# Who’s #1? The Science of Rating and Ranking

John Hamilton

4 May 2023

# POP QUIZ: Who won the 2022-2023 Superbowl?

# POP QUIZ: Who was the best team in the 2022-2023 season?

# Summary

**Starting Point**: there are many different methods of rating and ranking, each of which have their own theory and assumptions. In contrast to what our commercial stakeholders would have us believe, there is no “correct” way of determining value, worth, quality, etc.

**Goal**: to dive into the theory a bit to instill some thought behind the assumptions that we make

# Massey Method | Theory

Theory- the difference in ratings of two teams is expressed through some sort of score differential

$$r_i - r_j = p_{ij}$$

$$Mr = p$$

M: Massey Matrix

r: ratings vector

P: score differential vector

Solved using any method you wish (ie least squares)

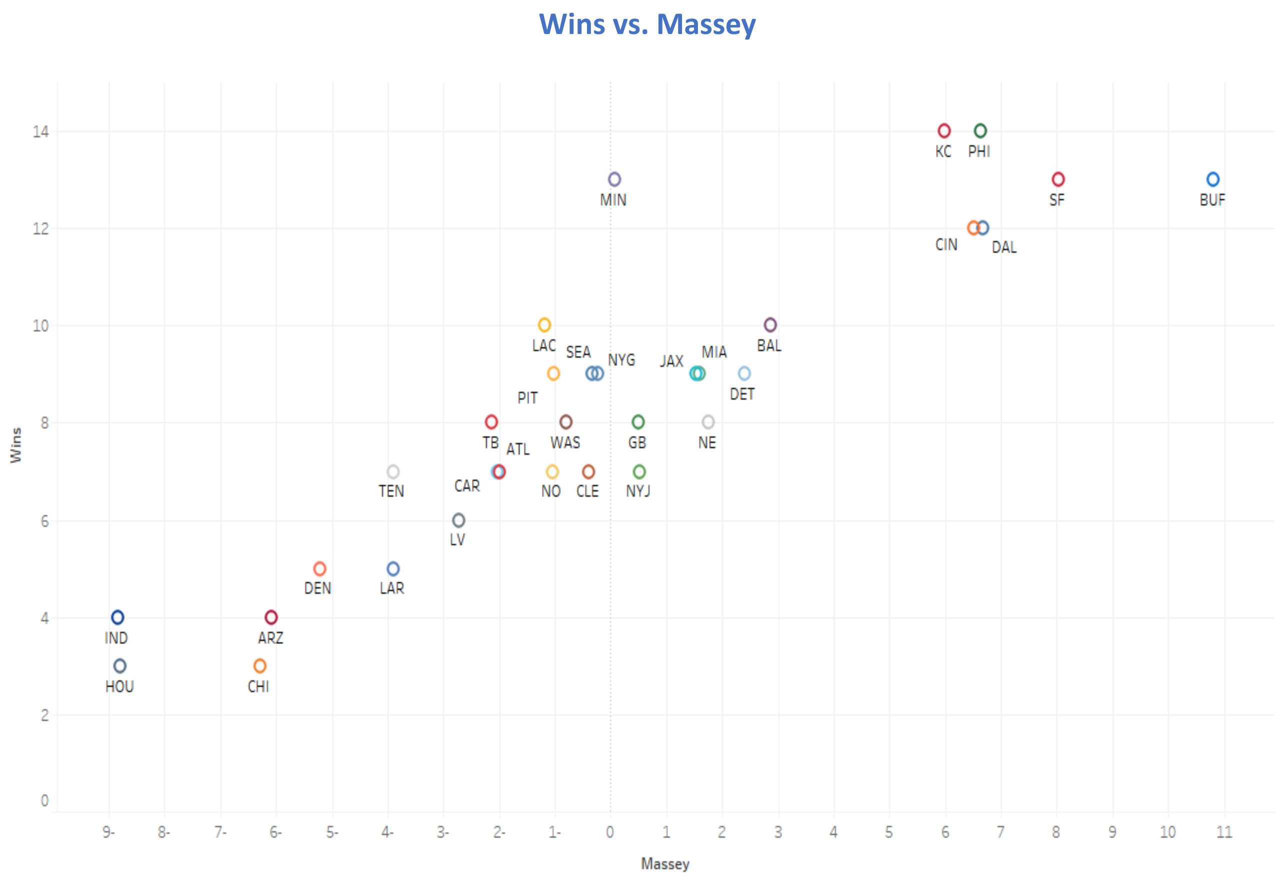
# games played

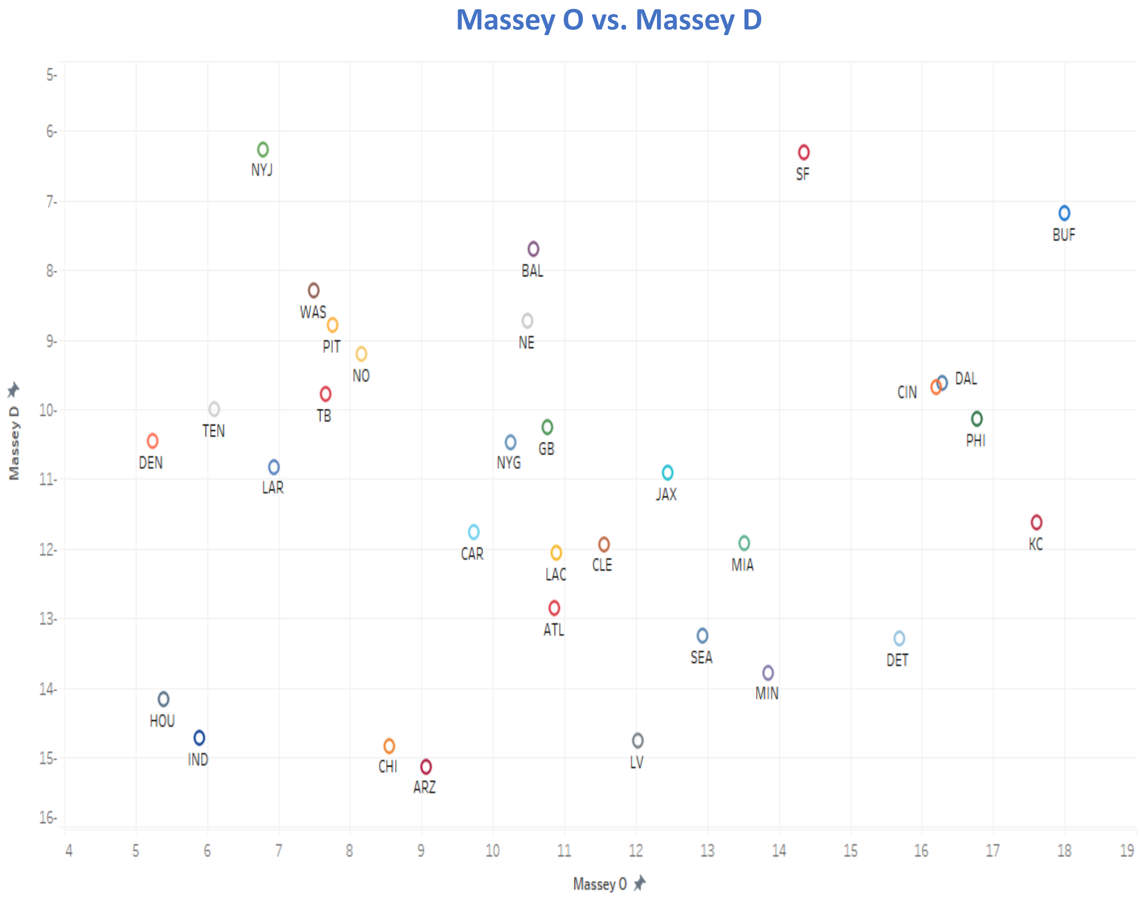
Less the ratings of the teams they played

$$\begin{array}{ccccc} \mathbf{M} & & \mathbf{r} & & \mathbf{p} \\ \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 3 & 0 & -1 \\ -1 & -1 & 0 & 3 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{pmatrix} & * & \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} & = & \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} \end{array}$$

These two teams didn't play each other, so fewer games and 0s where their respective rows and columns meet

# Massey Method | Results



# Massey Method | Results

Benefits

- Very simple- you could build using excel
- Easy to explain and intuitive

Main disadvantage: completely linear, so is thrown off by outliers (in this case, blowouts)

Fun fact: basis for one of the models that contributed to the old BCS rankings

# POP QUIZ: What is LaPlace’s rule of succession?

# Colley Method | Theory

Theory- flat winning % does not account for strength of schedule. Instead, incorporate Laplace's Rule of Succession to get:

$$r_i = \frac{1 + w_i}{2 + t_i}$$

Where:

$r_i$: rating

$w_i$: team wins

$t_i$: team games played

LaPlace's rule of succession:

$$P(X_{n+1} = 1 \mid X_1 + \cdots + X_n = s) = \frac{s+1}{n+2}.$$

Use: to estimate underlying probabilities when there are **few observations or events that have not been observed** to occur at all in (finite) sample data.

# Colley Method | Derivation

- The main benefit of this theorem is that it incorporates strength of schedule using a very simple formula

1. $$r_i = \frac{1 + w_i}{2 + t_i}$$ LaPlace's Rule of Succession

2. $$w_i = \frac{w_i - l_i}{2} + \frac{w_i + l_i}{2}$$ Decomposing 1. by adding $(-\frac{l_i}{2} + \frac{l_i}{2})$

3. $$= \frac{w_i - l_i}{2} + \frac{t_i}{2}$$ $\frac{w_i + l_i}{2} = \frac{t_i}{2}$

4. $$= \frac{w_i - l_i}{2} + \sum_{j=1}^{t_i} \frac{1}{2}$$ Rewriting 3

5. $$\sum_{j=1}^{t_i} \frac{1}{2} \approx \sum_{j \in O_i} r_j$$

6. $$w_i = \frac{w_i - l_i}{2} + \sum_{j \in O_i} r_j$$

Linchpin: the average rating is ½ , so you can sub in the sum of opponents' ratings for games played * ½

7. $$r_i = \frac{1 + \frac{w_i - l_i}{2} + \sum_{j \in O_i} r_j}{2 + t_i}$$ Therefore, a team's rating is dependent on the ratings of the teams they play against: strength of schedule

Some more fancy algebra (which I'm skipping) gives us:

$$Cr = b$$

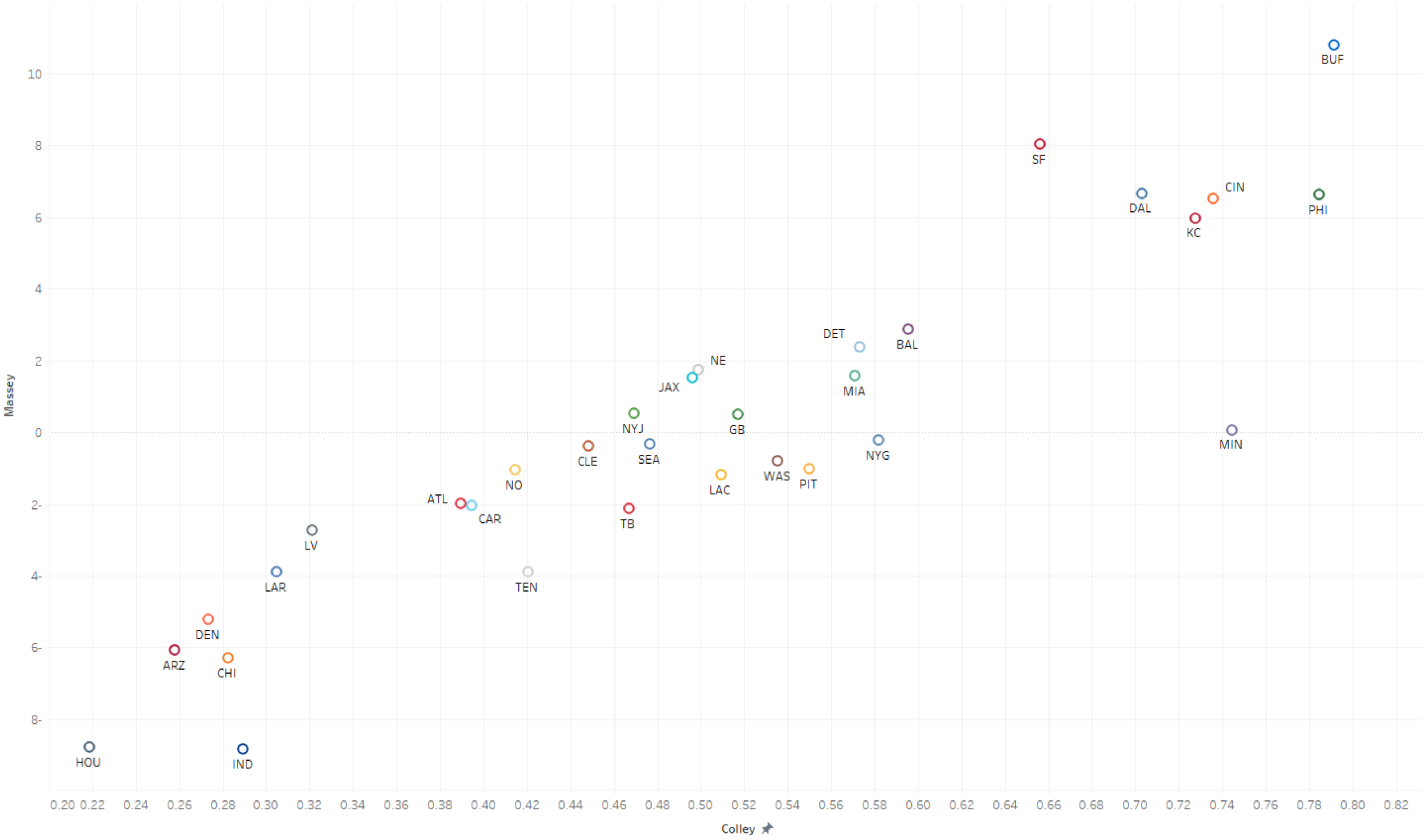
$C$: Colley Matrix

$$C_{ij} = \begin{cases} 2 + t_i & i = j \\ -n_{ij} & i \neq j \end{cases}$$

$r$: ratings vector

$b$: $b_i = 1 + \frac{1}{2}(w_i - l_i)$

# Colley Method | Results



# Colley Method | Results

Benefits
- Incorporates strength of schedule
- Not tied to game score
- Conservation property- average rating remains .5

Disadvantages
- Not tied to game score
- Conservation property doesn’t make sense if there is a skewed distribution of the quality of teams

Fun fact- also the basis for one of the models of the old BCS ranking system

# Combining Massey and Colley Methods

$$C_{ij} = \begin{cases} 2 + t_i & i = j \\ -n_{ij} & i \neq j \end{cases}$$

$$M_{ij} = \begin{cases} t_i & i = j \\ -n_{ij} & i \neq j \end{cases}$$

$$C = 2I + M$$

$$(2I + M)r = p\text{: Colleyized Massey Method}$$

Show results

# POP QUIZ: What does ELO stand for?

# Elo Method | Theory

Theory- Each player's performance is a normally distributed random variable whose mean can only change slowly over time- therefore the new rating should be a function of the old rating + a change (which is in turn a function of deviation from the expected mean)

$$r_{new} = r_{old} + K(S - \mu)$$

$$\mu_{ij} = \frac{1}{1 + 10^{-d_{ij}/\varepsilon}}$$

$$d_{ij} = r_i(old) - r_j(old)$$

*K*: adjustment factor- higher value gives more impact to one outcome

*S*: performance: % of points/games in a matchup, or just w vs. l

$\mu$: average expected performance based on difference in ratings

$S - \mu$: performance relative to expected performance

$\xi$: logistic parameter- effects the spread of the ratings

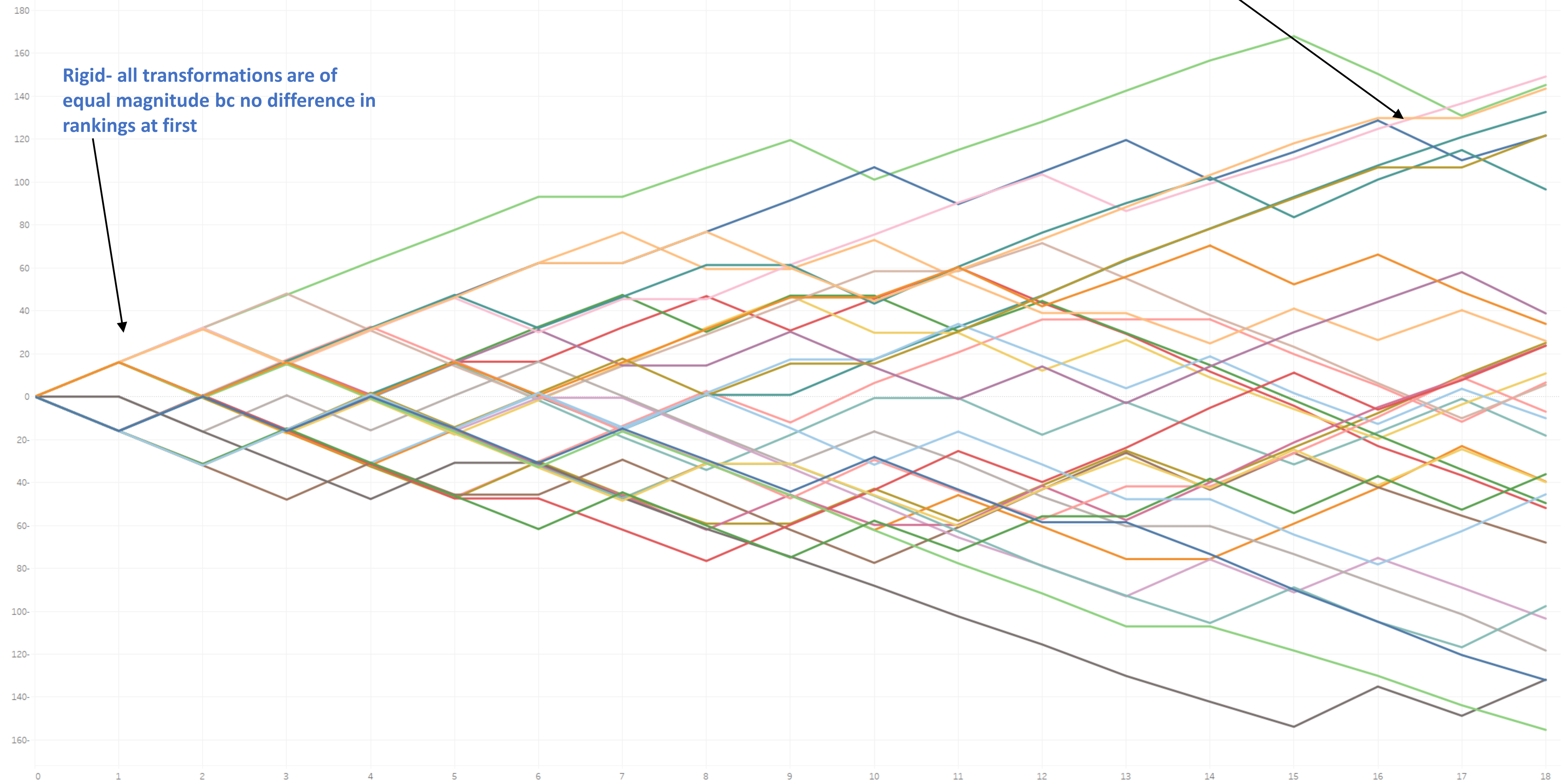
Implications:

- A difference in ratings of 0 implies an expected mean of a tie
- The better player winning will have a smaller impact on changes in ratings than the better player losing

More fluid- not all wins are equal anymore

# Elo Method | Results Over Time

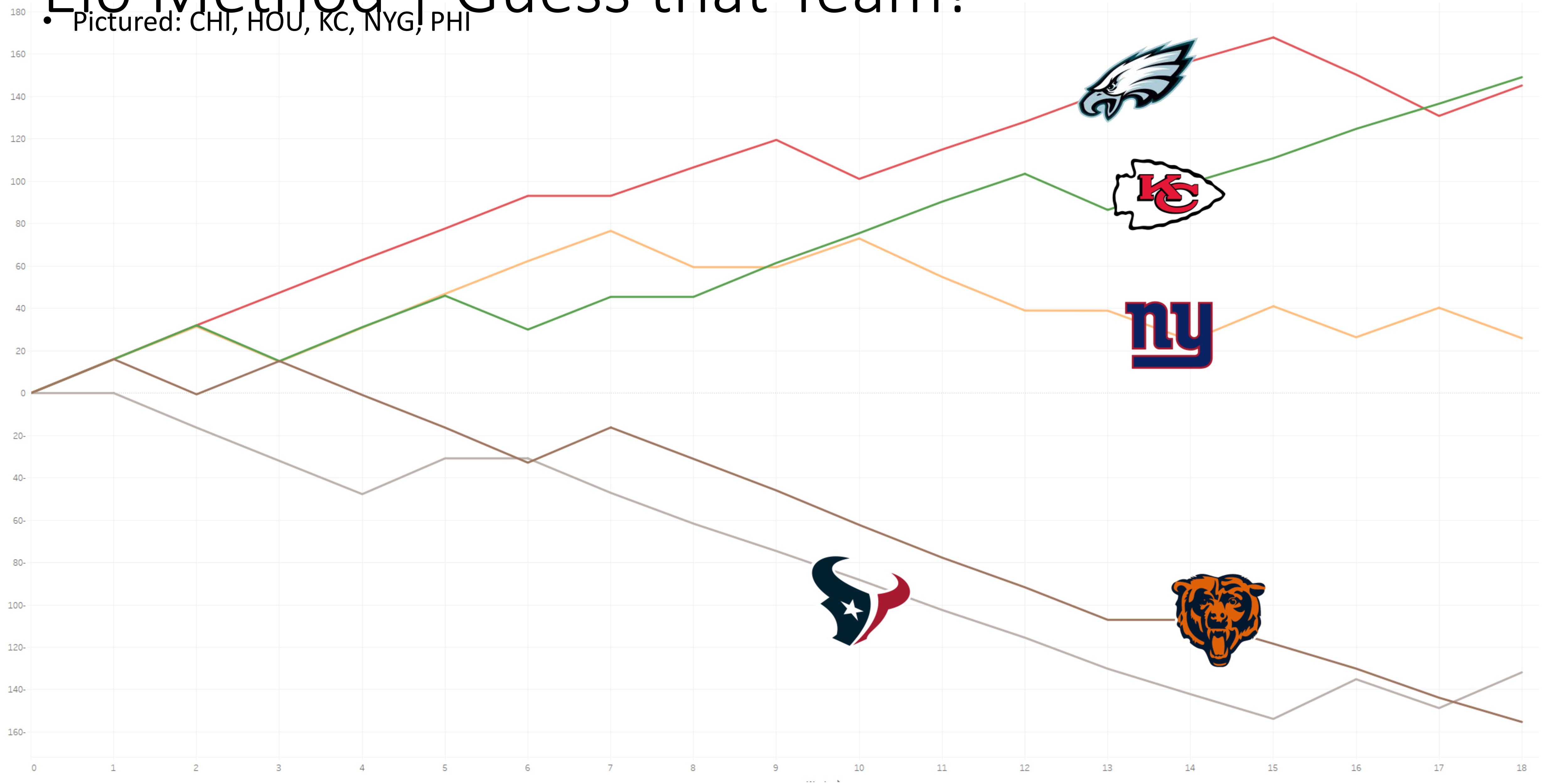
Rigid- all transformations are of equal magnitude bc no difference in rankings at first



# POP QUIZ: Guess the team!

# Elo Method | Guess that Team!

- Pictured: CHI, HOU, KC, NYG, PHI



# Elo Method | Results

Benefits:

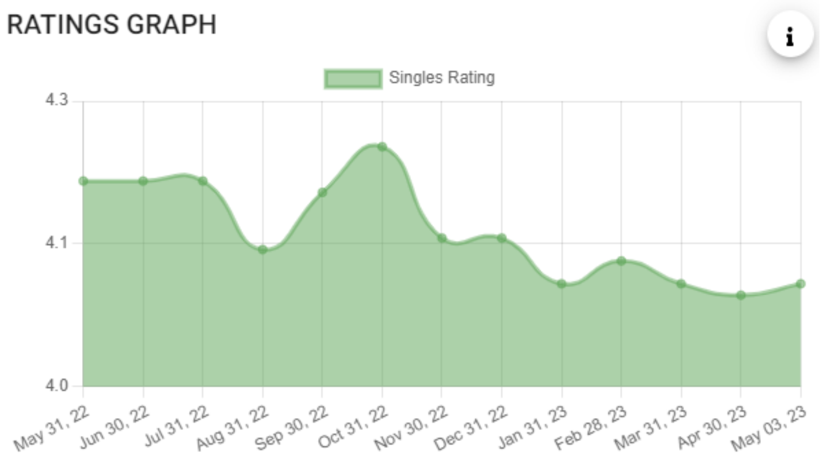
- Rewards weaker players for beating stronger players
- Shows evolution over time extremely well
- Good for individual competitions, you can respect the variability of one players performance
- Bad for anything where offense and defense is separated
- Doesn’t need same # games, can also handle asynchronous play

Main disadvantage: takes a lot of data before subtleties start showing, can be kind of swingy

Improvements: adjusting K over time, using points spread

Fun fact: invented for chess, used for squash (sort of)

**John’s Squash Ranking**



POP QUIZ: What is an Eigenvalue or an Eigenvector?

# Markov Method | Theory

Theory- every matchup between two teams is an opportunity for the weaker team to vote for the stronger team

Eigen vectors/values- For any matrix, you can find vectors for which the multiplication doesn't move the vector off of its span (same ratio)- just stretches or squishes it

$$Sr = \lambda r$$

Dominant eigenvalue $\lambda = 1$

Markov chain

- The probability of each event depends only on the state of the system of the previous event
- System is in a specific state at each point of time- can be represented by nodes or on a graph sometimes
- Movement described as probability
- **Steady state- movement (game data) doesn’t change vector (ratings)**

$$Sr = r$$

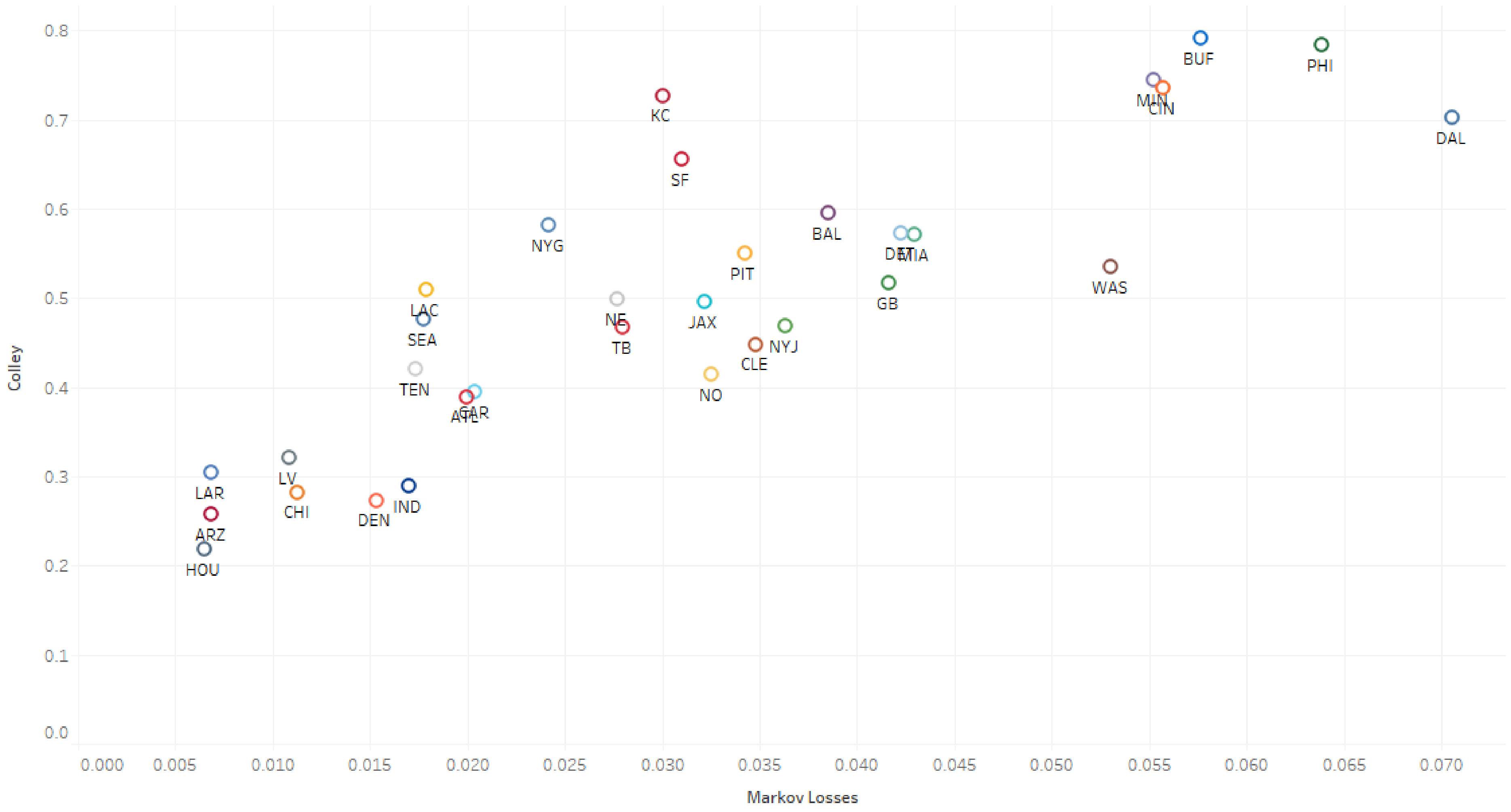
S- normalized voting matrix that is made stochastic

r- dominant eigenvector of S

Notes:

- Extremely intuitive- normalized to 100%- what % of our time do we spend with each team?
- Great to combine different metrics- simply add weighted S matrices before solving
- Very bad at handling undefeated teams
- Seems to respect beating good teams and losing to bad teams a little too much

# Markov Method | Results



# Markov Method | Results

Benefits

- Extremely intuitive conceptually- normalized to 100%- what % of our time do we spend with each team?
- Great to combine different metrics- simply add weighted S matrices before solving
- Big teams playing big teams seem to have the largest implications (more on that later)

Disadvantages

- Very bad at handling undefeated teams and not great at ties
- Can be challenging to interpret results

Application: Referral patterns!

# Offense-Defense Method | Theory

Theory- A team's offense is a product of their opponents' defenses, and must consider the quality of defense and their performance against them (and vice versa)

$$o_j = \sum_{i=1}^{m} \frac{a_{ij}}{d_i}$$

$$\mathrm{d}_i = \sum_{\mathrm{j}=1}^{m} \frac{a_{ij}}{\mathrm{o}_j}$$

$a_{ij}$: the number of points that j scored against I

$d_i$: i's defensive rating

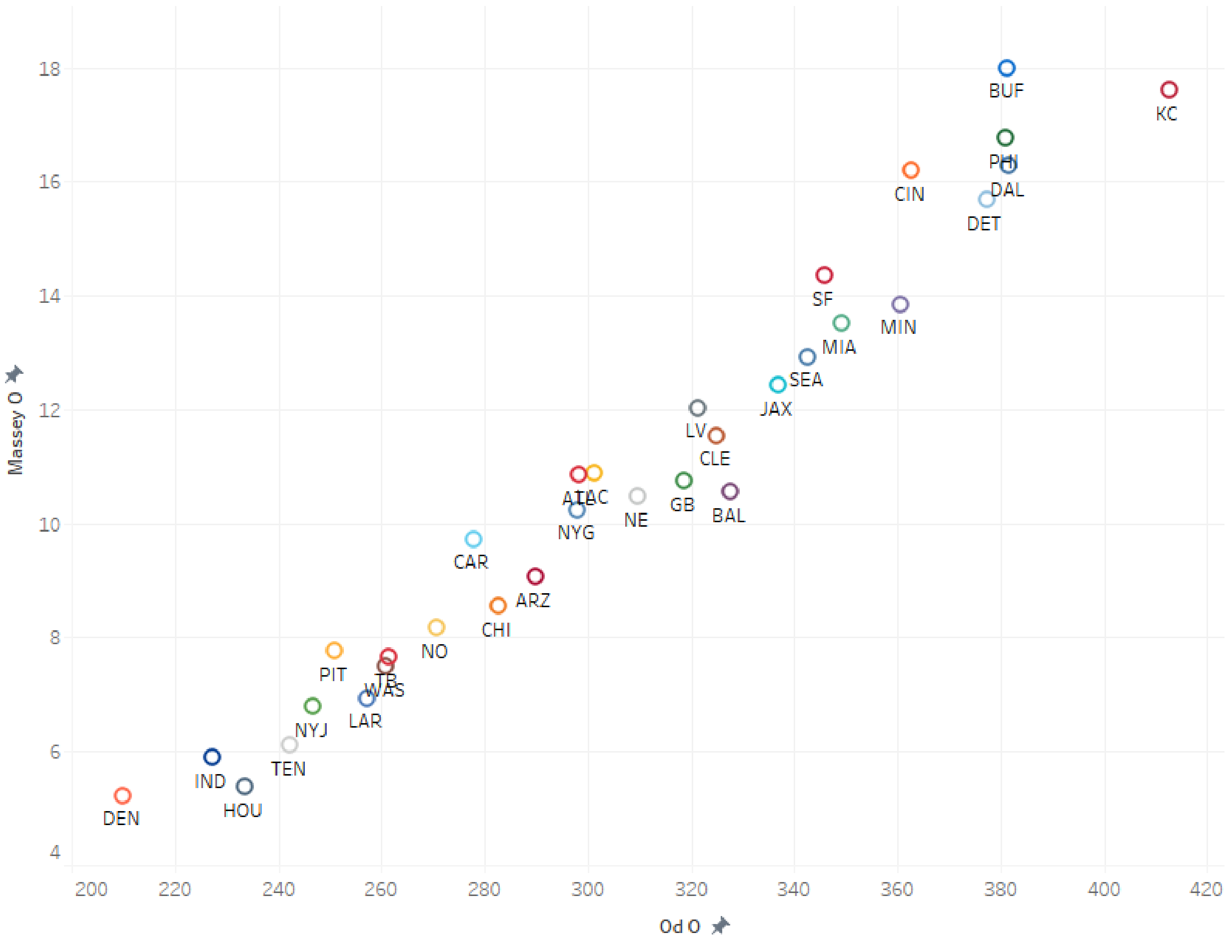
$o_j$: j's offensive rating

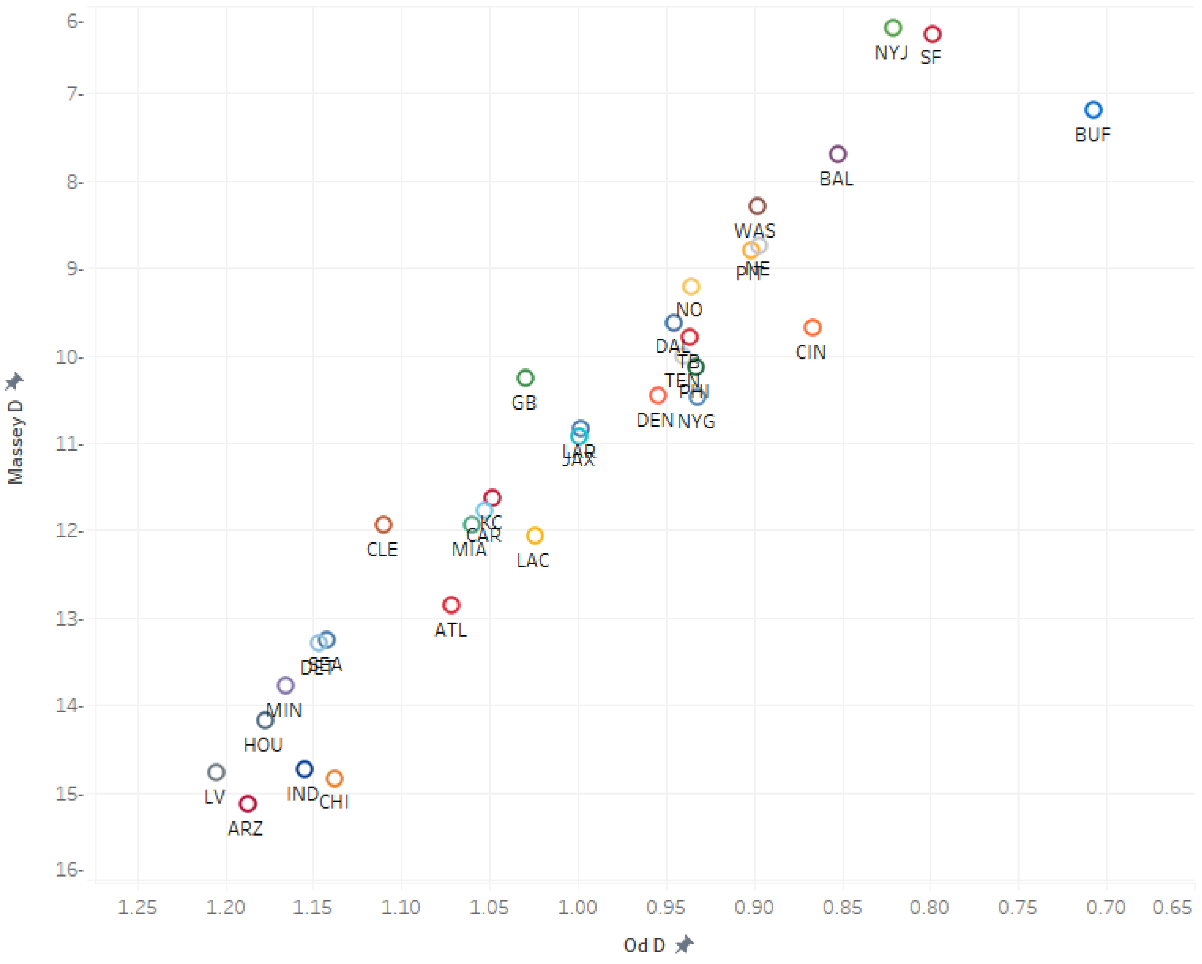
$$\vec{o} = A^T \vec{d}^{\div}$$

$$\vec{d} = A\vec{o}^{\div}$$

**Iteration required**

# Offense-Defense Method | Results



# Offense-Defense Method | Results

Very similar to Massey O vs. D methodology

- Massey solves for “closest” solution while O-D converges on “correct” solution
- O-D method has many mathematical nuances and unanswered questions
- Which is right for you? You have to look at how it plays out in the data!

# Sensitivity

### Colley

| Week | winner | home team | loser | PtsW | PtsL |
|---|---|---|---|---|---|
| 3 | Green Bay Packers | @ | Tampa Bay Buccaneers | 14 | 12 |
| 4 | Minnesota Vikings | @ | New Orleans Saints | 28 | 25 |
| 11 | Atlanta Falcons | vs. | Chicago Bears | 27 | 24 |
| 16 | Carolina Panthers | vs. | Detroit Lions | 37 | 23 |
| 12 | Washington Commanders | vs. | Atlanta Falcons | 19 | 13 |
| 2 | New York Giants | vs. | Carolina Panthers | 19 | 16 |
| 1 | Chicago Bears | vs. | San Francisco 49ers | 19 | 10 |
| 8 | Minnesota Vikings | vs. | Arizona Cardinals | 34 | 26 |
| 4 | Seattle Seahawks | @ | Detroit Lions | 48 | 45 |
| 17 | New Orleans Saints | @ | Philadelphia Eagles | 20 | 10 |

### Markov

| Week | winner | home team | loser | PtsW | PtsL |
|---|---|---|---|---|---|
| 10 | Minnesota Vikings | @ | Buffalo Bills | 33 | 30 |
| 2 | Dallas Cowboys | vs. | Cincinnati Bengals | 20 | 17 |
| 17 | New Orleans Saints | @ | Philadelphia Eagles | 20 | 10 |
| 8 | Philadelphia Eagles | vs. | Pittsburgh Steelers | 35 | 13 |
| 2 | Philadelphia Eagles | vs. | Minnesota Vikings | 24 | 7 |
| 6 | Buffalo Bills | @ | Kansas City Chiefs | 24 | 20 |
| 1 | Tampa Bay Buccaneers | @ | Dallas Cowboys | 19 | 3 |
| 16 | Dallas Cowboys | vs. | Philadelphia Eagles | 40 | 34 |
| 12 | Philadelphia Eagles | vs. | Green Bay Packers | 40 | 33 |
| 1 | Philadelphia Eagles | @ | Detroit Lions | 38 | 35 |

### ELO

| Week | winner | home team | loser | PtsW | PtsL |
|---|---|---|---|---|---|
| 16 | Carolina Panthers | vs. | Detroit Lions | 37 | 23 |
| 15 | Green Bay Packers | vs. | Los Angeles Rams | 24 | 12 |
| 16 | San Francisco 49ers | vs. | Washington Commanders | 37 | 20 |
| 17 | Seattle Seahawks | vs. | New York Jets | 23 | 6 |
| 14 | Los Angeles Chargers | vs. | Miami Dolphins | 23 | 17 |
| 16 | Jacksonville Jaguars | @ | New York Jets | 19 | 3 |
| 16 | Pittsburgh Steelers | vs. | Las Vegas Raiders | 13 | 10 |
| 16 | Cincinnati Bengals | @ | New England Patriots | 22 | 18 |
| 16 | Baltimore Ravens | vs. | Atlanta Falcons | 17 | 9 |
| 16 | New Orleans Saints | @ | Cleveland Browns | 17 | 10 |

Colley- more swayed by unexpected results- better team losing
ELO- same thing? harder to do sensitivity- all of the most important games are later
Markov- matchups between two good teams seem to matter the most

# Further Reading

*Who's #1? The Science of Rating and Ranking* | Amy Langville and Carl D. Meyer

*Mathletics* | Wayne Winston

*Moneyball: The Art of Winning an Unfair Game* | Michael Lewis

*Sprawlball: A visual Tour of the New Era of the NBA* | Kirk Goldsberry

*Scorecasting: the Hidden Influences Behind How Sports Are Played and Games Are Won* | Tobias Moskowitz and L. Jon Wertheim