

Analisi del campionato di basket NCAA (1985-2020)

Gabriele Dominici

Campionato di basket NCAA

Division I

Il campionato di basket dell'NCAA (National Collegiate Athletic Association) è diviso in tre categorie:

- Division I
- Division II
- Division III

La principale, la Division I, raggruppa circa 350 tra le migliori squadre di college degli Stati Uniti, divise in 32 Conference (8 o 16 squadre).

Tra i giocatori di queste squadre troviamo i futuri professionisti del campionato di basket più famoso al mondo, l'NBA

Regular Season

Il campionato NCAA comincia i primi di novembre e si conclude al termine di marzo. Nella prima parte di stagione, chiamata Regular Season, si distinguono due periodi:

- non conference schedule: in cui ogni college decide il proprio calendario, scegliendo che squadre affrontare
- conference schedule: in cui ogni college gioca almeno una partita contro ogni altra squadra appartenente alla sua Conference

Questo periodo è fondamentale per poter giocare la seconda parte di stagione

March Madness

La seconda parte di stagione è quella più importante ed è divisa anch'essa in due parti:

- conference tournament: in cui ogni squadra partecipa al torneo della propria Conference, nel quale viene definito il campione stagionale di tale Conference
- March Madness: in cui le migliori 64 squadre di college si sfidano per ottenere l'ambito premio

Quest'ultima si svolge tra marzo e l'inizio di aprile ed è uno degli eventi più seguiti dell'anno negli Stati Uniti

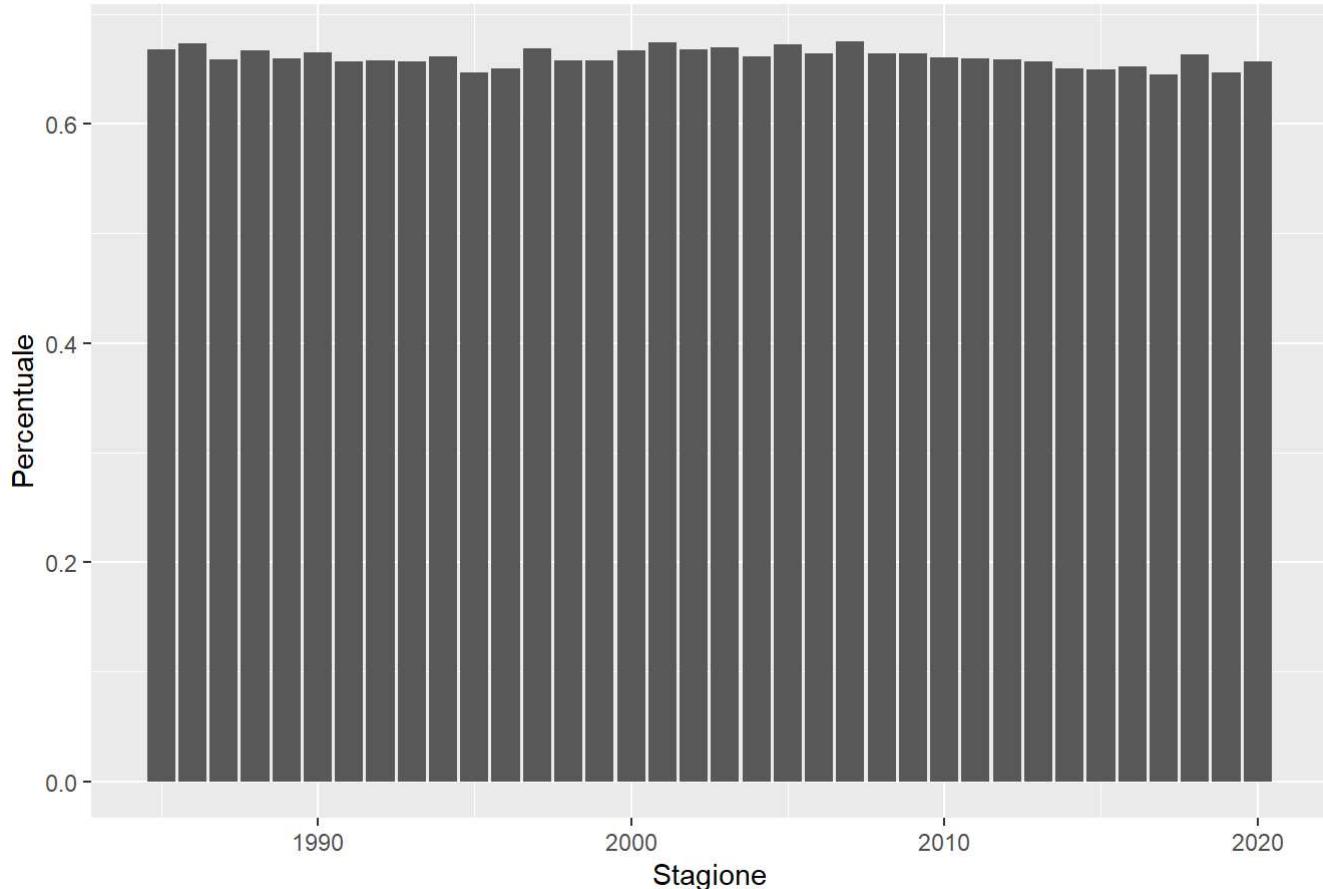
Analisi del campionato tra il 1985 ed oggi

Fattore campo

Il fattore campo è spesso considerato fondamentale, una marcia in più per la squadra di casa, che conosce lo stadio in cui viene disputata la partita, ma soprattutto ha il sostegno del pubblico amico.

Con il passare degli anni l'NCAA è sempre più seguita in America e, spesso, alcuni biglietti, per delle normali partite di college, vengono venduti a prezzi folli, più alti rispetto a quelli di alcune partite NBA. Le partite delle squadre più blasonate, inoltre, sono quasi sempre sold out, creando quindi un ambiente molto suggestivo per i giocatori, i quali si ritrovano a giocare sotto gli occhi di migliaia di persone.

Percentuale di vittoria della squadra di casa



Dal grafico possiamo notare che il fattore campo ha avuto sempre una componente fondamentale, durante lo svolgimento delle partite, il quale non è stato influenzato dall'incremento di interesse e dall'afflusso di tifosi.

Rapporto medio di vittorie in casa

```
## [1] 0.6607463
```

Squadre vincenti

In molti college è di fondamentale importanza il campionato di NCAA e molto spesso offrono ai ragazzi più promettenti delle High School, ma anche a giocatori provenienti dal resto del mondo, numerose borse di studio, per attrarli a giocare nelle loro sedi. Spesso però le squadre più blasonate sono anche quelle in grado di attrarre i talenti migliori e, di conseguenza, ogni anno avere squadre sempre più competitive.

Squadre con il win rate più alto dal 1985 ad oggi

## # A tibble: 10 x 10										
##	TeamName	Win	Lose	Points	Rate	PF	PS	DIFF	DIFFPG	Rank
##	<chr>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
## 1	Kansas	911	212	1822	0.811	89514	75813	13701	12.2	1
## 2	Duke	926	220	1852	0.808	94480	78349	16131	14.1	2
## 3	Kentucky	870	261	1740	0.769	87694	75668	12026	10.6	3
## 4	Arizona	841	272	1682	0.756	87856	76337	11519	10.3	4
## 5	Gonzaga	786	268	1572	0.746	81370	70405	10965	10.4	5
## 6	North Carolina	853	291	1706	0.746	93312	81461	11851	10.4	6
## 7	Syracuse	814	309	1628	0.725	85575	76399	9176	8.17	7
## 8	Xavier	767	319	1534	0.706	82990	74752	8238	7.59	8
## 9	Michigan St	765	332	1530	0.697	80411	71631	8780	8.00	9
## 10	Memphis	774	349	1548	0.689	84570	76372	8198	7.30	10

Nella tabella vediamo le squadre che dal 1985 ad oggi hanno performato meglio. Nelle prime posizioni sono presenti atenei di grandissima fama come Kansas (dove hanno giocato Paul Pierce, Joel Embiid, Wilt Chamberlain ...), Duke (Zion Williamson, Kyrie Irving, Jayson Tatum ...) oppure Kentucky (Anthony Davis, John Wall, DeMarcus Cousins) che negli anni hanno sfornato moltissimi talenti

Squadre che hanno performato meglio negli ultimi 10 anni rispetto a prima

## # A tibble: 10 x 7						
##	TeamName	Win	Lose	Rate	Rank10	Rank Change
##	<chr>	<int>	<int>	<dbl>	<int>	<int>
## 1	Harvard	211	95	0.690	26	176 150
## 2	Wofford	205	136	0.601	74	219 145
## 3	St Bonaventure	203	139	0.594	81	211 130
## 4	Northwestern	176	176	0.5	170	297 127
## 5	Buffalo	213	126	0.628	54	177 123
## 6	Lehigh	191	146	0.567	103	224 121
## 7	William & Mary	167	164	0.505	165	285 120
## 8	LIU Brooklyn	188	158	0.543	133	251 118
## 9	Stony Brook	218	122	0.641	45	161 116
## 10	San Diego St	260	87	0.749	8	120 112

Qui troviamo le migliori dieci squadre che hanno performato meglio negli ultimi 10 anni rispetto a prima. Alcune squadre nella precedente tabella erano posizionate molto in basso e, grazie a delle ottime ultime stagioni, sono riuscite a risalire il ranking. Menzione speciale la meritata Harvard, Buffalo e San Diego St.

Confronto con le posizioni delle top 10 squadre negli ultimi 10 anni rispetto a prima

## # A tibble: 10 x 7						
##	TeamName	Win	Lose	Rate	Rank10	Rank Change
##	<chr>	<int>	<int>	<dbl>	<int>	<int>
## 1	San Diego St	260	87	0.749	8	120 112
## 2	St Mary's CA	269	77	0.777	5	61 56
## 3	Wichita St	245	75	0.766	7	62 55
## 4	Villanova	276	84	0.767	6	20 14
## 5	Belmont	257	86	0.749	9	19 10
## 6	Gonzaga	309	48	0.866	1	5 4
## 7	Kansas	306	59	0.838	2	1 -1
## 8	Duke	298	68	0.814	3	2 -1
## 9	Kentucky	265	68	0.796	4	3 -1
## 10	Arizona	272	92	0.747	10	4 -6

In questa tabella vediamo come erano posizionate nella classifica dal 1985 ad oggi le squadre nella top 10 degli ultimi 10 anni.

Confronto con le posizioni negli ultimi 10 anni delle top 10 squadre dal 1985 ad oggi

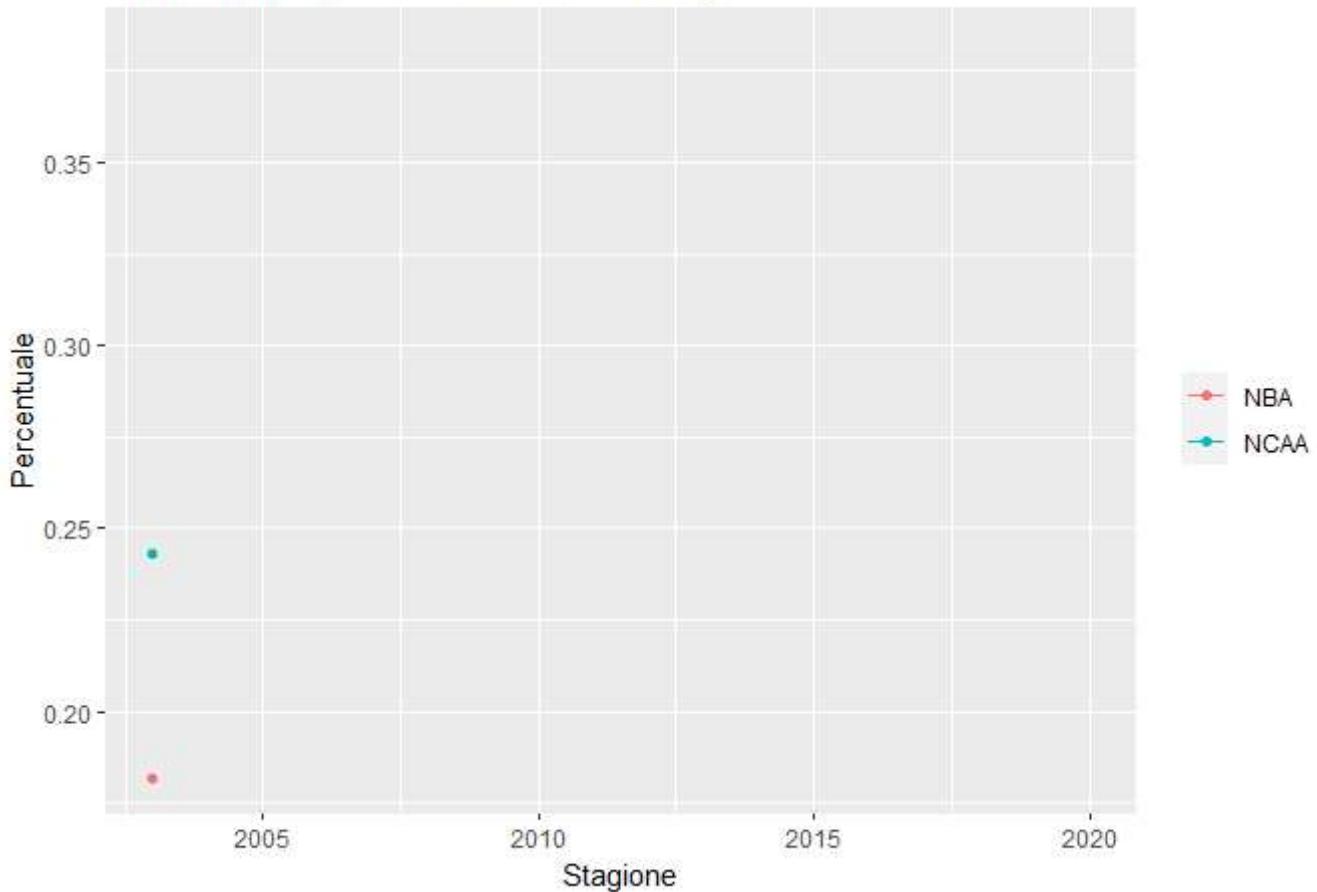
	TeamName	Win	Lose	Rate	Rank10	Rank	Change
	<chr>	<int>	<int>	<dbl>	<int>	<int>	<int>
## 1	Gonzaga	309	48	0.866	1	5	4
## 2	Kansas	306	59	0.838	2	1	-1
## 3	Duke	298	68	0.814	3	2	-1
## 4	Kentucky	265	68	0.796	4	3	-1
## 5	Michigan St	269	94	0.741	12	9	-3
## 6	Arizona	272	92	0.747	10	4	-6
## 7	North Carolina	261	106	0.711	22	6	-16
## 8	Syracuse	251	107	0.701	24	7	-17
## 9	Memphis	245	117	0.677	31	10	-21
## 10	Xavier	241	117	0.673	32	8	-24

Mentre ora, vediamo l'esatto opposto; notiamo, inoltre, la scalata di Gonzaga che si attesta in prima posizione negli ultimi 10 anni, in cui è stata per la maggior parte delle volte una contender, senza mai, però, vincere un titolo.

Cambiamento del gioco negli anni - Tipi di tiri effettuati

Con il passare degli anni, tra l'85 ed oggi, è anche cambiato il modo di giocare a basket, insieme alle regole e alle strategie in campo. Negli anni il tiro da tre nel basket è diventato sempre più presente, influenzando le difese, i movimenti in campo, ma soprattutto il modo di allenarsi, cercando di affinare anche questo aspetto, che in passato era un tipo di azione dedicato alle *guardie*.

Percentuale tiri da 3 nelle diverse stagioni

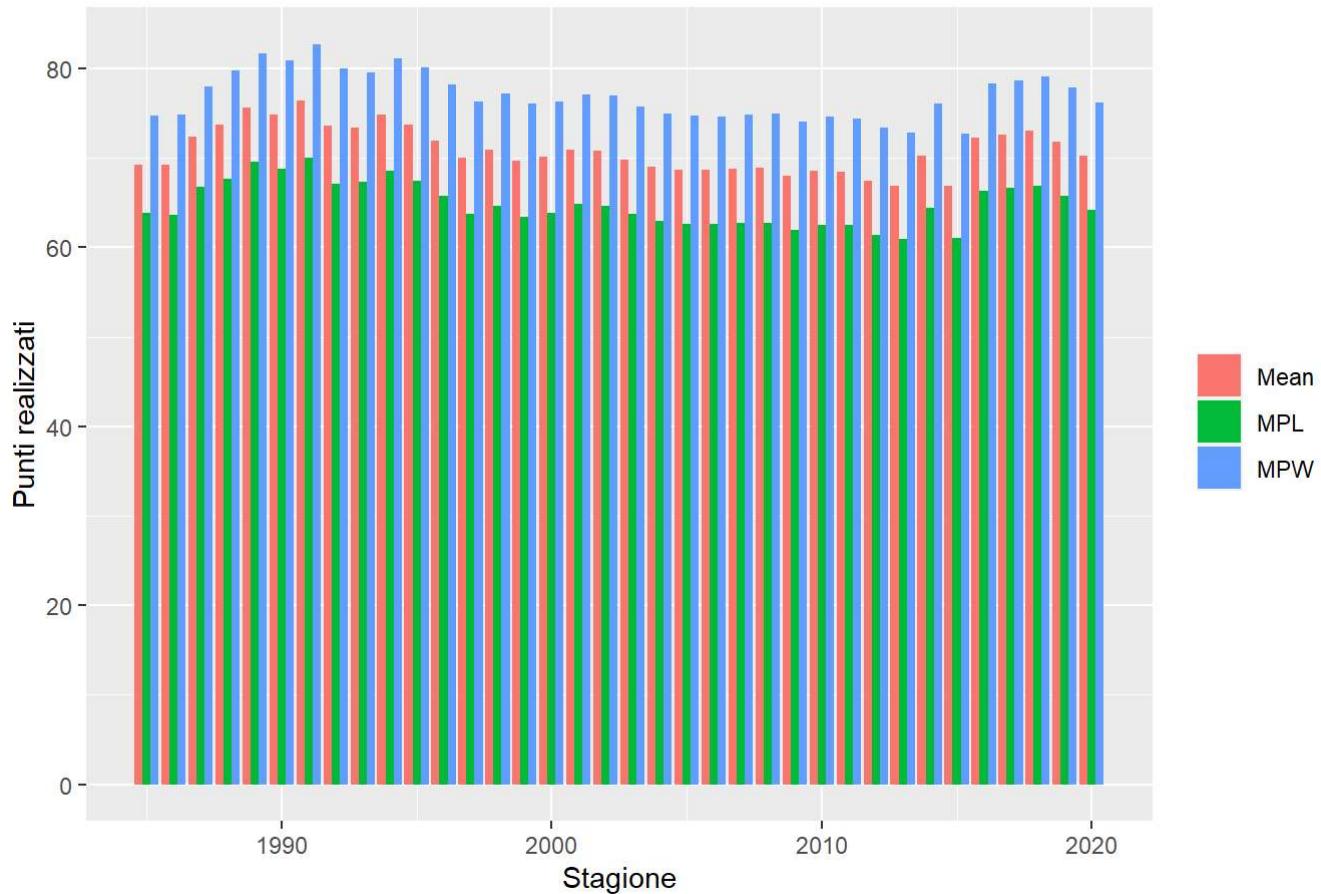


In questa animazione, che confronta la situazione in NBA con quella in NCAA, sul fronte della percentuale di tiri da 3 presi durante una partita, mostra molto chiaramente che in entrambi i casi c'è un incremento netto, nel primo caso si passa dal 18% al 38%, nel secondo dal 24% al 28%. Questo cambiamento ci può mostrare come il basket NCAA sia molto più equilibrato, nel gioco, rispetto a quello NBA che, negli anni, è passato dal basket giocato nel *pitturato* a quello dietro l'arco, passando da un'estremo all'altro. Questo cambiamento è iniziato nei primi anni 2000, ma si è instaurato in seguito ai draft della seconda decade, con l'arrivo nella lega di giocatori del calibro di Steph Curry, Klay Thompson e James Harden, i quali fondano la loro carriera su questo tipo di tiro.

Cambiamento del gioco negli anni - Punteggio

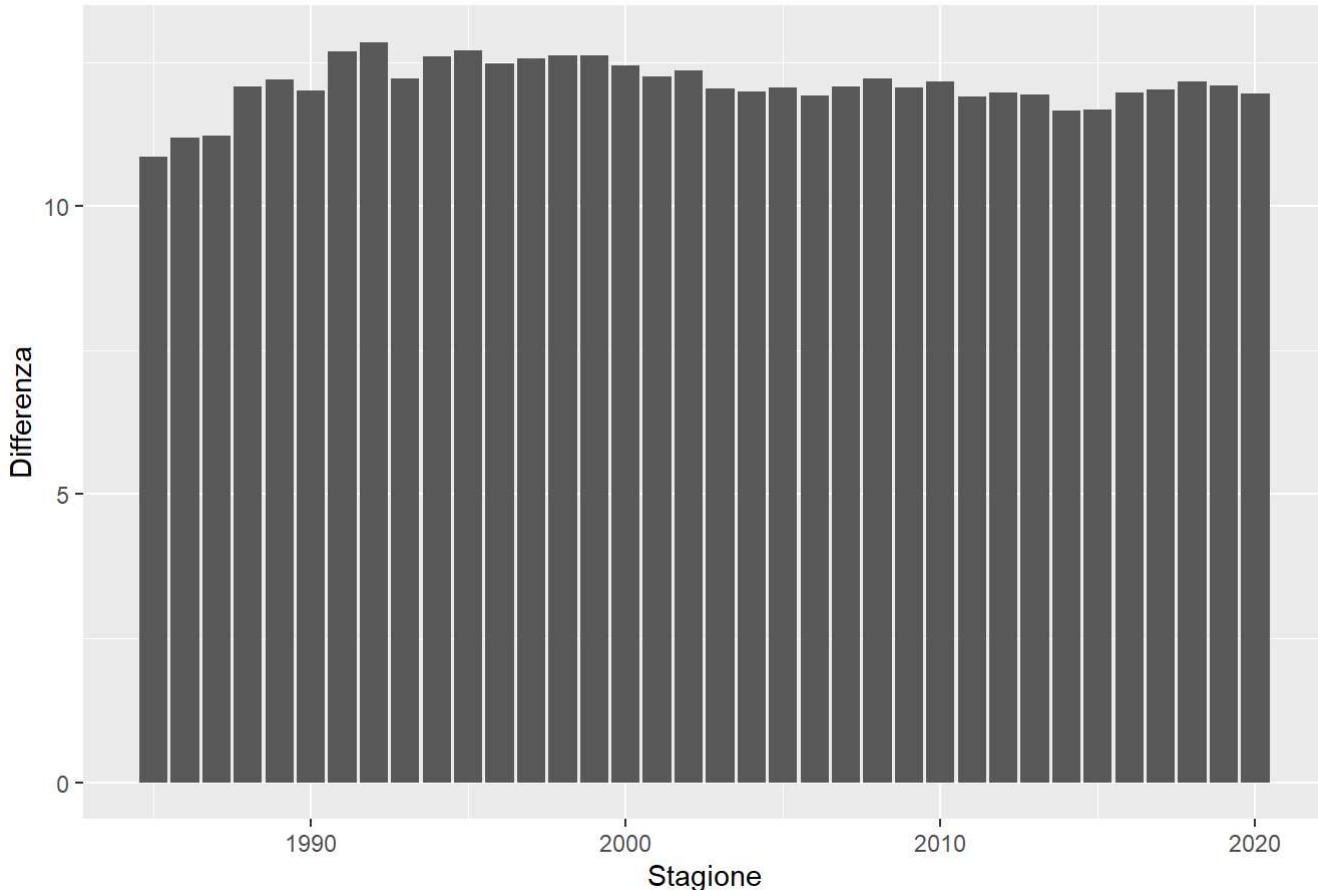
Le modifiche del gioco hanno influenzato però anche i punteggi?

Punteggio medio partite nelle differenti stagioni



I punteggi nel corso degli anni sono altalenanti, ma tutti i tipi di punteggi (Media punti segnati, Media punti segnati dalla squadra vincitrice e Media punti segnati dalla squadra perdente) non subiscono né trend rialzisti, né ribassisti, questo ci mostra che lo scarto medio delle partite negli anni è rimasto pressochè invariato (MPW - MPL).

Scarto medio nelle differenti stagioni

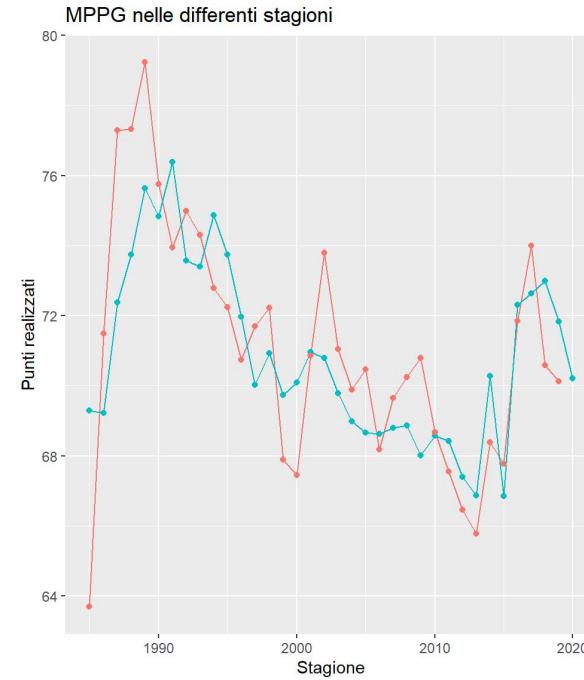
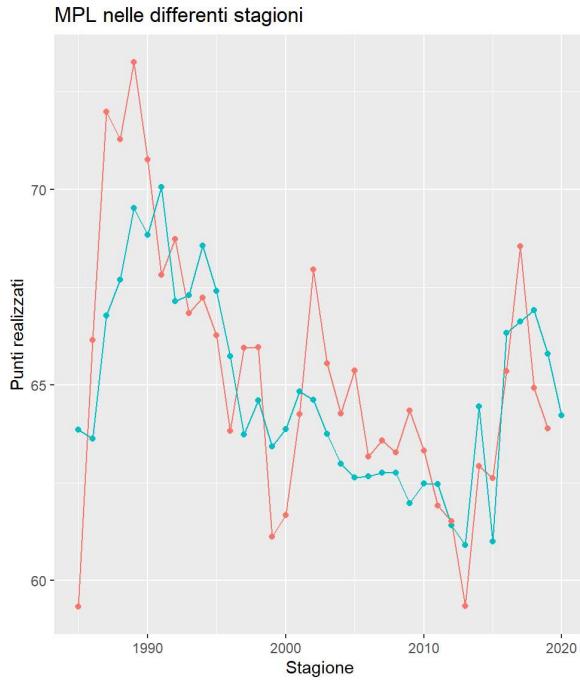
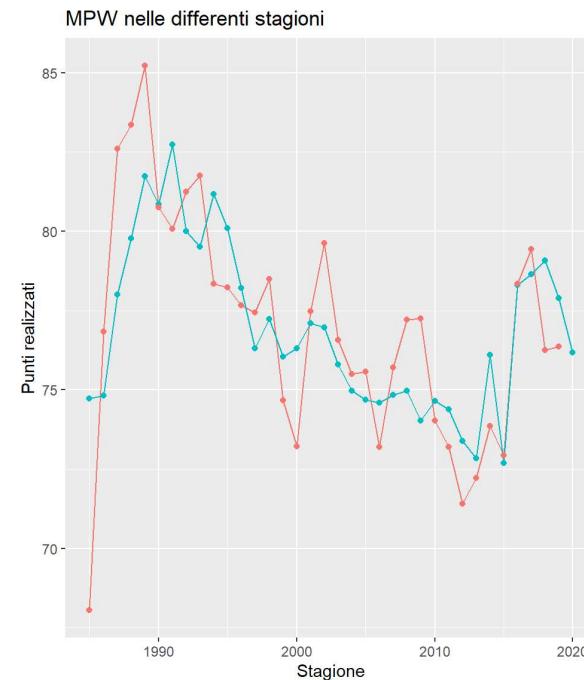
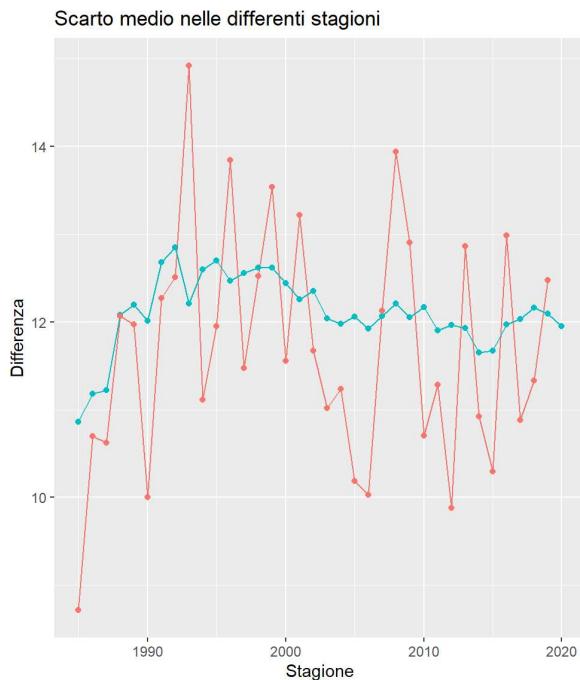


Cambiamento del gioco negli anni - Punteggio RS-MM

Il gioco è stato indubbiamente influenzato, ma nel complesso il punteggio medio delle partite è rimasto stabile. Nonostante questo c'è un'aspetto fondamentale da considerare: la RS è giocata allo stesso modo della MM? Le migliori 64 squadre della nazione si fronteggiano per ottenere l'ambito trofeo, le partite dovrebbero essere più serrate ed equilibrate.

- MPW = Media punti team vincente
- MPL = Media punti team perdente
- MPPG = Media punti per game

Diverse correlazioni tra i punteggi della RS e della MM



Nel grafico in alto a sinistra possiamo notare come lo scarto medio nelle partite della MM è molto altalenante negli anni, ma la maggior parte delle volte è al di sotto della media della RS, ciò conferma il presentimento di partite più equilibrate.

Negli altri 3 grafici vediamo come i punteggi medi durante la MM seguono l'andamento di quelli della RR.

Round MM

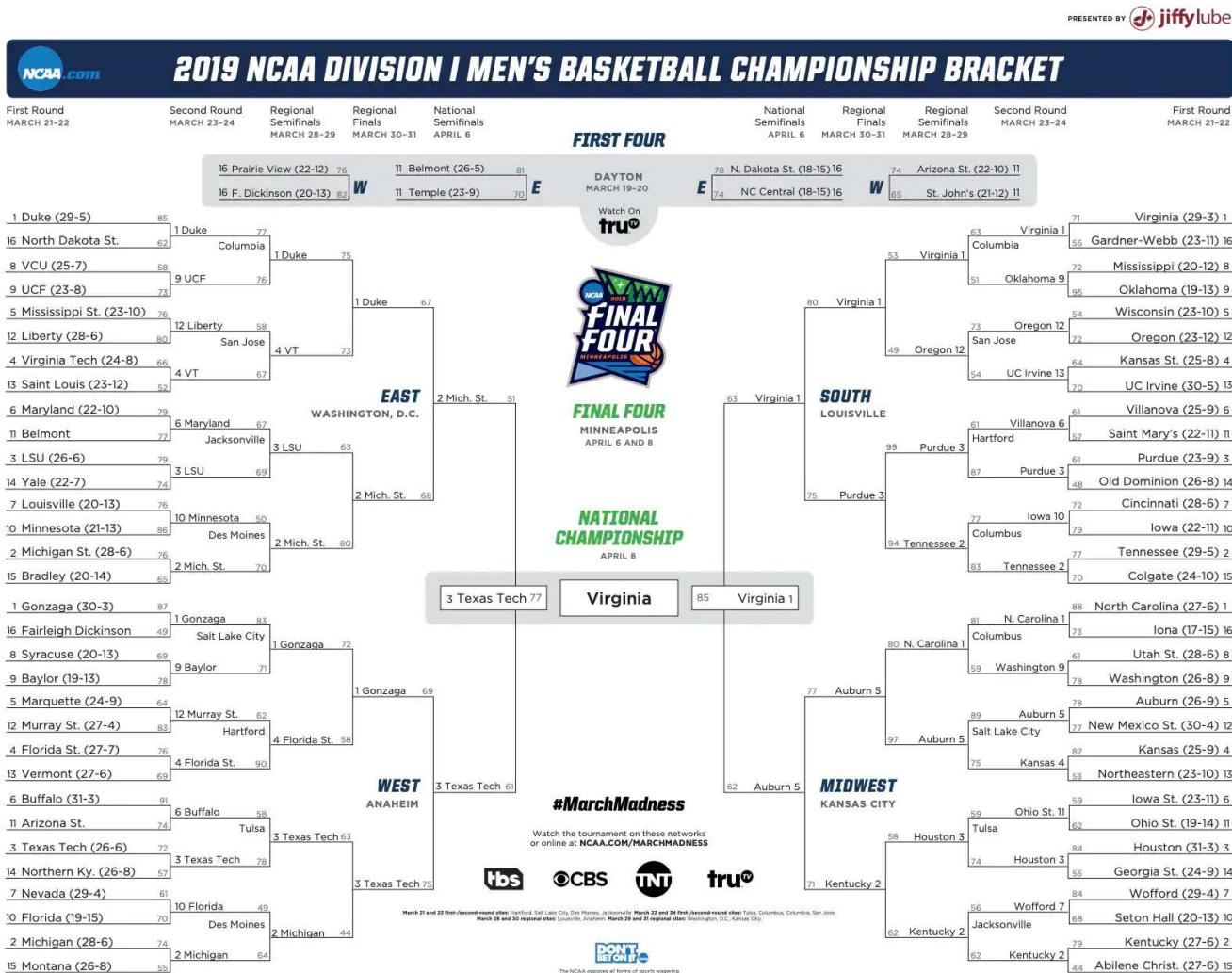
Il bracket della MM è diviso in 4 Regioni, che indicano la diversa provenienza dei college. Negli ultimi anni sono:

- East
- MiddleWest
- South
- West

Durante la March Madness si svolgono diversi round:

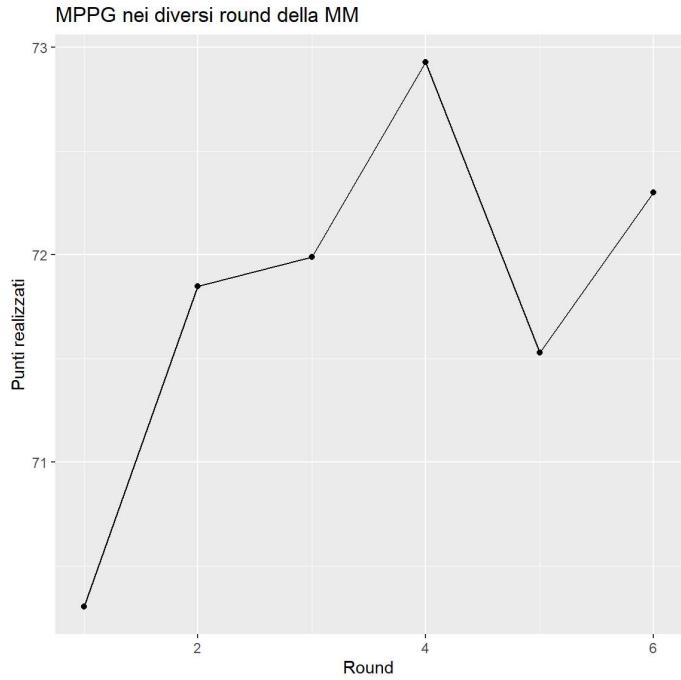
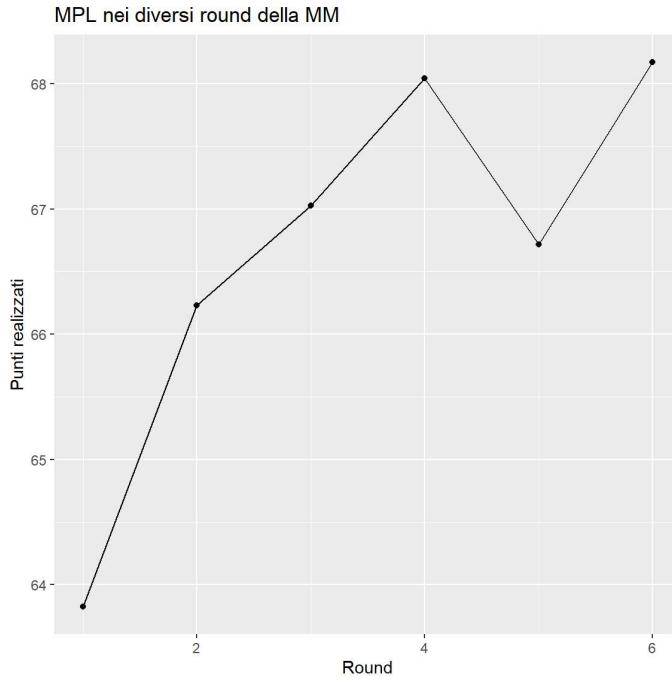
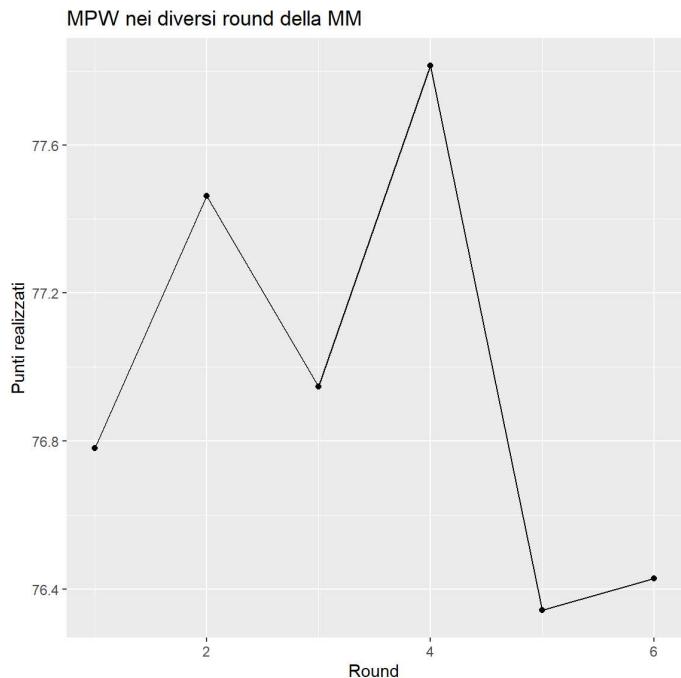
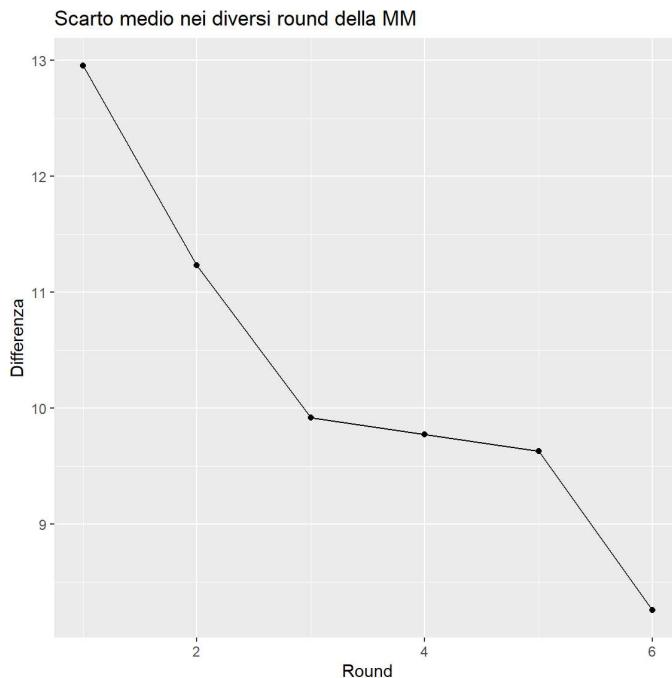
- top 64 - First round
- top 32 - Second round
- top 16 - Sweet Sixteen
- top 8 - Elite Eight
- top 4 - Final Four
- top 2 - Final

Bracket



Se tra RR e MM notavamo delle piccole differenze in termini di punteggio, a maggior ragione ce le aspettiamo tra i differenti Round

Analisi dei vari tipi di punteggio durante i round della MM



Infatti vediamo come con il passare dei round (più ci avviciniamo alla finale) e più lo scarto medio diventa piccolo, sintomo di partite sempre più agguerrite.

Abbassandosi lo scarto medio anche il punteggio medio della squadra vincente e quello della squadra perdente devono avvicinarsi; infatti nel grafico in alto a destra vediamo diminuire, leggermente il primo e, nel grafico in basso a sinistra, alzarsi di molto quello della squadra perdente. Azzardando una conclusione potremmo dire che, con il passare dei round, il livello difensivo rimane invariato, mentre quello offensivo si alza di molto.

Maggior numero di presenze alla MM

Un ateneo può partecipare alla March Madness in due modi:

- vincere il Conference Tournament
- venire scelti dai Commissioner (esperti di basket scelti ogni 5 anni) la domenica precedente alla MM

In questo modo vengono scelte le 64/68 squadre che parteciperanno al torneo conclusivo della manifestazione e, per ogni Regione, viene stilata una classifica dai Commissioner che indica l'assegnazione della seed (il posto all'interno del bracket)

Squadre con il maggior numero di partecipazioni alla MM dal 1985 ad oggi

```
## # A tibble: 10 x 2
##   TeamName      N_partecipazioni
##   <chr>              <int>
## 1 Duke                  34
## 2 Kansas                 34
## 3 Arizona                32
## 4 North Carolina          32
## 5 Kentucky                30
## 6 Michigan St              29
## 7 Syracuse                 28
## 8 Louisville                26
## 9 Oklahoma                 26
## 10 Purdue                  26
```

Le sopracitate Duke e Kansas negli ultimi 35 anni hanno perso solamente una convocazione alla MM. A seguire, comunque, troviamo atenei con un win rate molto alto, infatti 7 su 10 sono presenti nella top 10 del ranking dal 1985 ad oggi

Squadre con il maggior numero di partecipazioni alla MM negli ultimi 10 anni

```
## # A tibble: 10 x 2
##   TeamName      N_partecipazioni
##   <chr>              <int>
## 1 Duke                  10
## 2 Gonzaga                 10
## 3 Kansas                  10
## 4 Michigan St              10
## 5 Cincinnati                 9
## 6 Kentucky                  9
## 7 North Carolina                 9
## 8 Villanova                  9
## 9 Wisconsin                  9
## 10 Florida                  8
```

Non stupisce la presenza al primo posto di Duke e Kansas, ma lo fa la presenza al primo posto di Gonzaga, assente nella tabella precedente. Quest'ultimo fatto conferma, come già detto, le ottime prestazioni di questa franchigia negli ultimi 10 anni.

Squadre che hanno vinto di più

```
## # A tibble: 10 x 3
##   TeamName     Rank N_Championship
##   <chr>       <int>      <dbl>
## 1 Duke          2         5
## 2 North Carolina 6         4
## 3 Connecticut    18        4
## 4 Kentucky       3         3
## 5 Villanova     20        3
## 6 Kansas         1         2
## 7 Louisville    12        2
## 8 Florida        23        2
## 9 Arizona        4         1
## 10 Syracuse      7         1
```

Ai primi due posti troviamo due dei college che hanno partecipato a più MM, sono sempre stati considerati tra i migliori atenei d'America anche per la loro continuità e la solidità dimostrata negli anni. Nonostante questo, non sono le due squadre più vincenti dell'NCAA, in quanto UCLA ha vinto ben 7 titoli, ma tutti prima del 1985. Tra le 10 posizioni ce ne sono due particolarmente interessanti:

- alla posizione numero 3 troviamo l'Università del Connecticut con ben 4 titoli; negli anni ha partecipato a diversi tornei di fine stagione, ma negli ultimi 10 solamente a 4 concretizzandone 2 di essi
- alla posizione numero 5 c'è Villanova, un college che si è messo in mostra negli ultimi 10 anni, partecipando a 9 MM e vincendone 2 di esse.

Analisi 2018-19

Ho voluto analizzare la stagione 2018-19 da un altro punto di vista, avendo già visto in precedenza alcune statistiche riguardanti questa stagione. Ho calcolato, prima, il win rate di ogni college e poi l'ho confrontato con l'elo rate calcolato con la classica funzione, apportandone qualche modifica.

Vincente squadra in casa

$$\text{update} = (k * (1 - \mu * (0.66) / (0.32 * \mu + 0.34))) * 2 / (2 - (\text{score} / \text{maxDiff}))$$

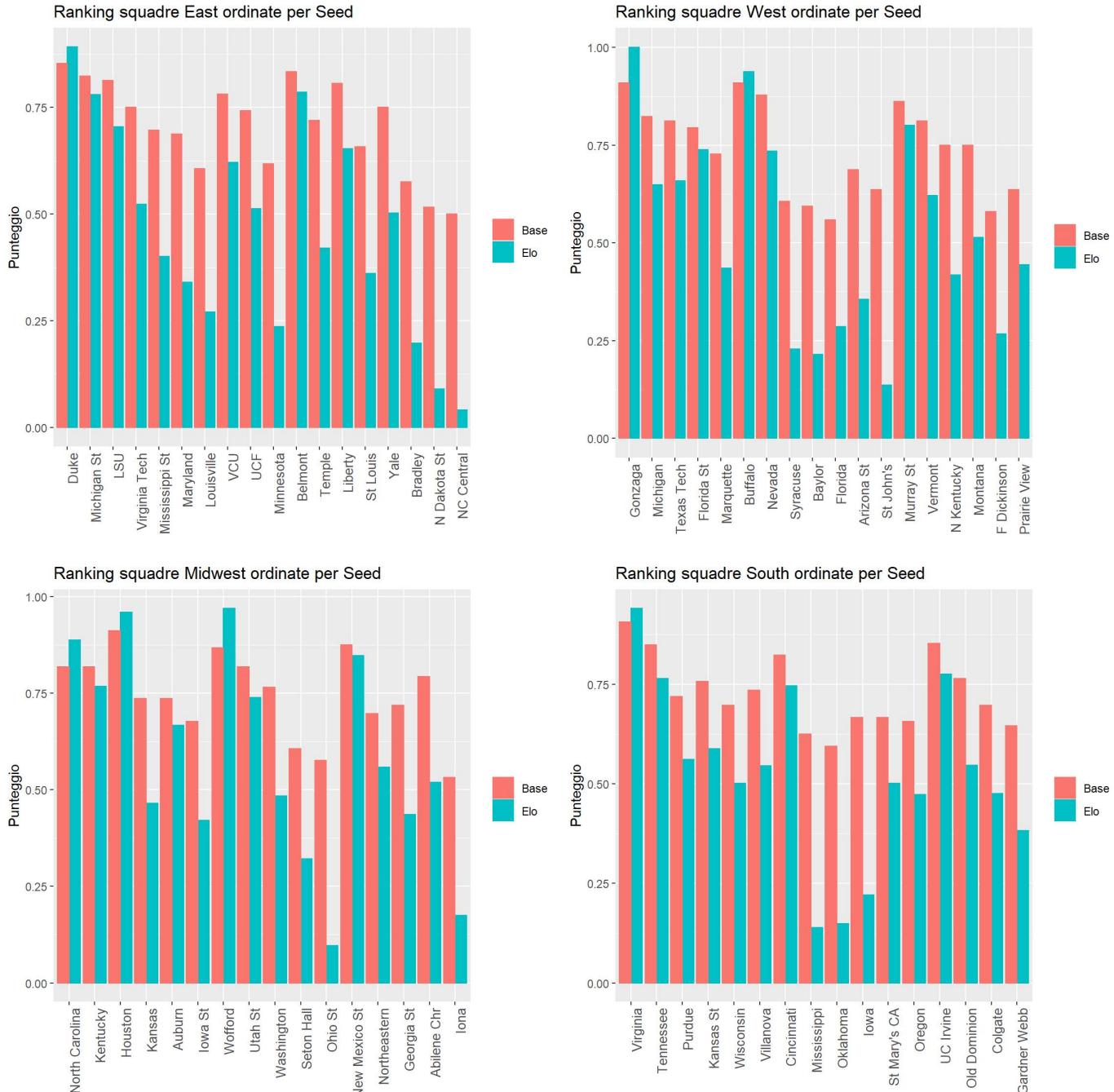
Vincente squadra fuori casa

$$\text{update} = (k * (1 - \mu * (1 - 0.66) / (0.32 * \mu + 0.34))) * 2 / (2 - (\text{score} / \text{maxDiff}))$$

- mu = calcolo match up con il classico elo rating system
- 0.66 = probabilità di vittoria da parte della squadra di casa
- score = risultato della partita come differenza tra il punteggio della squadra vincente e quello della squadra perdente
- maxDiff = mediana dell'1% dei game con la più alta differenza tra il 1985 e il 2018

In questo modo prendo in considerazione anche il fattore campo e lo scarto con cui è terminata la partita.

Le squadre qualificate alla MM, divise per Region ed ordinate per Seed



In questi grafici, uno per ogni Region, sono presenti le 68 squadre scelte dai Commissioner, per partecipare alla MM. In ogni grafico, gli atenei sono disposti rispetto alla Seed a loro assegnata (a sinistra la Seed n.1). Possiamo notare come le scelte fatte da questa commissione non rispecchiano né un ordinamento basato sul win rate, né uno basato sull'elo rate, il quale pesa in modo differente il risultato di una partita, correlandolo all'avversario, al campo di gioco e allo scarto. 32 squadre vengono ammesse attraverso la vittoria del Conference Tournament, alcune di queste leghe, però sono considerate di livello minore rispetto ad altre; questo spiega il perché ci sono squadre con un win rate alto (campioni di Conference) a cui, però, viene assegnata una seed più alta rispetto ad altre squadre, partecipanti a leghe più competitive, con win rate minore. Questa ipotesi è confermata da diverse vincitrici di Conference come Vermont, New Mexico ST, UC Irvine e Liberty.

Comparazione Win Rate - Elo Rate - Seed

Supponendo di dover predire i risultati del bracket dello scorso anno ho utilizzato tre metodi:

1. vincente chi ha il Win Rate maggiore
2. vincente chi ha l'Elo Rate maggiore

3. vincente chi ha la Seed minore

Numero di prediction corrette secondo i vari metodi

```
## [1] 40 43 47
```

Con questi risultati possiamo dire che:

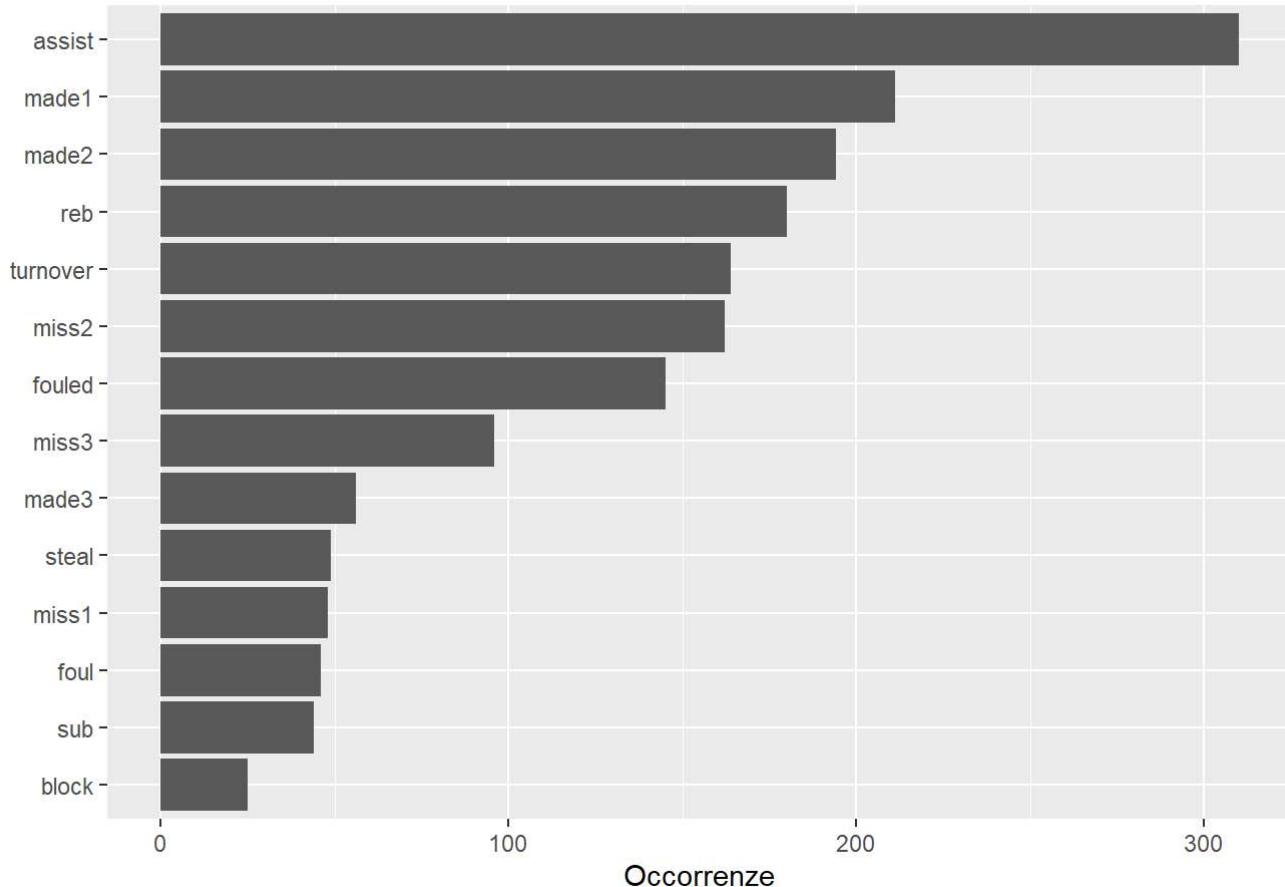
- il metodo che utilizza il win rate è molto povero, perché non cattura alcun tipo di information dalla partita, se non il semplice risultato (Vittoria - Sconfitta)
- il secondo metodo cerca di essere più ricco del precedente, in quanto viene influenzato anche dal tipo di avversario, dal tipo di campo che, come abbiamo visto, può influenzare il risultato, dal risultato effettivo della partita (lo scarto), ma alcune sfumature non riesce a coglierle. Come abbiamo visto, però, se le squadre giocano solo contro atenei della stessa Conference, avremo un ranking veritiero all'interno della Conference, ma non della Region
- il terzo metodo, quello realizzato utilizzando le opinioni di alcuni esperti (i Commissioner), è il più efficace, perché utilizza informazioni non ricavabili dai dati, come la composizione di una squadra, la forma fisica dei giocatori.

Nessuno dei tre metodi, però, riesce a predire i, così chiamati, *upset*, ossia quando una squadra con un Seed molto alto riesce a battere una con un Seed molto basso; è proprio questo il bello del basket, anche una squadra, tra le meno favorite, può ambire al titolo e diventare la *Cinderella* del torneo.

Ja Morant

Nella stagione NCAA, dello scorso anno, mi ha particolarmente emozionato un giocatore: Ja Morant. Un *playmaker*, classe '99, molto atletico ed esplosivo, con delle doti da passatore invidiabili da chiunque, un giocatore estremamente completo. Lo scorso anno era la stella di Murray State, e, insieme ai suoi compagni, ha portato l'ateneo a giocarsi un posto nelle *Sweet Sixteen*, perdendo contro Florida State.

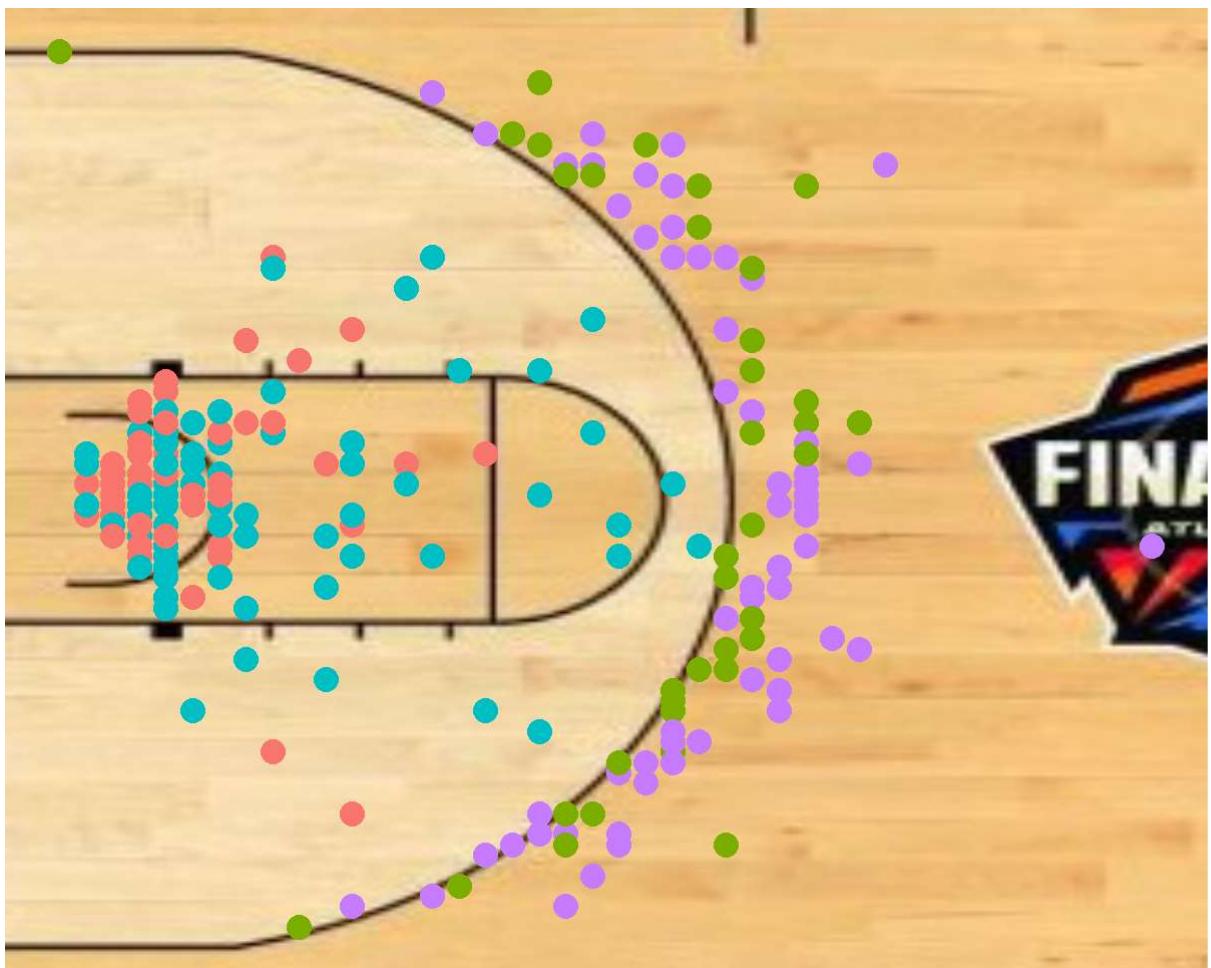
Tipi di azioni effettuate da Ja Morant



Da questo grafico possiamo notare diverse cose:

- l'estrema capacità di giocare per i compagni, effettuando numerosi assist (voce con più occorrenze). Infatti proprio per questo motivo si classifica alla sesta posizione di tutti i tempi, dell'NCAA, per numero totale di assist in una stagione e per assist medi a partita.
- una altissima percentuale di tiri liberi ($\text{made1} / (\text{made1} + \text{miss1})$), che lo fa entrare al terzo posto nella classifica dei giocatori che hanno segnato più tiri liberi in una partita, in NCAA (21 su 21)
- che la percentuale di tiri da 2 è positiva, ha segnato numerose triple, che mostrano, però, che non è una delle sue caratteristiche principali, data la non altissima percentuale realizzativa
- il numero importante di rimbalzi presi, nell'arco della stagione, dimostrando la sua competenza in diversi ambiti. Non a caso, con 3 *triple-doppie* realizzate (andare in doppia cifra in tre statistiche differenti in una partita), è nella top 10 dei giocatori che ne hanno realizzate di più

Distribuzione su metà campo tiri Ja Morant

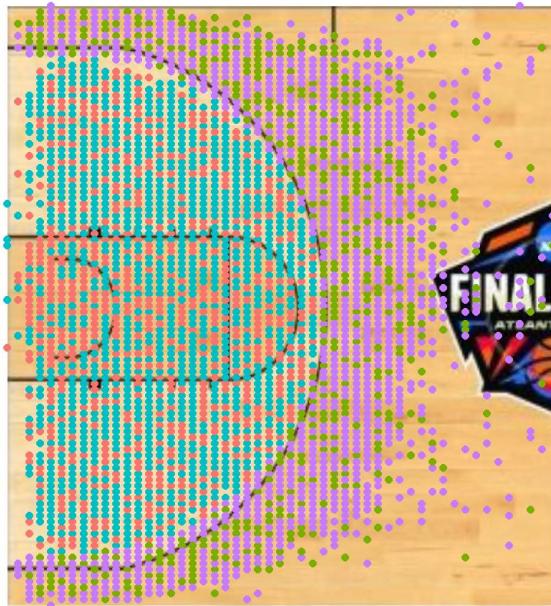


● made2
● made3
● miss2
● miss3

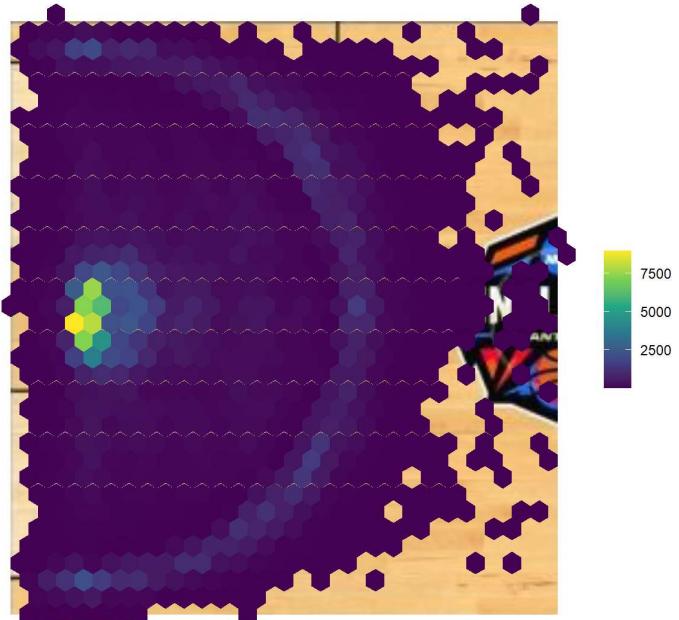
In diverse partite, non tutte, i tiri dei giocatori vengono registrati insieme alla posizione in cui vengono effettuati. Possiamo notare come Ja Morant, grazie al suo atletismo, predilige una conclusione al *ferro*, rispetto ad un *tiro dalla media*; i *tiri dall'arco* non mancano, nonostante il tiro da 3 non sia un suo pregio.

Distribuzione su metà campo tiri effettuati - Stagione 2018-19

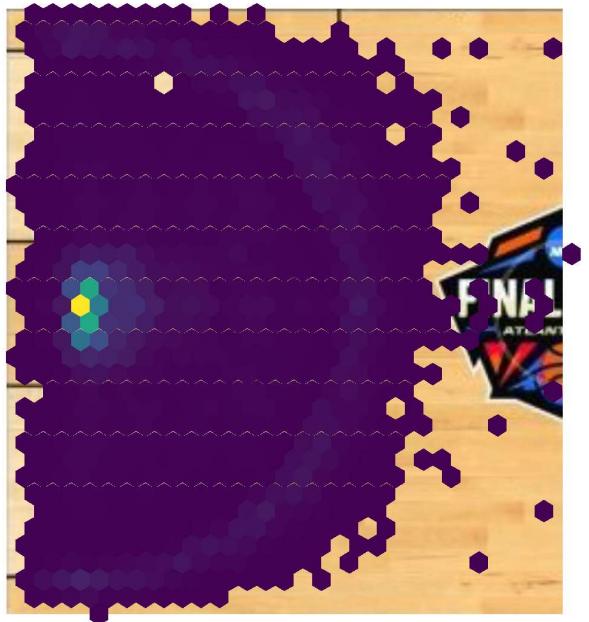
Tiri registrati durante la stagione 2018-19



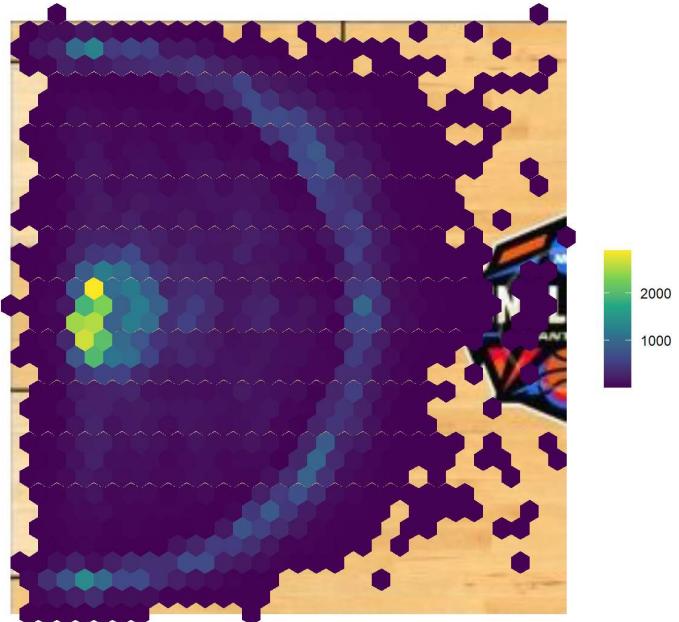
Heatmap dei tiri effettuati



Heatmap dei tiri realizzati



Heatmap dei tiri sbagliati



IL primo grafico ci dà poche informazioni sulle aree da cui i giocatori tirano maggiormente, sembrano distribuiti molto equamente sulla superficie del campo. Negli altri tre grafici, invece, utilizzando una *heatmap*, riusciamo a estrapolare meglio informazioni notando che:

- i tiri dalla media sono quelli che vengono meno utilizzati, di conseguenza le aree più sfruttate sono quelle sotto canestro, con i tiri ad alta percentuale, e quelle dietro l'arco da 3 punti, confermando l'ipotesi fatta in precedenza, sul cambio di tipologia di gioco
- ci sono numerosi tiri realizzati sotto canestro e, rispetto al resto del campo, a ridosso della linea da 3 punti
- numerosi tiri sbagliati sotto canestro, causati dai diversi contatti in cui ci si può imbattere, soprattutto in quella zona del campo, ma come vediamo dalle varie scale di colore sono estremamente minori di quelli segnati. Un'altra zona, che si nota particolarmente, è sempre quella dietro l'arco

STATISTICHE GIOCATORI 2018-19

Infine ho realizzato una Shiny App in cui è possibile visionare le statistiche dei vari giocatori, della stagione 2018-19, avendo anche la possibilità di vedere la distribuzione dei tiri realizzati da essi durante la stagione

Shiny applications not supported in static R Markdown documents