

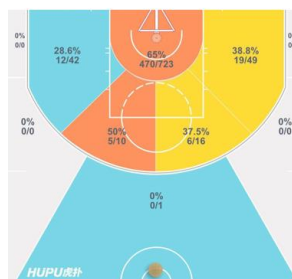
NBA 球员属性再定义

如果提及半个世纪以来最顶尖的中锋，天勾贾巴尔、大鲨鱼奥尼尔肯定是人们口中频率最高的名字之一，可是如果提及这十年来 NBA 最顶尖的中锋，人们可能感到错愕——什么？这个世代还有中锋吗？

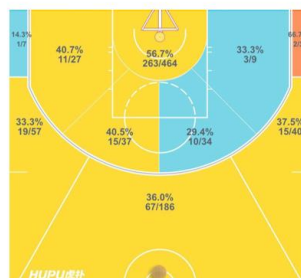
篮球运动员位置模糊化的说法在十年前就屡现纸端，只是在小球风暴盛行的当下暴露得更加明显。随着勇士王朝的建立，小、快、灵的小球战术席卷全联盟，再加上“魔球理论”（崇尚三分，摒弃低效的中远距离两分）的 Buff 加成，许多球员在做好本职工作的同时，还打破了传统位置属性的藩篱，同时具备两个及以上位置的攻守职能。

以中锋球员为例，有人称，姚鲨之后再无中锋。谈及姚明，我脑海里浮现的一幕是，阿尔斯通迅速将球推过半场后在三分区外盘旋，等待大吨位中锋姚明缓慢地回撤参与进攻。然而在今天，热衷防守反击所带来的攻击端提速将传统中锋置于尴尬境地，移动速度缓慢使得他们无法参与转换进攻和回撤防守。要么改变，要么消亡，这是现代篮球给每一位球员出的选择题，因此，许多后卫-前锋型、中锋-前锋型球员应运而生。

小考，我让你看看传统中锋的数据！



那在他们心里，我到底算什么？？？



接下来我们从数据描述方面来观察传统位置分类下的球员差异，本案例使用的是 2017~2018 赛季 NBA 常规赛数据集，该数据集记录了出场次数不低于 21 次的球员数据，包含了基本信息、基础指标和投篮指标三个维度，具体变量说明如表 1 所示。

表 1 数据说明表

变量类型	变量名	详细说明	取值范围	备注	变量类型	变量名	详细说明	取值范围	备注
基本信息	姓名	-	-	-	基础指标	进攻篮板	定量水平	0~5.1	-
	队伍	定性变量 (30水平)	-	NBA联盟共30支球队		助攻	定量水平	0.1~10.3	-
	位置	定性变量 (5水平)	SF、SG、PG、C、PF	-		失误	定量水平	0~5	-
	场次	定量变量	1~82	只取整数		抢断	定量水平	0.1~2.4	-
	胜场	定量变量	1~83	只取整数		盖帽	定量水平	0~2.6	-
	负场	定量变量	1~84	只取整数		个人犯规	定量水平	3.8~0.4	-
	出场时间	定量变量	0.5~37.8	-	投篮指标	突破得分	定量变量	0~11.5	-
	赛季	定性变量 (2水平)	2017~2018	只取整数		接球投篮得分	定量变量	0~11.9	-
基础指标	命中数	定量水平	0.2~10.4	-		急停跳投得分	定量变量	0~11	-
	出手数	定量水平	0.6~21.1	-		三秒区得分	定量变量	0~11.6	-
	命中率	定量水平	28.9~73.1	-		低位触球得分	定量变量	0~9	-
	防守篮板	定量水平	0.3~1	-		顶角得分	定量变量	0~4.3	-

● 球员基础指标

球员的基础指标囊括了球员攻防两端的表现，我们从中选取较有代表性的得分、助攻、防守篮板、命中率四个指标进行箱线图对比。

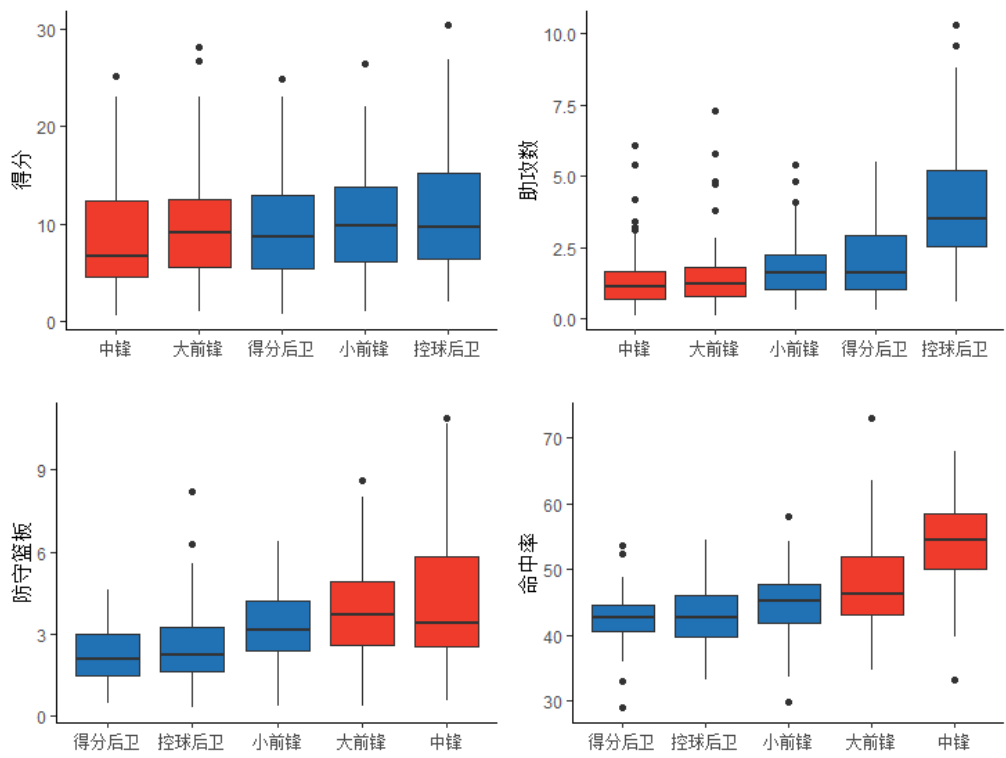


图 1 基础指标箱线图

得分数据是观众评价一个球员最直观的指标。以中位数计，中锋得分水平明显低于其他四个位置球员，控球后卫略胜一筹。中锋相较于其他位置球员而言，得分空间较小，短距离两分的得分效率远赶不上外线球员滂沱的三分雨。

助攻数意味着球员对于整个团队进攻的贡献程度，有些球员虽然得分不高，但是助攻数奇高，天赋异禀隆指导就是个中翘楚。从图上看，控球后卫的助攻水平遥遥领先，这与控球后卫的组织属性密切相关，观察比赛并在适当时候助攻队友得分，重剑无锋，大巧不工。



防守篮板体现了球员的防守能力，投篮是得分的手段，而篮板是投篮的前提。在防守篮板指标中，中锋和大前锋是毫无疑问的领头羊，毕竟强壮的身体优势和篮下有利空间使得他们比起后卫球员更加容易触及下落的篮球。

命中率，虎扑 JRS 们最常用的开会材料，哪怕得分较高，惨不忍睹的命中率还是会将球员送到 JRS 的刀子嘴下，比如，威少 Emmm……由于进攻位置多位于三秒区内，中锋和大前锋的命中率高与其他位置球员，“微微出汗”饼皇卡佩拉最有发言权，他所做的就是把自己手中的球轻轻送入篮框，而得分后卫囿于进攻空间较远的限制小幅落后。



● 球员投篮指标

投篮指标体现了球员的得分方式，突破、急停跳投、接球投篮、顶区跳投，不同位置的球员有着不同的投篮偏好。

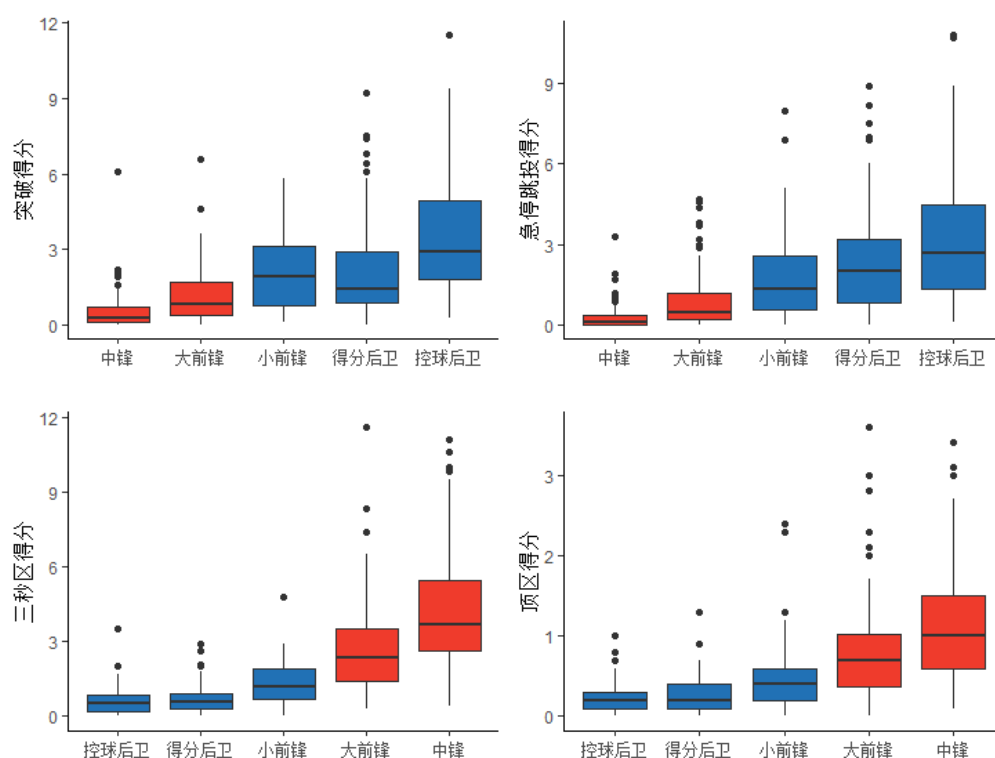


图 2 投篮指标箱线图

突破得分与急停跳投得分，这是后卫们最喜欢的得分方式，要么以迅雷之势孤军深入，要么手起刀落瞬间出击，直接且果决，不仅赏心悦目而且效果奇佳。而相较于灵活的后卫球员，锋线球员的得分则更需要耐心，三秒区的重锤细凿，顶区的不疾不徐，大个子们用自己的方式演绎另一种得分方式，不分上下，各有千秋。

从上面两幅统计图中，我们观察到大前锋和中锋、得分后卫和小前锋的数据指标分布趋同，出手方式相似，说明这两个位置的球员在球场上的作用可能极其相似或是重合，我们可以尝试用因子分析和聚类分析对球员位置属性进行再定义。

针对数据集里的 16 个球员技术指标进行综合评价，对技术指标进行了相关性分析，分析结果如图 3 所示。从图中明显观察到，部分变量的相关

性较强，比如突破得分和急停跳投得分、进攻篮板和盖帽等，这说明技术指标之间确实存在某种相关性，可以从中提取公共影响因子。

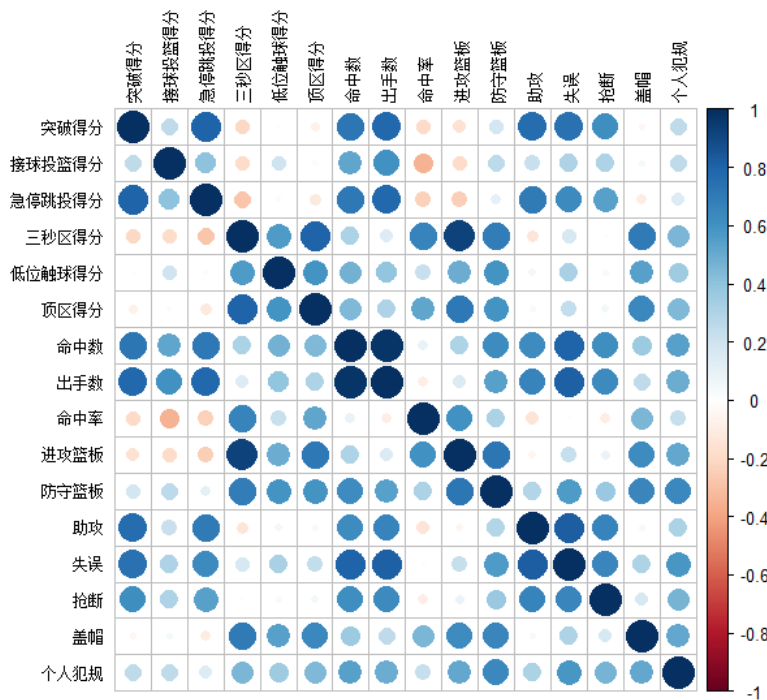


图 3 技术指标相关性

随后，本案例对 366 名球员的 16 个技术指标进行因子分析，最终保留了两个公共因子，累计方差贡献率达到 70.73%。从图 4 中，我们可以看到，突破得分、接球投篮得分、急停跳投得分、命中数、出手数、助攻、失误和抢断被归为同一类，其余指标自成一类，因此我们可以将两个公共因子分别命名为进攻因子和防守因子。

	进攻因子	防守因子	共性方差
突破得分	0.88	-0.13	0.8
接球投篮得分	0.55	-0.04	0.3
急停跳投得分	0.86	-0.21	0.79
命中数	0.86	0.42	0.92
出手数	0.93	0.25	0.92
助攻	0.84	-0.04	0.71
失误	0.87	0.3	0.84
抢断	0.76	0.1	0.58
三秒区得分	-0.12	0.95	0.92
低位触球得分	0.18	0.67	0.48
顶区得分	0.03	0.85	0.73
命中率	-0.25	0.66	0.5
进攻篮板	-0.07	0.91	0.83
防守篮板	0.39	0.79	0.78
盖帽	0.08	0.81	0.66
个人犯规	0.44	0.6	0.55

图 4 因子载荷矩阵

进攻和防守，是篮球中最基本的两方面内容，根据因子得分我们给出了在进攻因子和防守因子的上得分最高的 30 名球员，他们分别代表了联盟在进攻端和防守端表现最出色的球员。

进攻端上，新科 MVP 詹姆斯-哈登毫无争议地高居榜首，得分机器威少紧随其后，“假期苦练” 奥拉迪波和利拉德也获得了较好的排名，而库里和欧文由于伤病的影响排名与名气略有出入。



图 5 进攻因子 TOP30

防守端上，浓眉哥和庄神平分秋色，护框狂魔亚当斯凭借一手好板占据一席之地，引人注目的是字母哥扬尼斯-阿德托昆博，在进攻端和防守端的表现都名列前茅，天赋爆表的身体优势让他得以在新生代球员中占据领袖位置。

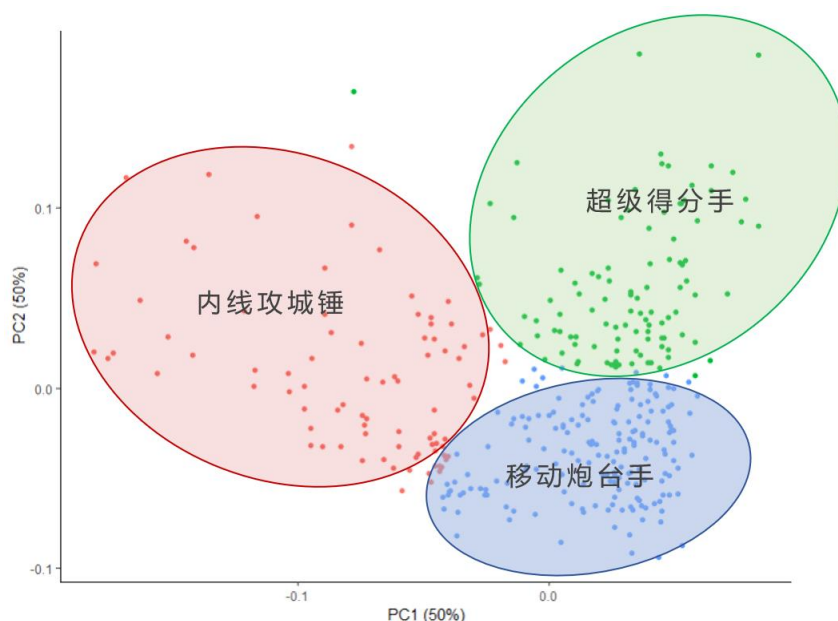


图 6 防守因子 TOP30

“我认为场上不再有五个位置，可能现在可以简化成三个位置，处理球的人、侧翼以及大个子。这很重要，我们的球员这些年来变得越来越全面了。”

——凯尔特人主帅史蒂文斯

进行因子分析后，我们尝试用 K-Means 聚类对球员进行分类，结果显示球员被分成了三个类别：超级得分手、移动炮台手和内线攻城锤。

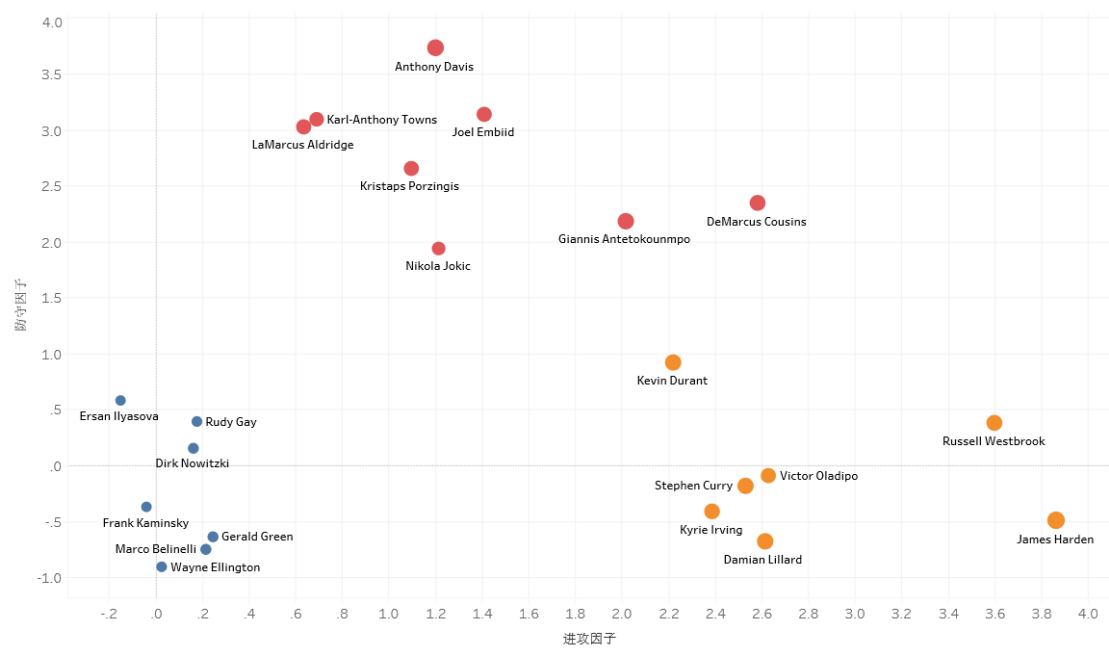


从球员新旧位置对比中看出，超级得分手包括了五个位置球员，得分后卫居多，长于得分擅于组织，场均得分与助攻数最高。超级得分手中盛产得分机器，根据上赛季的数据，詹姆斯-哈登是联盟中进攻杀伤性最高的球员，外有后撤步三分，内有欧洲步上篮，实在不行了还能跟卡佩拉内外连线着实令人头疼。而虎扑帅哥杜兰特则是攻防俱佳的典型，媲美中锋的超长臂展和站立摸高使得他在防守端也不落下风，一手无解中投也是将对手安排得明明白白。

内线攻城锤，顾名思义主攻内线，包括大前锋、小前锋和中锋等锋线球员，以防守为主，篮板数和三秒区得分独占鳌头。从图上看，浓眉哥、大帝和考辛斯是这类球员中较为突出的典型，除了在内线翻江倒海，他们在三分线外还有一手好篮，大帝更是被称为会投三分的奥拉朱旺，运动能力强、有持球能力和三分能力，这些新时代的大个子无疑是现代篮球的宠儿。

移动炮台手，包括除中锋以外的所有位置球员，此类球员进攻手段较为

单一，缺乏较强的自主进攻能力，多靠跑位赢得空间队友助攻投篮得分，因而场均得分较低。休斯顿火箭是联盟中最适合移动炮台手的球队体系之一，拥有两大全明星控卫，哈登具有极佳的传球技能，而保罗从来不是个人进攻优先的后卫，传球助攻是他的习惯，因此火箭队可以在外线埋伏两到三名移动炮台手，比如杰拉德-格林、莱昂-安德森、特雷沃-阿里扎甚至是PJ-塔克。这些移动炮台手仿佛一颗颗定时炸弹，他们并不占球权，出手极快并且有高效的三分投射能力，虽然没有力挽狂澜的壮举，但是却创造出了极致的空间环境，让对手顾此失彼。



一个内线攻城拔寨的大个子，两个里突外投的超级得分手，再加上两颗埋伏在外线的定时炸弹，看起来是一支球队的完美阵容。但是这仅仅是从数据方面考量，球场上的技战术演绎和 X 因素的存在使得每一场比赛结果都充满了不确定性，这也是篮球的魅力所在。

本次案例虽然对球员有了一个初步的聚类，但是并未运用胜利贡献值（WS）、霍林格球员效率指数（PER）、真实胜负值（+/-）等高阶数据，若能够多加运用，能使聚类结果更加精确。除此以外，聚类结果中出现了一些极端异常值，如何处理区分此类异常球员有待进一步思考。