# Understand Basketball Gambling through Data

As the NBA shifts towards capitalizing on the recent sports betting legalization boom, I thought it would be a good time to take a deep-dive into the data of sports betting. [Discuss history of sports betting in NBA] My aim is not to discuss the inefficiencies in Vegas lines or build a comprehensive betting model, but instead to i) visualize historical betting odds to help convey the mechanics of sports gambling, ii) retrospectively show where and when Vegas has been wrong, iii) translate sports gambling outcomes to how we watch basketball.

Before we look at the data, let’s review the vocabulary needed for betting on sports, and basketball more specifically. There are three main types of bets that can be made on any matchup:[[1]](#footnote-1)

* **Moneyline:** Simply, who will **win**?
* **Spread:** How many **points** will the favorite win by?
* **Total:** How many **total points** will be scored in the game?

Vegas sportsbooks also use two tools to encourage betting and minimize their own risk:

* **Odds:** The payout given a certain bet. Bet payouts are often expressed as “American odds” or as “Decimal Odds.” See below for decimal and American odds given different stakes and payouts. Note that Decimal odds multiplied by the stake will be the total payout, including the original stake, while the American odds are expressed to only include profit.**[[2]](#footnote-2)**

****

* **Lines:** For Spread and Total bets, the bettor will bet above or below some number called the “Line.”
  1. For Spreads, the line is the margin of difference between the two teams determined by Vegas, with a minus (“-“) representing a favorite and a plus (“+”) representing the underdog.
  2. For Totals, the line is the total number of expected points of both teams combined, which bettors can bet either “Over” or “Under” that total. Below is an example of betting odds for Moneyline, Spread, and Total bets:

****

Within each of the three betting types above, there are mechanisms Vegas sportsbooks use to encourage betting on both sides to minimize risk of an upset and maximize their own profit.[[3]](#footnote-3) Before the game starts, Moneyline bets can adjust payouts, or **odds**, to encourage betting. Spreads and Totals can also adjust odds, but more often adjust the **lines** to encourage or discourage betting. For spreads, a line adjustment would change the amount the favorite will win by. For Totals, a line adjustment would change the total amount of points in the game.

Below is a hypothetical showing how moneyline odds change prior to the beginning of a hypothetical game between the Lakers and the Nuggets. This table shows that Vegas changes the lines to maximize their return for either the scenario in which the Lakers win or the Nuggets win:[[4]](#footnote-4)



Note that the Expected Value column assumes that Vegas expects the Lakers have a 65% change to win (while the nuggets have a 35% chance to win). This example speaks to an important artifact of sportsbook: **different bettors can get locked into different odds.** In that sense, **timing is everything** in securing a bet with good odds. If you had a strong expectation that the Lakers would win against the nuggets above, you would far prefer to get odds at -150 than -180, so bettors often monitor odds throughout the day.

Below is a similar example that looks at betting on the spread instead of the moneyline. Notice that there are two mechanics changing now: the odds and the line.

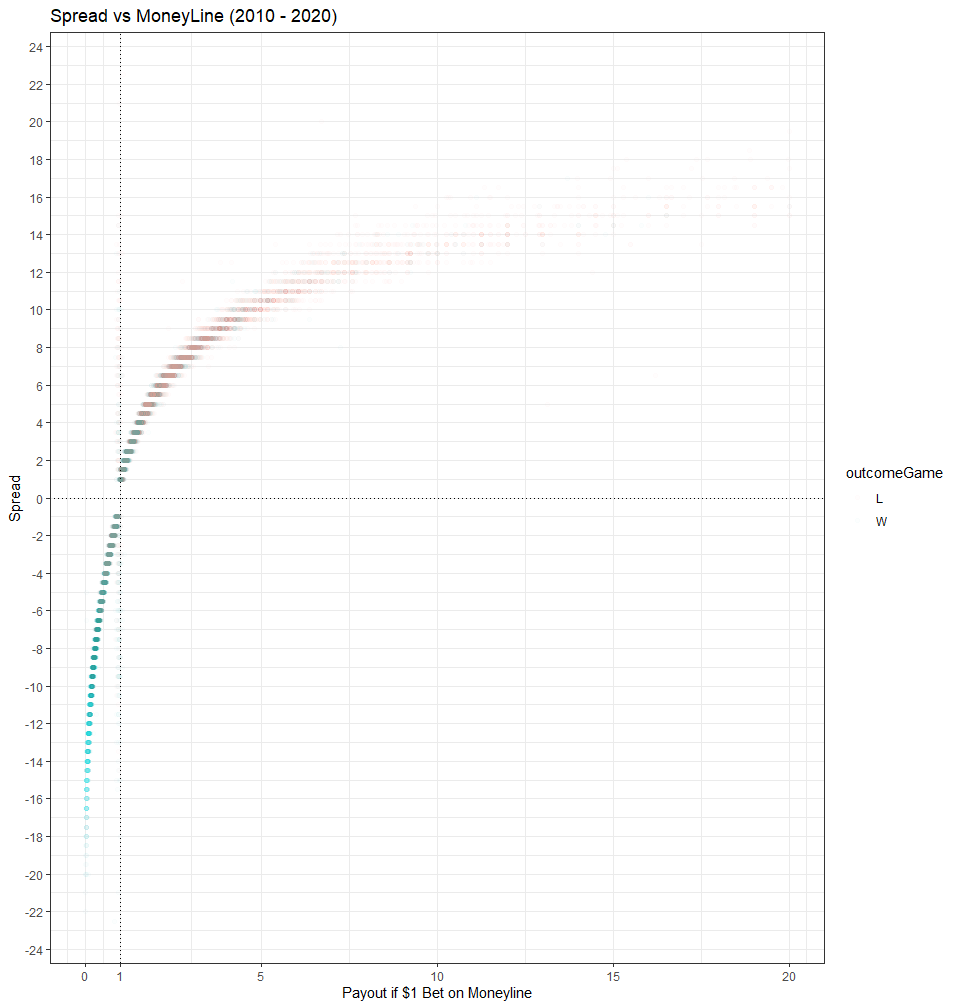


Below is a short summary of the Bet Types, Definitions, Examples and Tools

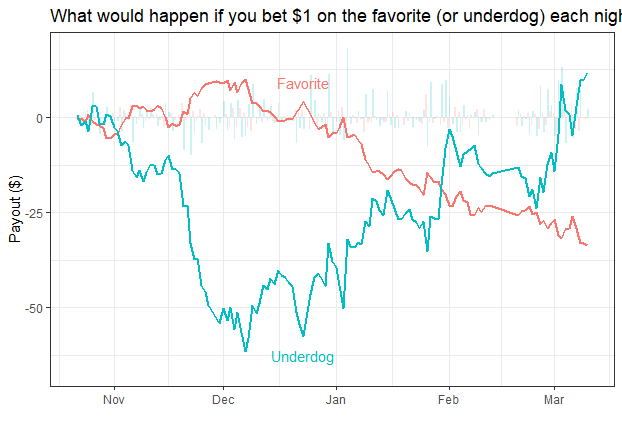


## The Spread and the Moneyline

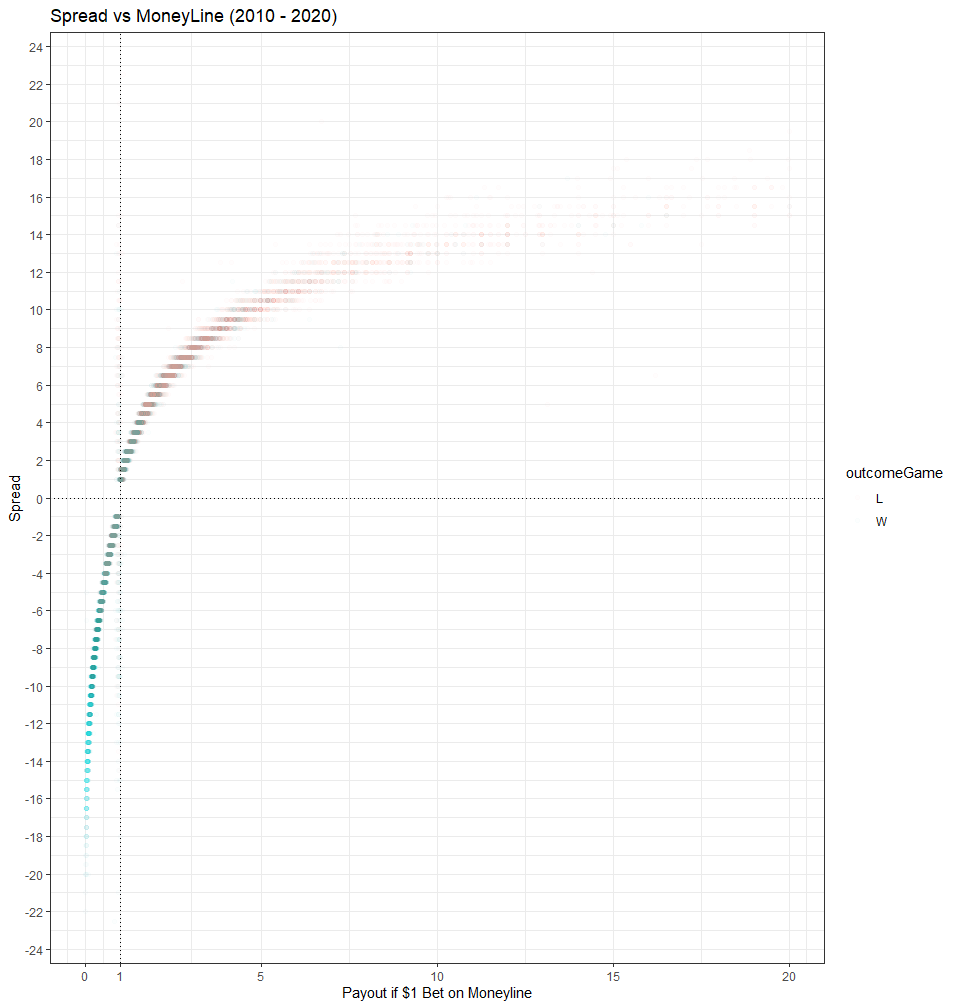
One thing I wondered when I was learning about these concepts was how the lines, specifically the spread and the moneyline, reflected each other. Below is a graph of the moneyline odds versus the spread for every game over the last 10 years.



The graph above shows that a ton of games are between -3 and +3.



* Above is the payout if you put $1 on all Favorites (Red) or all Underdogs (blue)
* As probably expected, when the underdogs hit, payouts can be quite a bit larger than when favorites hit



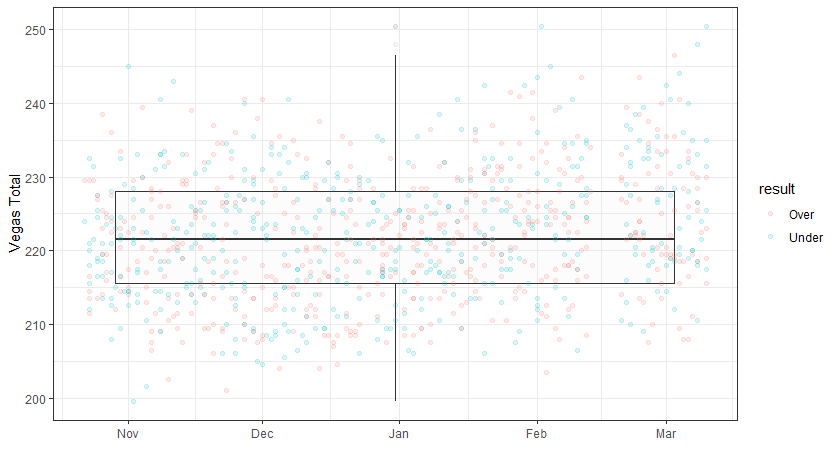
Above the is the relationship over the past 10 years between the Moneyline odds and the spread. Bet returns are often expressed as “American odds,” where, for example, +150 (Bet $100 to win $150) or -110 (Bet $110 to win $100). Instead, I like expressing odds as a decimal, where +150 would be 1.50 (150/100) and -110 would be 0.91 (100/110). Another way to think about decimal odds is the profit from betting $1 on a game. The reason I like odds expressed as a decimal is that it’s easier to compare underdogs and favorites on the same scale. While many [outlets](https://www.investopedia.com/articles/investing/042115/betting-basics-fractional-decimal-american-moneyline-odds.asp) include the initial bet in the decimal (+150 would be 2.5 instead of my 1.5), I think it’s a bit cleaner to just look at profit.

There is one basic rule of separating the favorites from the underdogs in the graph above: a team is a favorite if their moneyline odds are less than 1 (see bottom left quadrant) and are underdogs if their moneyline odds are greater than 1 (top right quadrant). Favorites with moneyline odds less than 1 have spreads that are **negative,** meaning they are favored to win a game by some number of points. Underdogs (positive odds) have positive spreads, meaning they are expected to lose my some number of points.

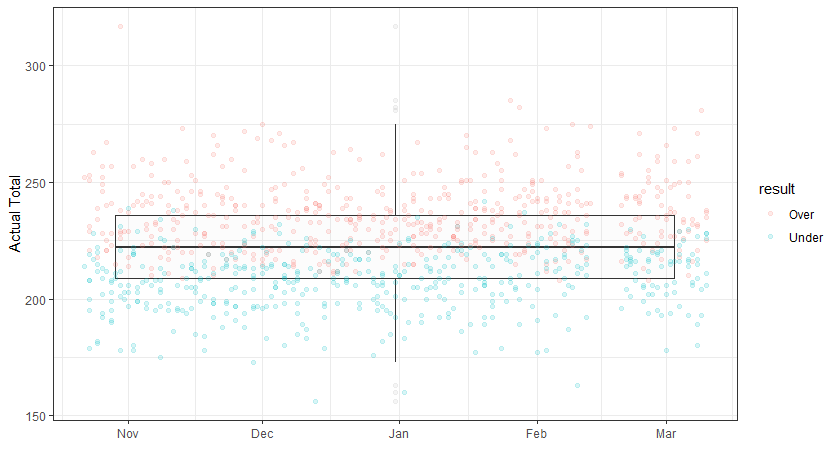
The relationship between moneyline odds and spreads appear to be a logarithmic relationship. Note that red dots are wins and blue dots are losses. [Explain more on what I think is going on. There has been a lot of consistency between the years]

**Totals**

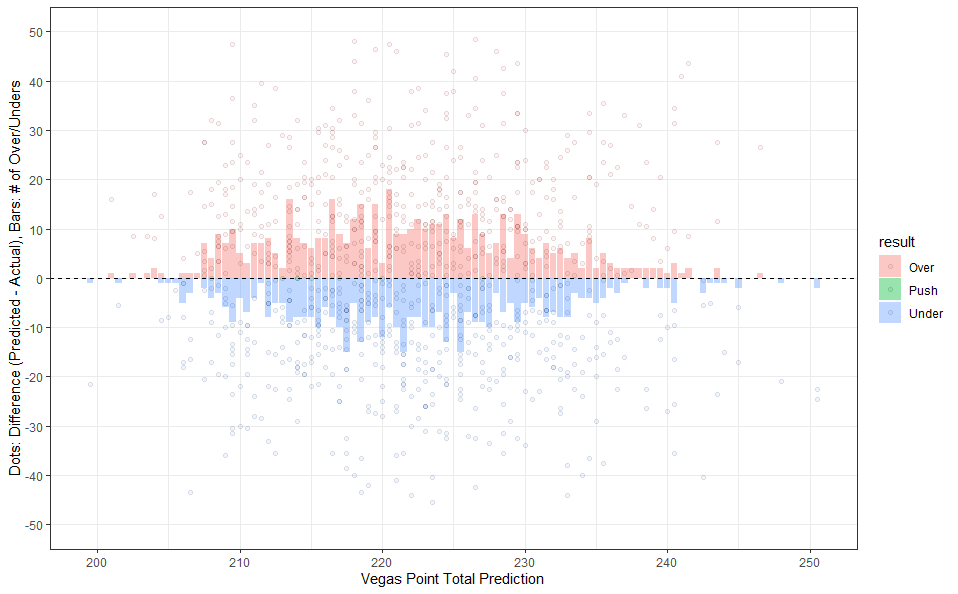
Here is what the total lines look like so far this year (along with the boxplot showing median, quartiles, and 95. The median predicted score appears to be around 221.



Here is what the actual totals look like over time. The median actual total appears to be around 224, very close to the sbr total prediction.



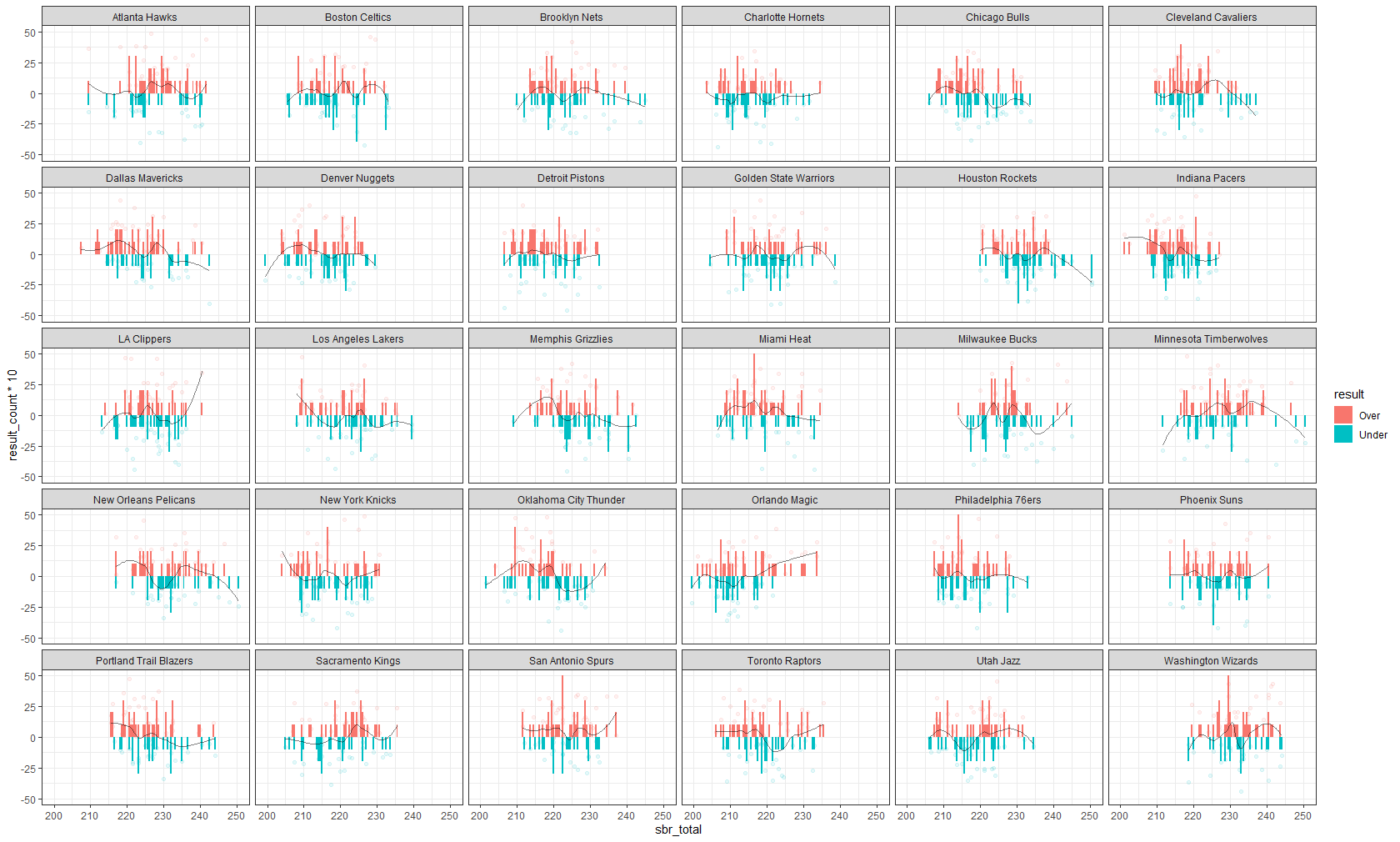
Let’s bring the two together to see if there is any Vegas Point Total that appears to be more vulnerable (or less vulnerable) to going over (or under):



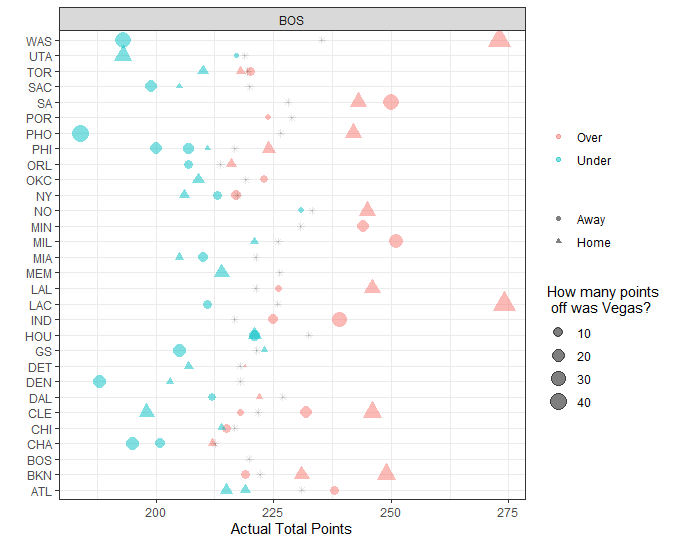
Here, the x axis shows the different Vegas Point Total lines before games started. It looks like most lines are set between 210 and 240 points (confirming the previous graph). The Y axis here tells us two things: i) **bars** tells us the number of overs (red bars) or unders (blue bars) at a given point total prediction, and ii) **points** tell us the difference between expectation and reality individual games (**points**). For the times where the expected total is equal to the actual total, there is a green point on the dotted line.) The sum of each individual red point is the total number of overs and the sum of each individual blue point is the total number of unders. [make some notes on the observations here. Maybe discuss how this has changed over time.]

To make this concrete, it looks like when Vegas has predicted totals of 220 points, there have been 3 overs, 12 unders, and one push. The largest over was 23 points over 220 (243 points) whereas the largest under was 28 points under 220 (192 points).

What does this look like by team? Let’s take a look:



This is a lot to look at, and doesn’t tell me anything about home vs away. Let’s look at the Celtics specifically in a different visualization:



It looks like the Celtics have generally played teams below their average.

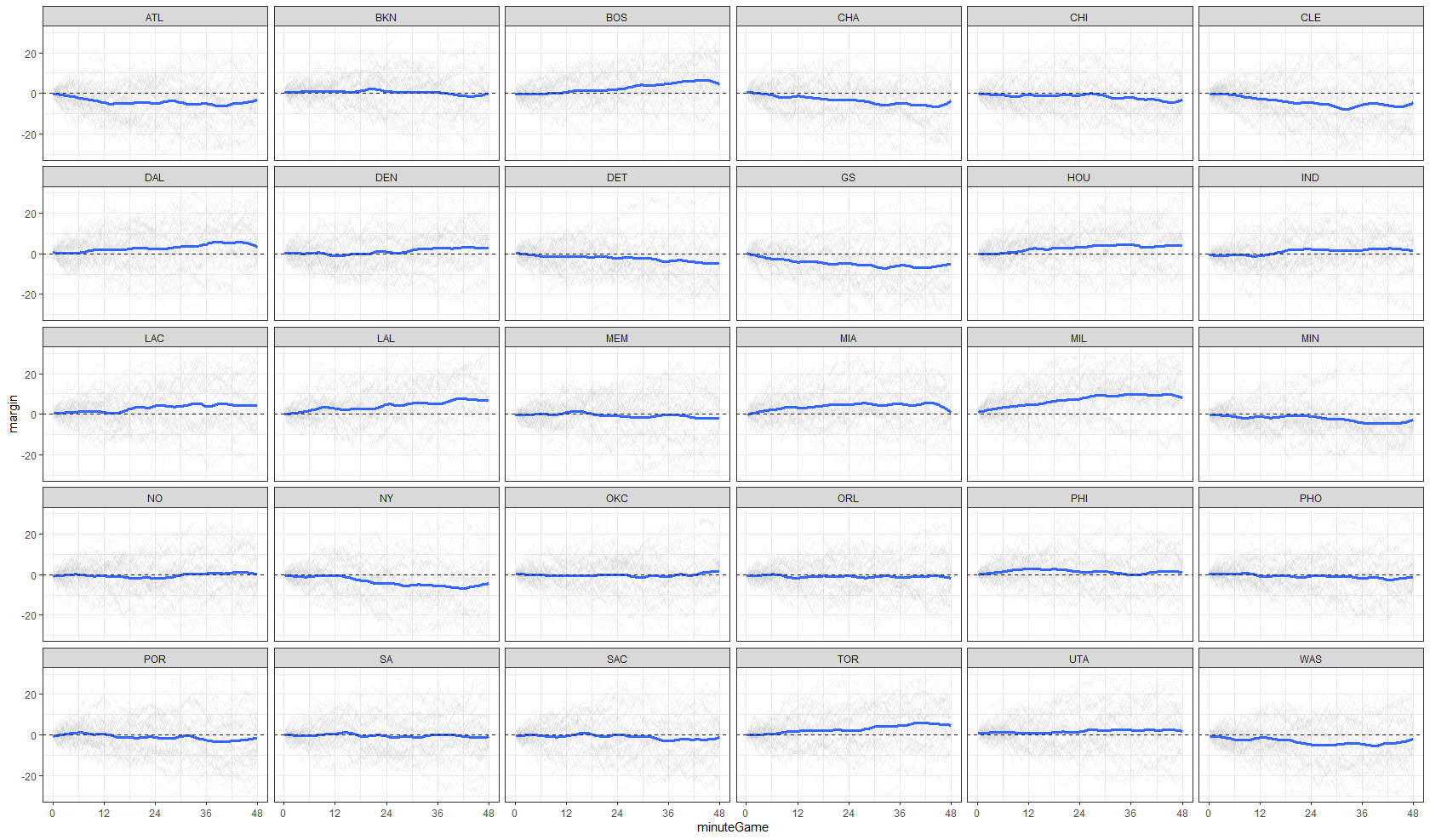
**Spread**

**Moneyline**

It would be really cool to see the same graph above ( the bar/point chart except the x axis would be ML Ratio. I think the Y Axis would stay the same. It probably makes sense to bin the ML Ratio (-100 to -120, -121-150, 151- 200, -200 – 300, -300 - -400, -400-500

**Live Betting Observations**

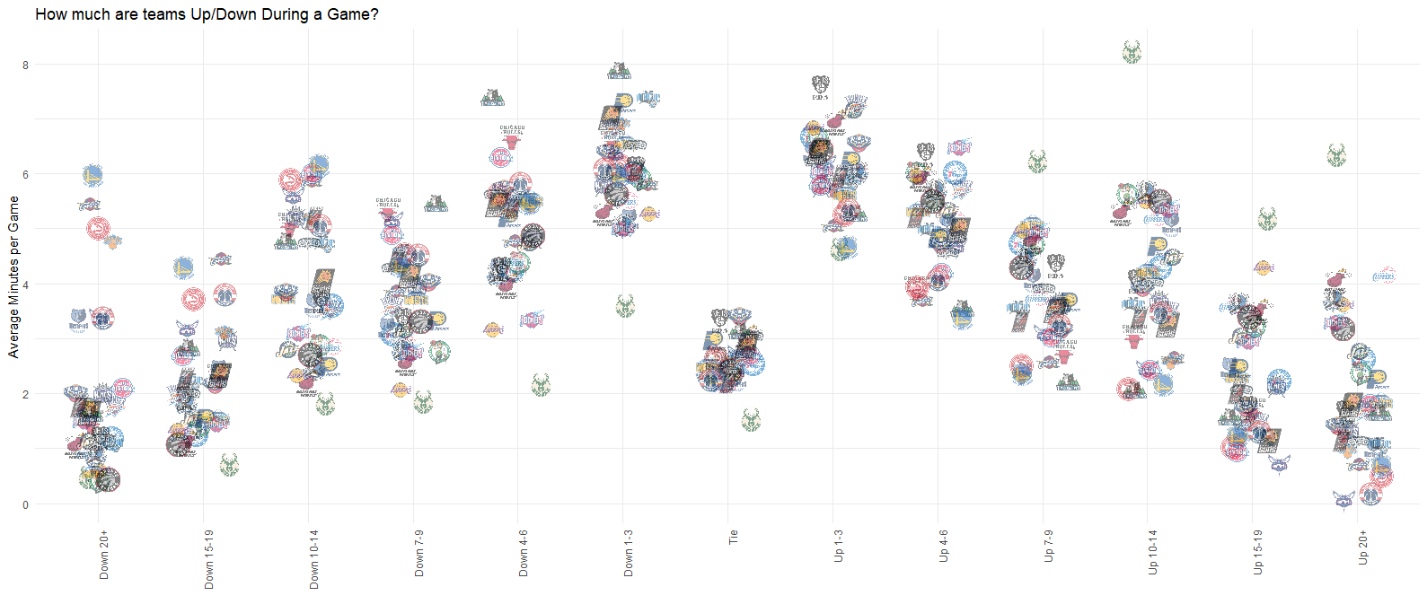
Margins for teams can fluctuate greatly. Even when a team comes in as a heavy favorite, Vegas will sometimes react to large lead changes and adjust live lines such that they become further apart from the original line. How can we take advantage? Let’s first look at how the margin, or lead, for each team changes through the game:

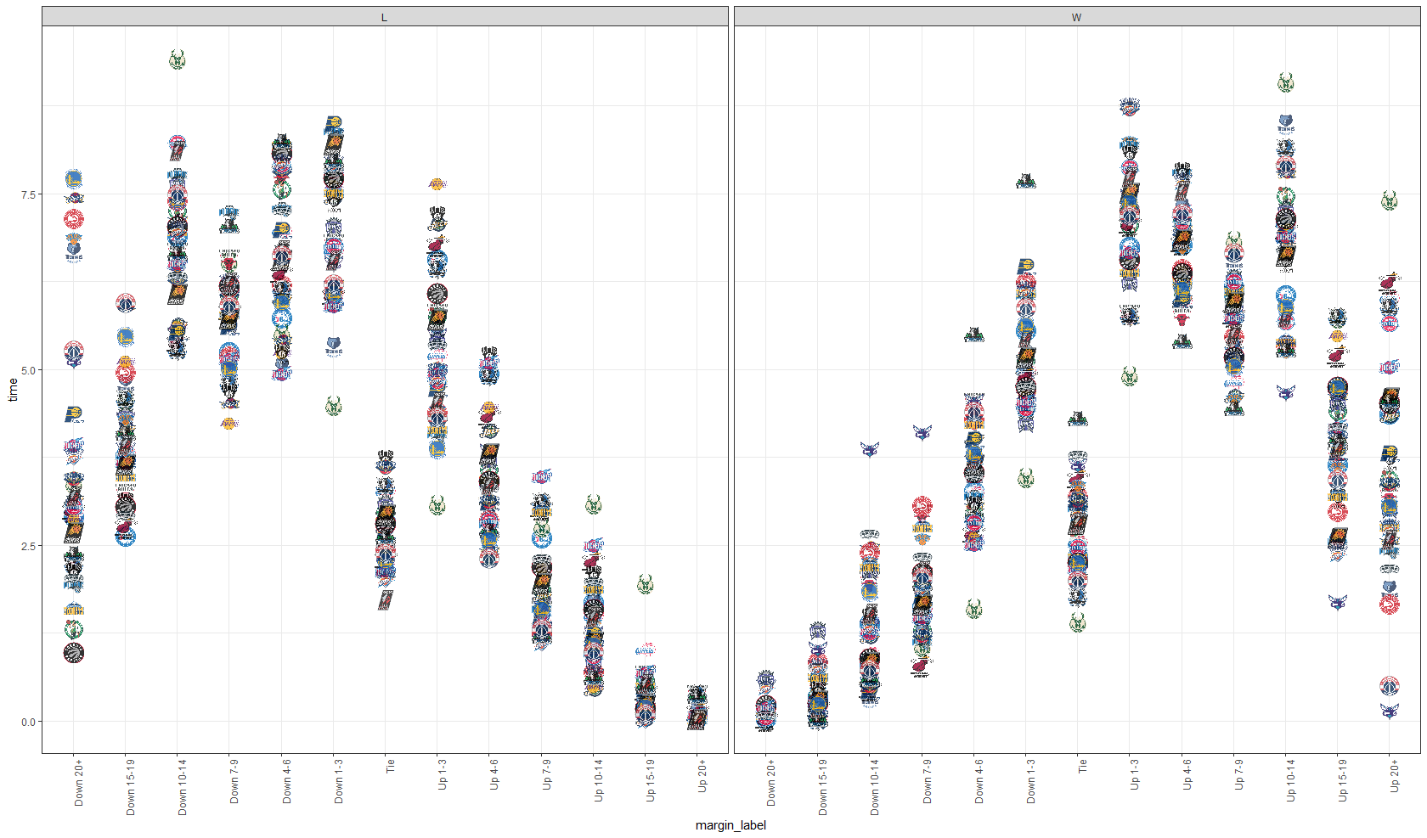


How does this look for Wins vs losses?



How much times do team spend down/up a certain number of points in wins and losses?





It might be cool to know what the % of the game (amount of time / 48 mins) winning by a certain amount (or losing by a certain amount) means for hitting the line.

If the rockets are -9 and they were up by 9+ points 45% of the time in the first 3 quarters, how often do they win?

Maybe a histogram of Point proportions by team on wins and on losses (I guess average?). Bins: Tie. 1-3 points (1 possession). 4-6 points (2 possessions). 7-9 points (3 possessionss). 10-12 points (4 posessions). 13+ (5+ possessions).

**Headers:**

Today’s Line

Team Comparisons [control which teams you’re looking at, the period of time, etc.]

Live Betting and Game Fluctuations

Season

Comparisons

1. For the purposes of this discussion, we will ignore more complicated ways of combining multiple bets (e.g., parlays and teasers), shifting lines through “buying” points, betting on individual bets (“prop” bets), or betting on future events (“Futures”). [↑](#footnote-ref-1)
2. **American Odds Formula** =

   For Stake < Profit: 100 \* Stake/Profit.

   For Stake > Profit: 100 \* Profit/Stake.

   **Decimal Odds Formula** = Profit/Stake + 1. [↑](#footnote-ref-2)
3. Let’s say the best team in the league is playing the worst in the team in the league. What stops the potential bettors from all taking the heavy favorite? For the Moneyline, there is only one way to encourage/discourage betting on a given team: payout or “odds”. [↑](#footnote-ref-3)
4. Note – this is an oversimplified example that ignores any additional fees or taxes charged by the sportsbook. [↑](#footnote-ref-4)