Although I have no practical experience in sports analytics, I am passionate about sport and maintain my own involvement on summer weekends as the vice-captain of a club side. Therefore, two of my ideas relate to quintessentially Australian sports which I enjoy either watching or playing.

## Idea 1) AFL team success by player behaviour

### Background

The Australian Football League (AFL) is a football code which currently has 18 competitive teams. This contact sport is played by 18 on each side plus four subs each that rotate on and off as required. The basic aim is to kick more goals than the opposition, with a goal worth six points. At any one time, the ball is in dispute, in clear possession, or airborne.

AFL appears as a hybrid of NFL, rugby and soccer, is continuous in nature – thus requiring and rewarding supreme levels of fitness – and apart from scheduled quarter breaks the only stoppages in play occur when a) a goal is scored, b) the ball is “trapped” in dispute and is thrown up by the umpire, c) a player catches (“marks”) the ball from a kick of at least 15 metres distance, or d) a player is penalised for an action.

### Start With the Why

AFL is a data-rich sport, with every touch of the ball and individual effort recorded by data company, Champion Data. In a single game played over approximately 1¾ hours, an elite midfield player will run around 18-20km including 3-5km of sprinting, touch the ball 20-30 times, tackle an opponent 5-10 times and perhaps take some marks and kick some goals. AFL is a sport requiring many aspects: skill, fitness, strength, decision-making, teamwork, tactics and effort.

If I was coaching an AFL side in pre-season, beyond the strategies, structures and tactics, I would want to know:

* Should my players be pushed harder for fitness or for skills? (and therefore, how should I set up my assistant coach structure?)
* What is more influential – skill or effort?
* What specific skills or effort would I focus on? Which ones are likely to be influential not just on one game but on a season?
* Is it better to have players with similar skill / effort attributes or players with different attributes in the same team?

### Data

Player data is freely available by year from 2001-2017. Measures of success could be finals results, ladder finish position, win/loss ratio.

Data will require manual effort to copy but is in a friendly format to collate in Excel before wrangling and feature engineering in R. AFL’s website restricts queries to 11 variables at a time so 2x extractions for each page will be required. Web scraping is plausible but unlikely to speed the process.

## Initial analysis

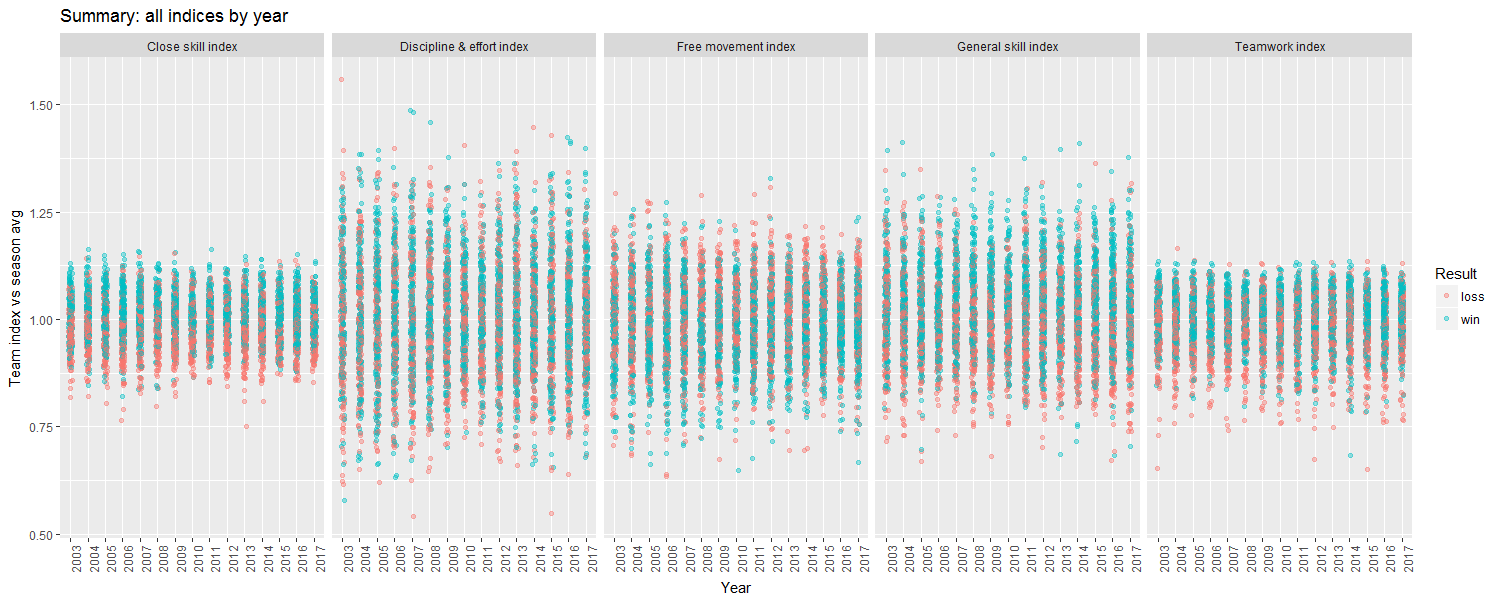
* removed draws from the data set (not meaningful to analysis of “how to win”)
* calculated season averages for each player statistic – glossary of terms:

|  |  |  |
| --- | --- | --- |
| **Disposals** – number of times a player kicked or handballed the ball | **Kicks** – disposing of the ball by kicking it, usually towards a team mate or goal | **Marks** – catching a ball which has been kicked a minimum of 15 metres |
| **Handballs** – disposing of the ball by using a closed fist to hit it out of the opposite hand | **Goals** – kicking the ball through the taller centre posts | **Behinds** – kicking the ball through the shorter side posts, or if a “goal” is touched by a player before crossing the line, or if a ball hits the taller posts |
| **Hit outs** – (like a jump-ball in basketball) when the umpire bounces or throws the ball up for a contest, a player wins a clear “tap” | **Tackles** – getting hold of a player who is in possession of the ball, must not be below the knees or above the shoulders | **Rebounds** – repelling the ball from the team’s defensive 50m arc – not used currently |
| **Inside 50s** – getting the ball inside the attacking 50m arc (distance from goal) | **Clearances** – when the umpire bounces or throws the ball up for a contest, the ball is cleared to the obvious advantage of the team | **Clangers** – an obvious mistake, an unforced error, e.g. dropping a mark under no pressure |
| **Free kicks** – receiving a free kick due to an opponent being penalised – not used currently and not likely to be a useful measure as it is caused by an act of the opponent | **Free kicks against** – giving away a free kick to the opposition due to being penalised | **Brownlow votes** – number of best & fairest votes (similar to MVP) – not useful and not controlled by the players |
| **Contested possessions** – winning the ball in a contest and disposing of it cleanly | **Uncontested possessions** – winning the ball in open space and disposing of it cleanly | **Contested marks** – a mark taken against one or more opponents, or a mark taken in a pack of both opponents and teammates |
| **Marks inside 50** – a mark taken inside the attacking 50m arc | **One percenters** – an act which shows extraordinary courage or effort | **Bounces** – upon taking possession of the ball a player can run with it but must either bounce it or dispose of it within 15m |
| **Goal assists** – as the name suggests, assisting a teammate’s goal | **% Played** – percentage of game time the player spent on the field – not used currently | **Subs** – permanent substitution of a player out/in from the game due to injury, fatigue or strategy – not likely to be a useful measure |

* for each team, each round, each season, calculated team averages for each statistic
* calculated indices at team level across five areas (in all cases, the index is calculated by performing the per round, per team sum and dividing by the competition average of the same calculation for that season):  
    
    
  + close skill – measures of a team’s ability to win a contested ball
  + discipline & effort – measures of a team’s discipline to avoid infringing rules and effort in making plays under physical duress
  + free movement – measures of a team’s free-running style, i.e. playing a continuous and flowing style of football
  + general skill – measures of a team’s general efficiency winning and using the ball
  + teamwork – measures of a team’s ability to work together and give each other options

Ratios have been multiplied by 10 within log functions to return a more workable result. The composite for discipline needed to multiply the first part by two to ensure the second part didn’t have too much influence on results (as this gave too large a spread of values).

### Plotting indices against wins and losses, over the years:



* Looks like potential positive relationships between close skill, general skill and teamwork with winning
* Less clear whether discipline and effort correlates with winning

### Plotting indices against wins and losses for finals games only, over the years:



* Possible stronger relationship between close skill and winning during finals games
* Possible stronger relationship between teamwork and winning during finals in 2015 and 2017

### Plotting difference between winning and losing team indices, over the years:

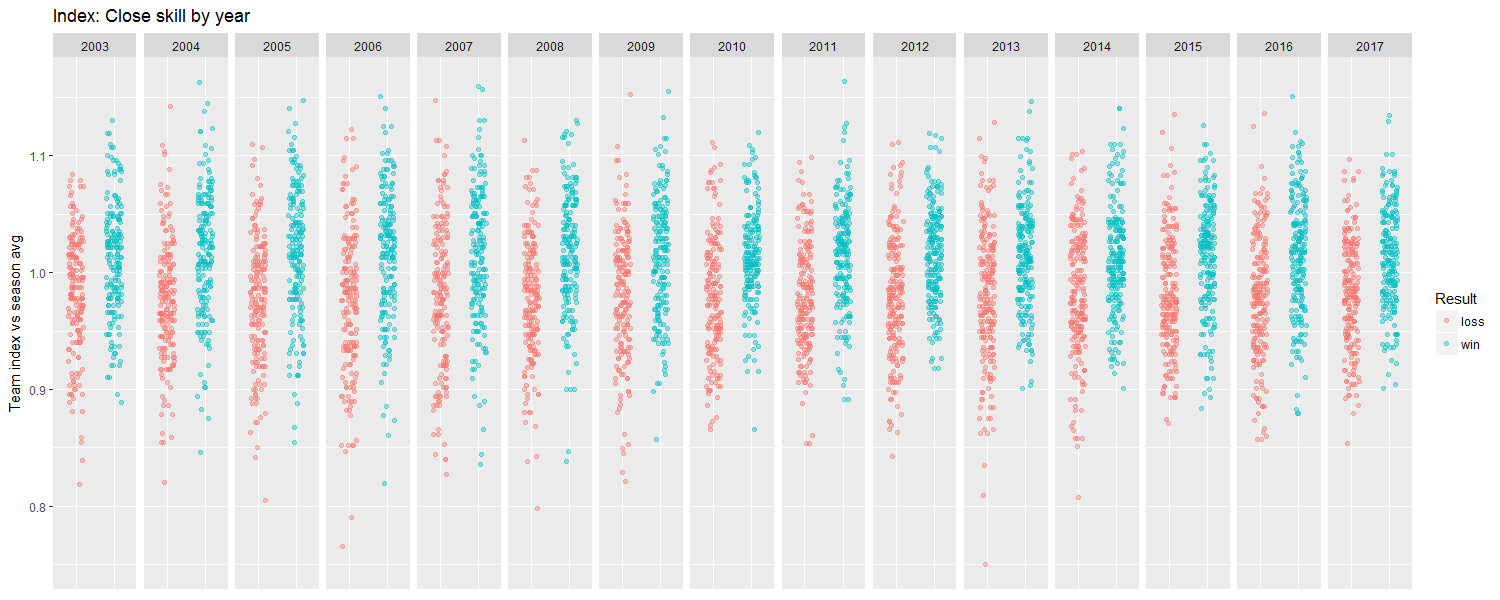


* Confirms potential positive relationships between close skill, general skill, teamwork and winning
* Also highlights a possible negative relationship between free movement and winning

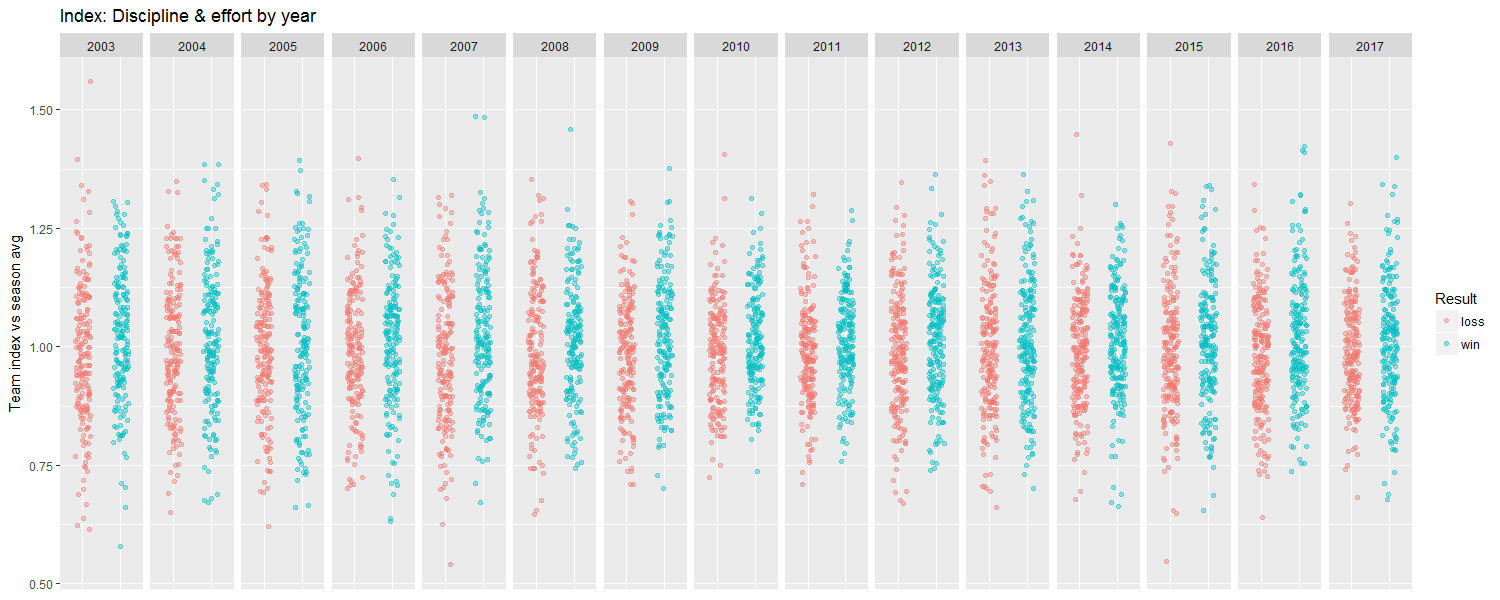
### Possible ways to cut the analysis for better insights:

* bring in opposition statistics as direct comparators (needs to be done to better validate the results visualised here)
* use margin of victory, not just binary win/loss (needed)
* consider impact of top or bottom 20% of players in each side (optional)
* cut different indices, e.g. “win the ball”, “supply the ball”, “stop opposition’s supply”
* see how results change if games with wide margins are removed, i.e. what are the differentiators of a close contest? (optional)

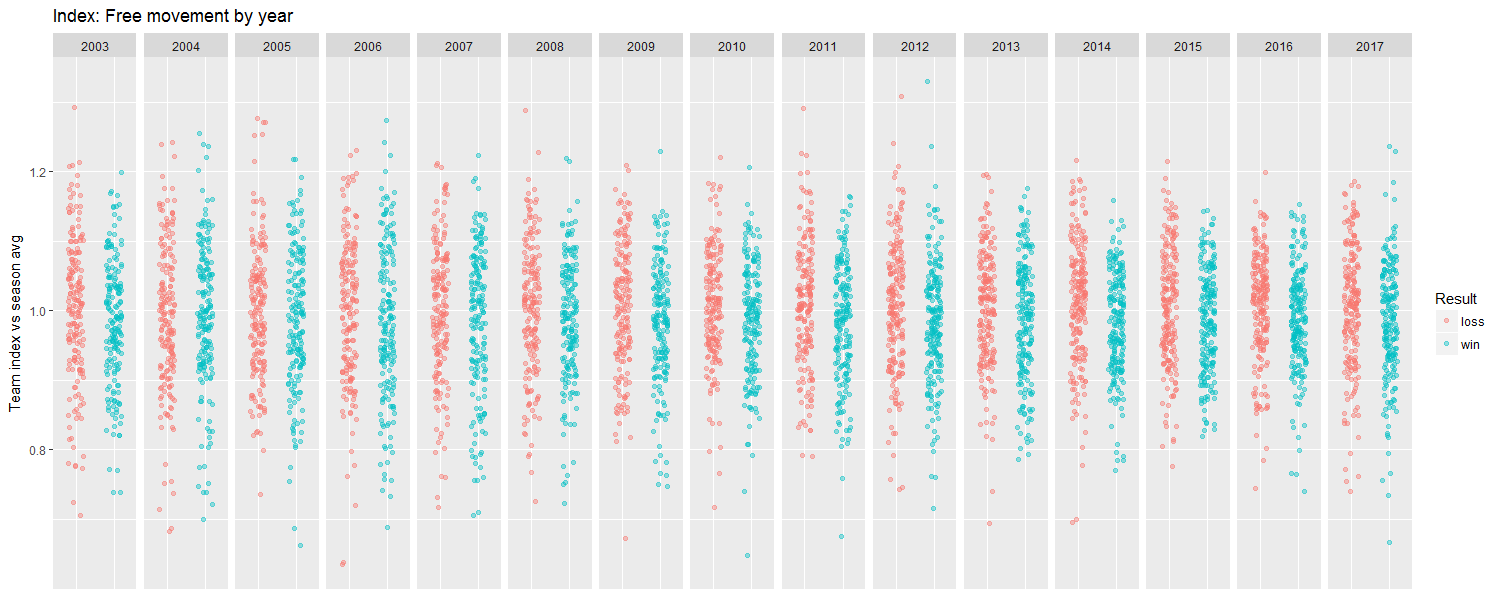
### Plotting each index over the years:



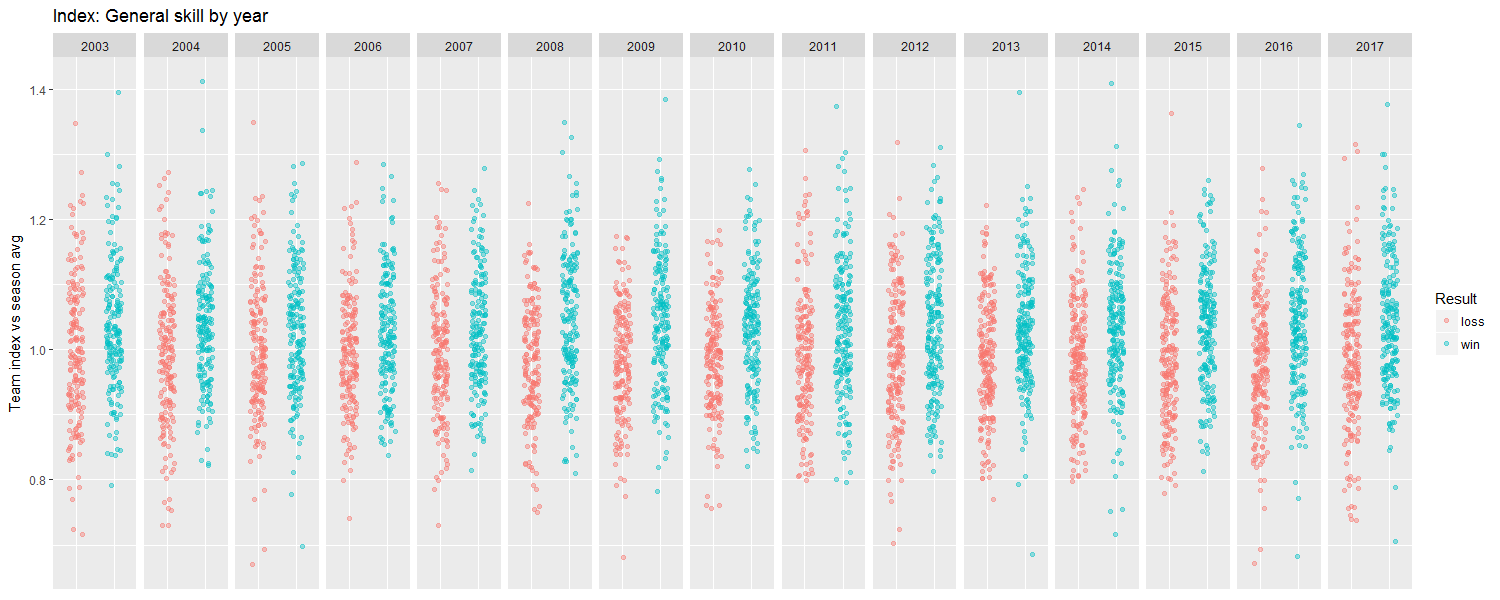
* probable positive relationship



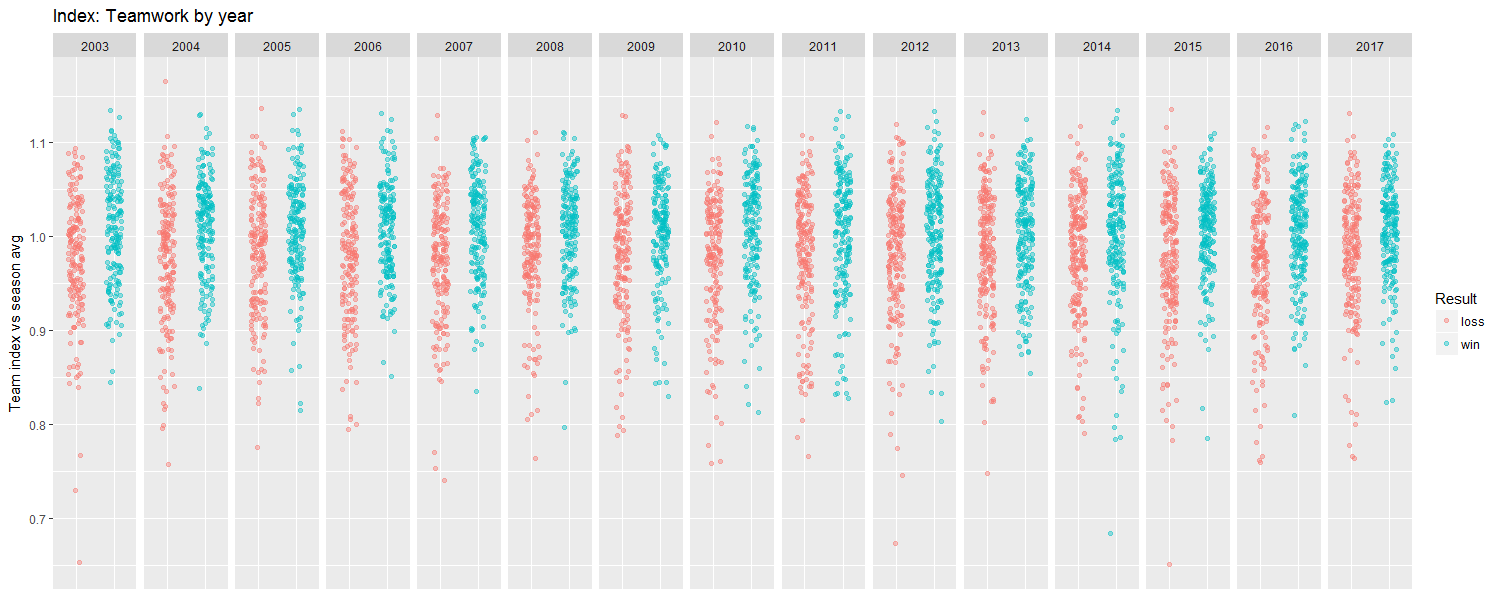
* no relationship



* possible negative relationship



* possible positive relationship



* possible positive relationship