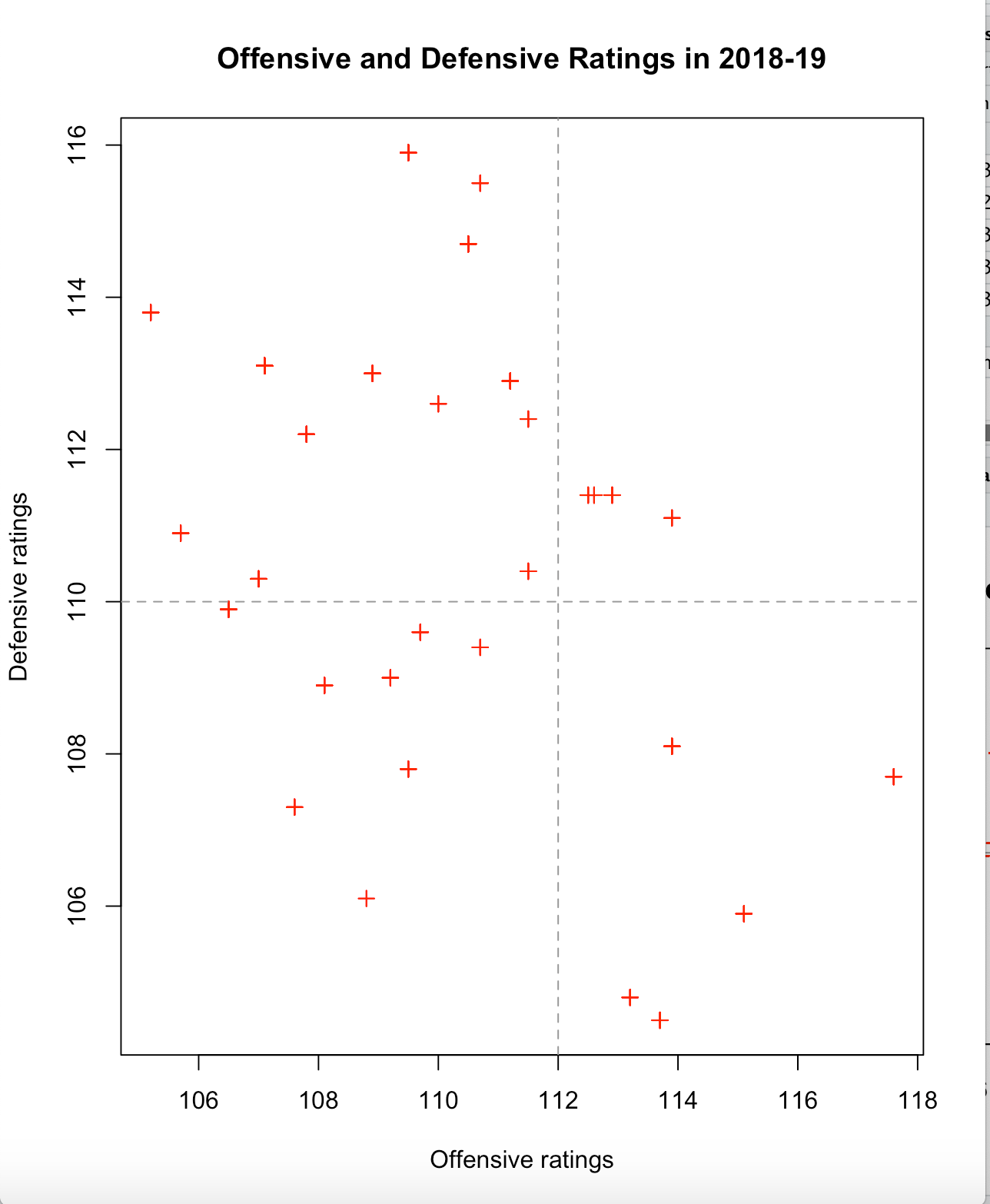
数据说明：（本文数据收据每个球员和球队每100次控球内的所有基础数据。）

1.首先考虑个体球员数据：

Possession控球：一次控球开始于某个球队的球员拿到球权，开始进攻，结束于该队的球员失去了球权。失去球权的方式包括（1）该球员投篮命中，球权转换的对方球员；（2）该队球员防守失败，球权转换（3）该队球员进攻犯规，球权转换。在一场比赛中两队控球机会和时长基本持平，为了赢得比赛，球队应该在每场比赛特定的控球机会中寻得更高的得分。所以根据球员在控球时间内的得分可以很好的评判球队的整体表现。

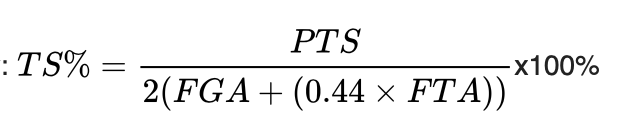
Possession = 0.5 \* (([Tm](https://www.basketball-reference.com/about/glossary.html#team) [FGA](https://www.basketball-reference.com/about/glossary.html#fga) + 0.4 \* [Tm](https://www.basketball-reference.com/about/glossary.html#team) [FTA](https://www.basketball-reference.com/about/glossary.html#fta) - 1.07 \* ([Tm](https://www.basketball-reference.com/about/glossary.html#team) [ORB](https://www.basketball-reference.com/about/glossary.html#orb) / ([Tm](https://www.basketball-reference.com/about/glossary.html#team) [ORB](https://www.basketball-reference.com/about/glossary.html#orb) + [Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id) [DRB](https://www.basketball-reference.com/about/glossary.html#drb))) \* ([Tm](https://www.basketball-reference.com/about/glossary.html#team) [FGA](https://www.basketball-reference.com/about/glossary.html#fga) - [Tm](https://www.basketball-reference.com/about/glossary.html#team) [FG](https://www.basketball-reference.com/about/glossary.html#fg)) + [Tm](https://www.basketball-reference.com/about/glossary.html#team) [TOV](https://www.basketball-reference.com/about/glossary.html#tov)) + ([Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id) [FGA](https://www.basketball-reference.com/about/glossary.html#fga) + 0.4 \* [Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id)[FTA](https://www.basketball-reference.com/about/glossary.html#fta) - 1.07 \* ([Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id) [ORB](https://www.basketball-reference.com/about/glossary.html#orb) / ([Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id) [ORB](https://www.basketball-reference.com/about/glossary.html#orb) + [Tm](https://www.basketball-reference.com/about/glossary.html#team) [DRB](https://www.basketball-reference.com/about/glossary.html#drb))) \* ([Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id) [FGA](https://www.basketball-reference.com/about/glossary.html#fga) - [Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id) [FG](https://www.basketball-reference.com/about/glossary.html#fg)) + [Opp](https://www.basketball-reference.com/about/glossary.html" \l "opp_id) [TOV](https://www.basketball-reference.com/about/glossary.html#tov)))

Offensive and Defensive Ratings进攻得分和防守失分率: 率代表每次控球的得分（失分），进攻得分率和防守失分率是根据每100次控球的总得分或者对手球队得分。根据分别考虑进攻得分和防守失分可以全面的评判一个球队的能力和长处，而不是将目标只局限于得分高低。下图展示了2018-19赛季中顶级30个球队得分率和失分率的分布图。下边的球队防守能力更强，因为对方球队的得分较低；右边的球队进攻能力很强，因为进攻得分很高。



Pace Adjustment 进攻效率调整：一个获得100次球权的 球队，比一个获得80次进攻机会的球队多25%的机会进行投篮，助攻，篮板等，所以当两个球队有不同的控球次数时，需要调整获得的数据。不同的控球次数来源于球队不同的进攻风格，若一个球队的步伐相对缓慢，称为有更少的进攻机会。根据调整球队进攻步伐获得更统一的衡量标准。

True Shooting Percentage实际命中率: 相对于衡量一个球员所有的投篮命中率，实际命中率更全面涵盖一个球员通过投篮得到所有的分数，包括二分球，三分球和罚球得分，该指标可以很客观的评判运动员的命中率高低，从而衡量一个球员的命中率对球队的贡献。

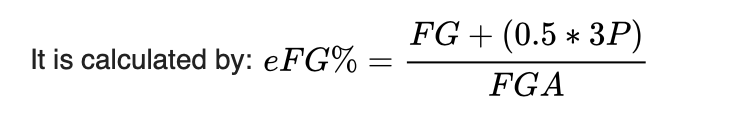


True shooting Rate = PTS/2(FGA+0.44\*FTA)

(用中文打公式)

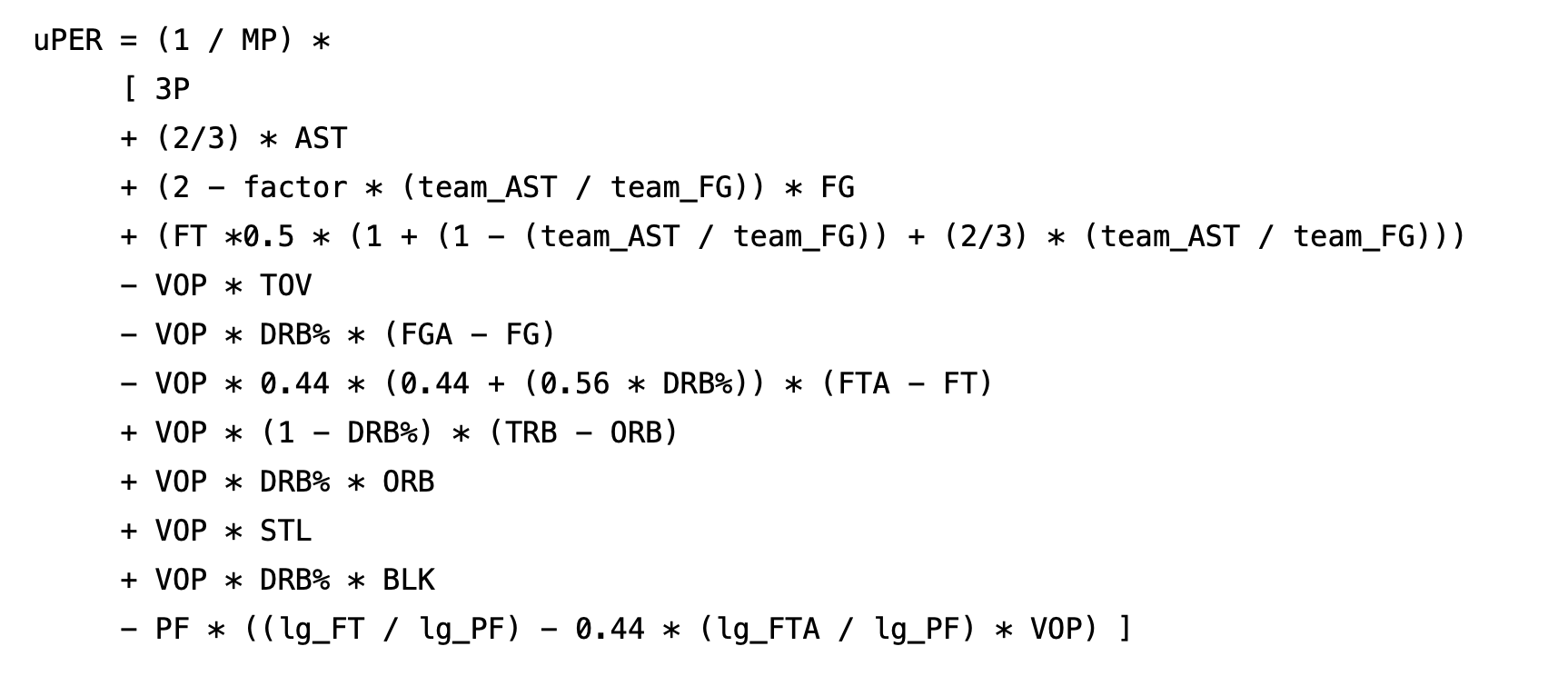
注释：系数0.44是根据一个球员每100次控球中2分球和3分球的出手比例计算而来的。如果100次控球中所有的投篮都是2分球，该系数应该是0.5，如果100次控球中所有投篮出手都是3分球，该系数为0.333.

Effective Field Goal Percentage有效得分率：衡量命中率之后需要考虑每次命中的实际得分，由于3分球出手命中后得3分，2分球出手后命中得2分，不能用统一的标准来计算所有命中的得分。计算有效得分率，可以更好的衡量球员得分对整体球队的贡献。



Effective Field Goal Percentage= FG+(0.5\*3P)/FGA

PER：衡量球员效率的指标Per Efficiency Rating. 指标根据NBA官网公开的各球员、球队、联盟的基础数据计算得出。总结球员对球队的正向贡献和负向贡献，即在衡量球员得分的指标前赋予一个正的系数，在衡量球员失分的指标前赋予一个负的系数，再将所有加权后的指标相加，再除以根据球队和联盟数据构建的指标对球员的数据进行规范化。PER最终得到球员每分钟的贡献率，该指标不应该受到球员上场时间，球队进攻类型，整个联盟赛制变化的影响。该指标客观有效的衡量球员的比赛效率，可以成为衡量球员对球队贡献指标之一。



Vop替换成alpha, Factor替换成beta，DRB替换成gamma

factor = (2 / 3) - (0.5 \* (lg\_AST / lg\_FG)) / (2 \* (lg\_FG / lg\_FT))

VOP = lg\_PTS / (lg\_FGA - lg\_ORB + lg\_TOV + 0.44 \* lg\_FTA)

DRB% = (lg\_TRB - lg\_ORB) / lg\_TRB

这三个参数分别代表联盟的(factor)，联盟的

注释：由于在大多数开源的数据中更多记录了一个球员进攻表现（更容易记录）如进攻篮板球，三分球得分，罚球得分等；记录球员防守表现的数据更少（不容易被记录）。所以该指标对一个防守很强的球员更有倾向。在实际应用中，两个同样优秀的球员，其一打球风格更激进，其二防守表现突出，可能PER在得出第一位球员更优秀的结论。

Usage percentage利用效率：是衡量每个球员在每一次掌握球权时对球权的利用率。根据射门次数、罚球次数、平局次数或失误次数等数据构建出该指标。通常控球次数多的球员有更高的利用效率，因此一般个子大的球员对球的控制能力更强，打出更高的利用效率。利用效率越高，相对球员效率PER越低。优秀的球员可以在最少的控球次数中打出最高的得分。

Usage percentage = 100 \* (([FGA](https://www.basketball-reference.com/about/glossary.html#fga) + 0.44 \* [FTA](https://www.basketball-reference.com/about/glossary.html#fta) + [TOV](https://www.basketball-reference.com/about/glossary.html#tov)) \* ([Tm](https://www.basketball-reference.com/about/glossary.html#team) [MP](https://www.basketball-reference.com/about/glossary.html#mp) / 5)) / ([MP](https://www.basketball-reference.com/about/glossary.html#mp) \* ([Tm](https://www.basketball-reference.com/about/glossary.html#team) [FGA](https://www.basketball-reference.com/about/glossary.html#fga) + 0.44 \* [Tm](https://www.basketball-reference.com/about/glossary.html#team) [FTA](https://www.basketball-reference.com/about/glossary.html#fta) + [Tm](https://www.basketball-reference.com/about/glossary.html#team) [TOV](https://www.basketball-reference.com/about/glossary.html#tov)))

2.球队指标

a. 根据球员指标计算球队指标：

按照每支球队整个赛季中各个球员在场时间将球员排序

将每个球员的PER和上场时间相乘

将每队上场时间最长的12个球员的上述指标与上场时间相乘后的指标相加得到

该计算方法的意义：

1. 由于PER是一个以分钟为单位的指标，通过与球员的上场时长相乘可以估计出球员在整个赛季中对球队贡献。
2. 一些球员通过选秀等方式会被交易到其他球队，所以要计算球员上场时间最长的12个球员，更充分可以反映球员对球队的贡献。因为这12个球员是球队中比较稳定的中流砥柱.
3. 该计算方法也可以自动给PER更高的球员赋予更大的权重，因为PER更高的球员上场时间更长，而PER较低的球员的上场时间会相对较短，所以时长作为权重，可以很好的平衡球员之间的效率。就算是某个PER很高的球员在赛季中受伤，他的时长也会限制他对球队的贡献。如史蒂芬库里在19-20赛季中因伤暂停比赛半年，尽管他是一个价值很高的球员，他的上场时长受到影响，他对球队的贡献

b.球队自身数据：

* Team\_Rank：球队排名，依据球队赢得比赛的次数，与球队整个赛季中各个方面的表现得到的整体排名。排名综合考虑球队的进攻，防守，总得分和整体数据得出。
* Adjust\_MoV：Margin of Victory 比分差距是用来衡量比赛中获胜队伍和失败队伍比赛分数差距的统计量，参考MOV可以快速看出比赛是否激烈，获胜队伍胜利的是否显著，大的比分差距代表比赛中获胜方在比赛中强势压制对手，而小的比分差距说明两个球队水平接近，比赛相当激烈。在本数据集中，MOV指标是取一个队在整个赛季比赛中球队MOV指标的平均值。如果一个球队的Margin of Victory指标是正数，说明该球队在整个赛季中表现更好。然而在比赛中由于每个球队的进攻风格不同，需要按照进攻强度来调整对应的MOV值。尤其是不同的队伍会根据遇到的对手球队调整进攻步伐和阵容等，所以我们利用更好反映球队强弱的Adjust\_Mov数据来建模。历史上最大的比分差是68分，是1992年克利夫兰骑士队和迈阿密热火队的比赛，当时骑士队是最强的球队之一。
* Adjust\_NRtg：如果只单纯根据Offensive Rating和Defensive Rating来衡量一个球队的优秀程度多少有一些偏差，这时需要引入球队净得分来更全面的衡量。Net Rating是球队每100次进攻中得分与失分的差。用来衡量球队风格，和更全面的评判球队的表现。一个球队如果在整个赛季中净得分大于0，说明进攻性更强。当然球队的净得分和球员的表现有很大的关系。球队的净得分均值是0.
* Win\_ratio:在整个赛季的比赛中，赢得的比赛与参加的比赛之比作为胜率。胜率对衡量球队在下一个赛季的表现有着很重要的作用。可以帮助教练调整训练节奏和步伐，帮助球员在选秀中找到一个更合适自己的队伍，可以帮助粉丝们在赌球中有一个更好的判断。胜率的平均值在0.5.