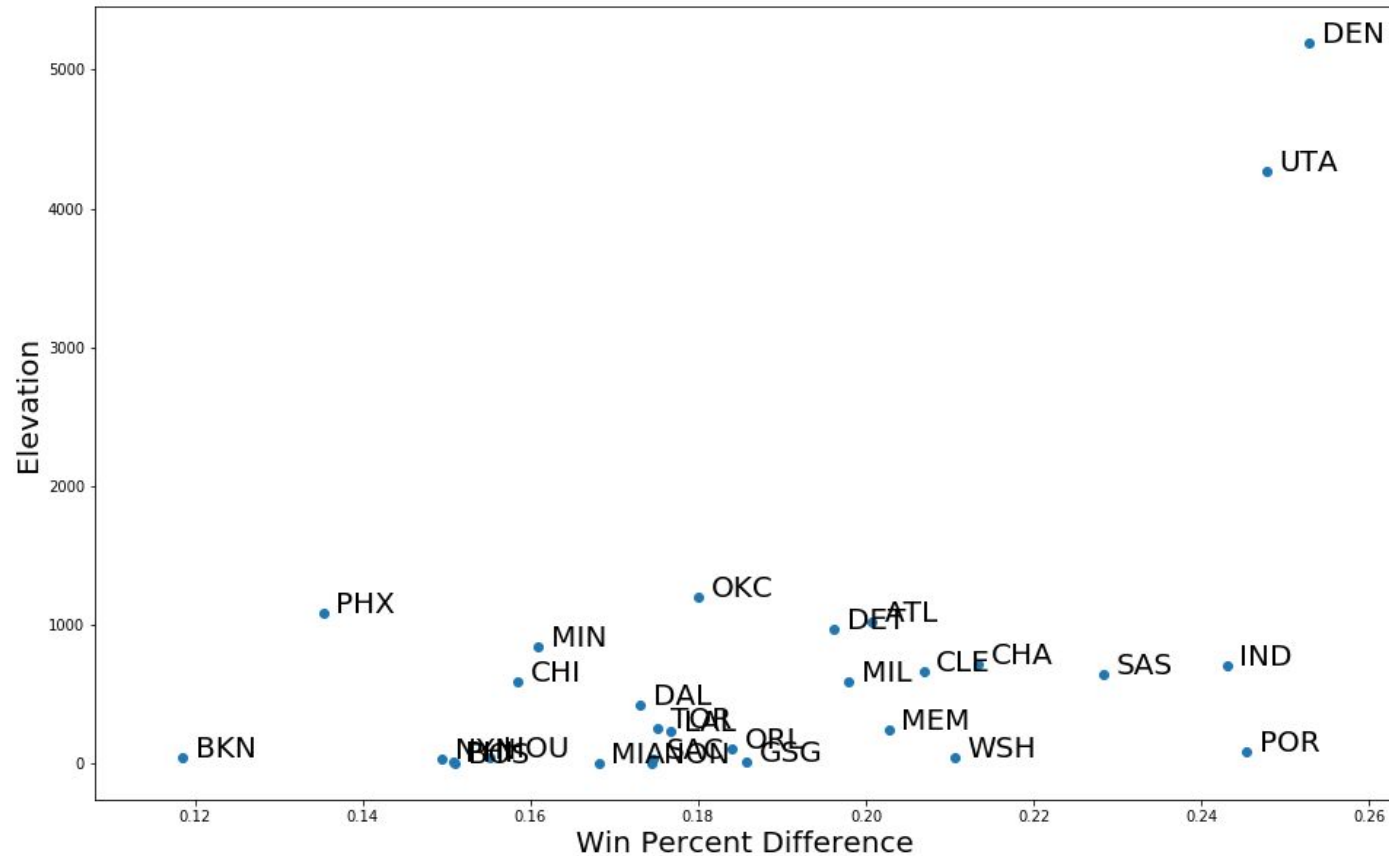
The idea is that if you are playing a 7pm game on the west coast that feels like 10pm for the east coast team, and if you are playing a 7pm game on the east coast, that feels like 4pm for the west coast team.

If you are planning to compete in a physical activity I imagine you would prefer to compete at 4pm than at 10pm.

Elevation and HCA



Win Percent Difference vs Elevation



Takeaways

- Attendance has no discernable effect
- Time Zones might initially seem to have a modest effect, but when accounting for the two highest elevation teams both being in the Mountain Time Zone, it looks like there is little to no effect on HCA
- Elevation seems to have an effect once a certain threshold is reached

Maximize HCA

- Build your stadium as high as possible, expansion teams look at Mexico City, or perhaps Provo, Utah
- Defend the three point line, try to force the other team to settle for 2 point field goals
- Maybe try a stadium in the Pacific or Central Time Zone, just in case...

Minimize Away Court Disadvantage

- Practice/train in high altitude or simulated high altitude conditions
- Shoot a lot of threes
- Play at a fast pace, try to get a high score differential (put the game away before the 4th quarter)



Skills Developed

- Pandas
- Matplotlib
- Hypothesis Testing: Z-scores

Honorable Mentions:

Github and Google Slides