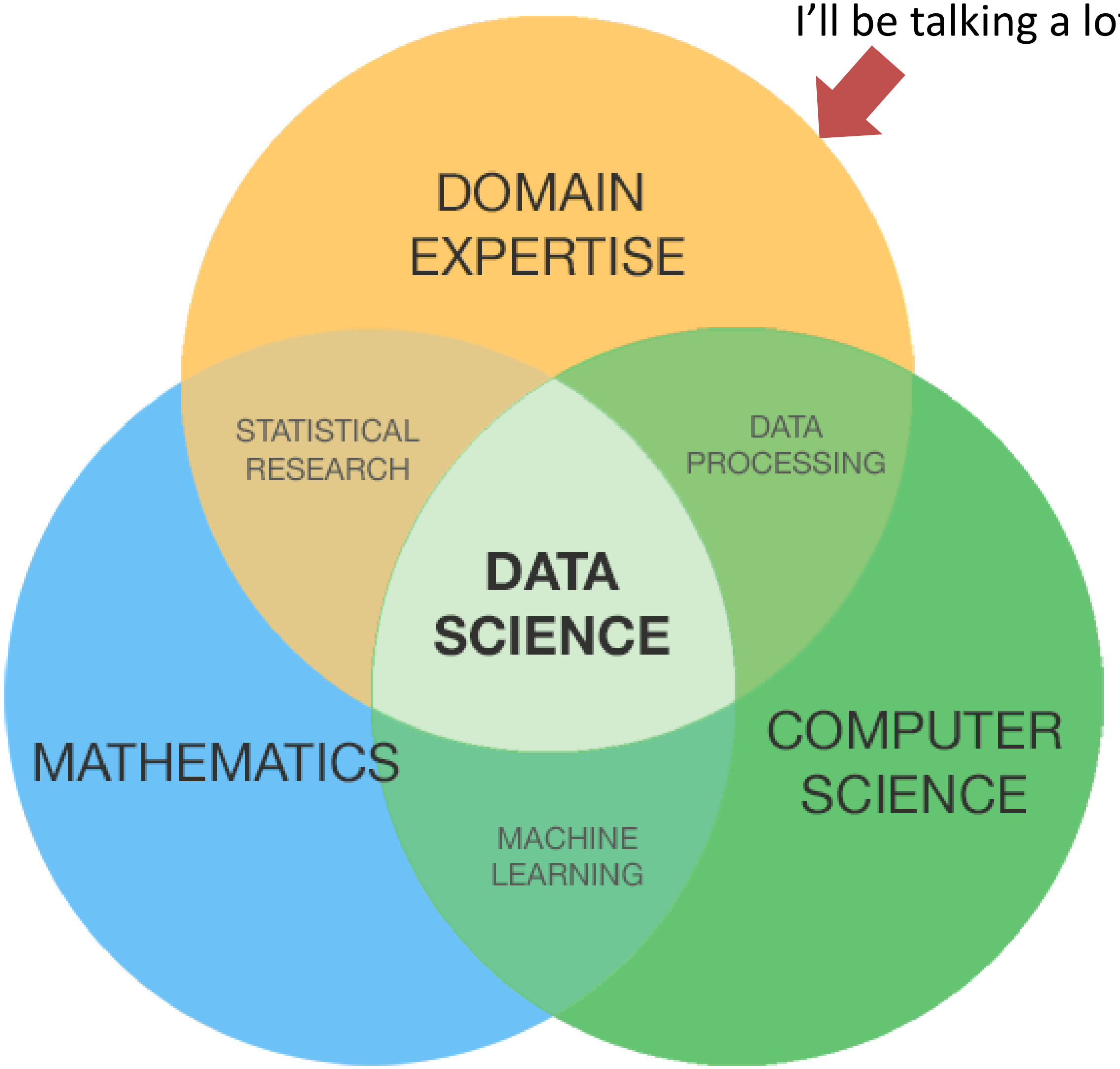
Applying Data Science to Australian Rules Football

Some Examples and Reflections

What Is Data Science?

I'll be talking a lot about this today



Source: www.datanimi.com

Where's The Science?

Science is a systematic enterprise that builds and organises knowledge in the form of testable explanations and predictions …

Wikipedia

AFL Explanations and Predictions

is the team currently on top of the league ladder really the best team right now?

which team was the best ever?

by how much is team X win likely to defeat team Y on Saturday? How confident should I be about that prediction?

if my team is leading by X points with M minutes to be played, how confident should I be that they'll win?

will team Z make the finals?

how many types of Grand Final are there?

Application #1: Team Ratings System

is the team currently on top of the league ladder really the best team right now?

which team was the best ever?

The Objective

Construct an ELO-style team rating system for VFL/AFL teams that:

Uses only historical results as inputs (ie requires no player data)

SIMPLE TO CONSTRUCT

Provides offensive and defensive ratings that have a straightforward interpretation (eg are measured in points or scoring shots)

READILY INTERPRETED

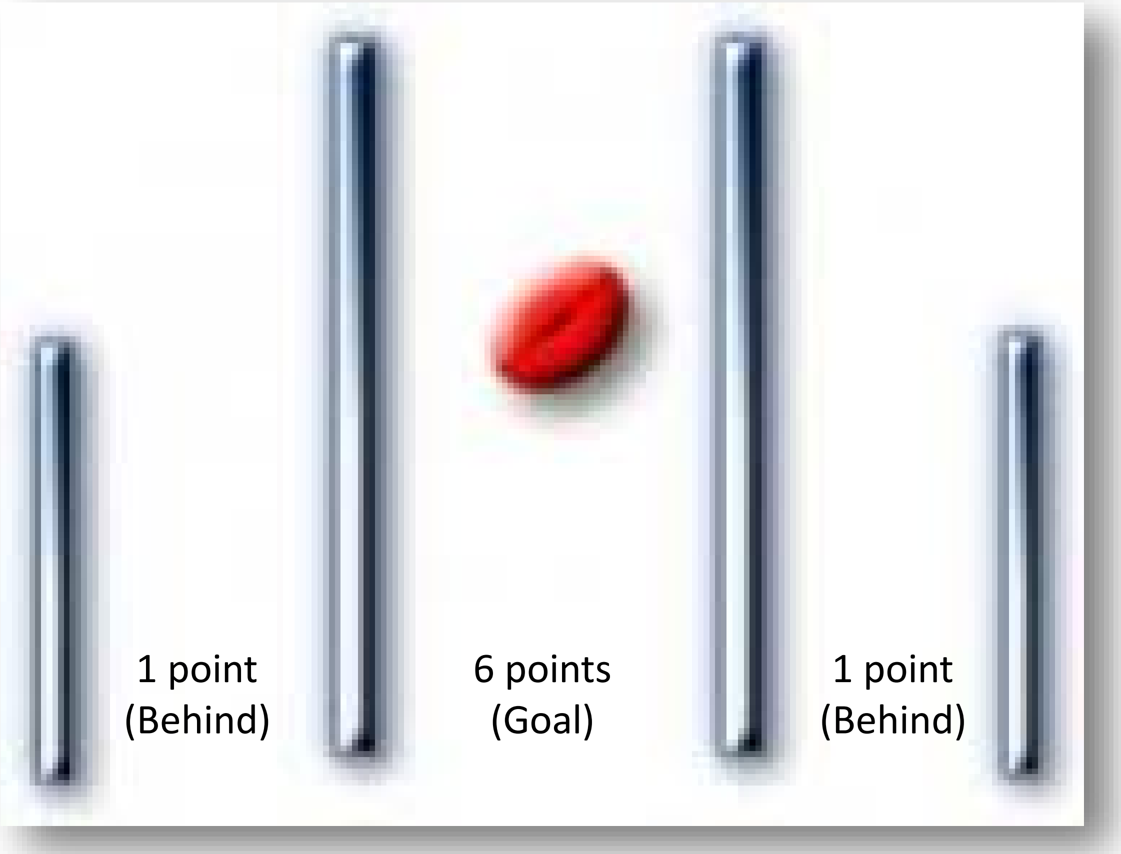
Can be applied across the entire history of the VFL/AFL (1897 to 2016 = 14,993 games)

BROADLY APPLICABLE

Provides acceptably accurate forecasts of game margins (home score less away score)

UNAMBIGUOUSLY MEASURABLE

Scoring in Australian Rules



- Scoring Shots = Goals + Behinds
- Conversion Rate = Goals / Scoring Shots

Set of Updating Equations

General principle of an ELO system is that ratings be updated based on performance relative to expectation

New Defensive Rating = Old Defensive Rating +
k x (Actual Defensive Performance – Expected Defensive Performance)

Actual Defensive Performance =
All-Team Scoring Shot Average –
Adjusted Opponent's Scoring Shots

Expected Defensive Performance =
Own Defensive Rating – Opponent's Offensive Rating + Venue Adjustment / 2

Adjusted Opponent's Scoring Shots =
min(Scoring Shot Cap, Actual Opponent's Scoring Shots)

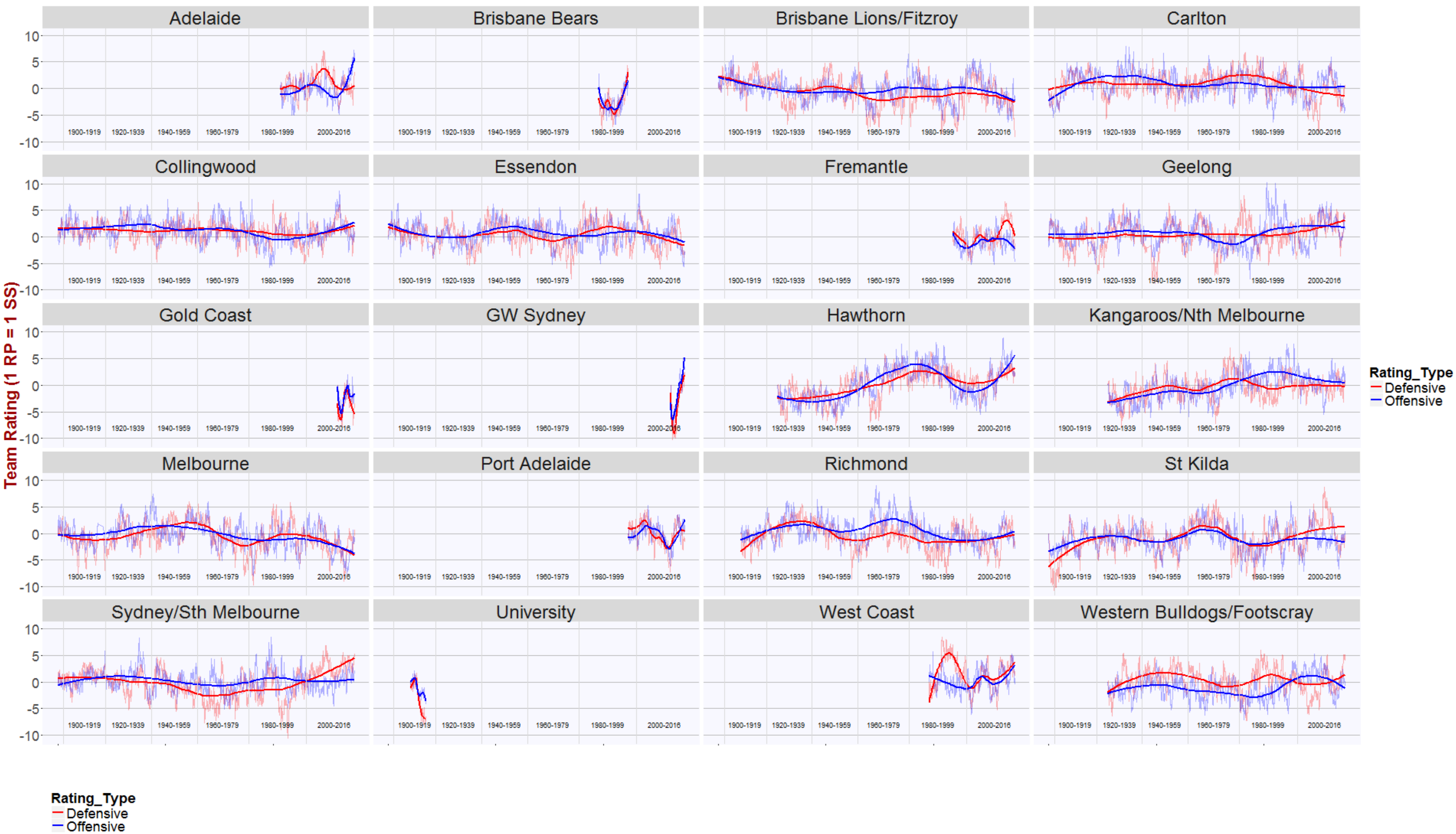
Venue Adjustment =
Net Venue Performance + Net Travel Penalty

All teams have a Venue Performance Rating at every venue (set to 0 for first 30 games)

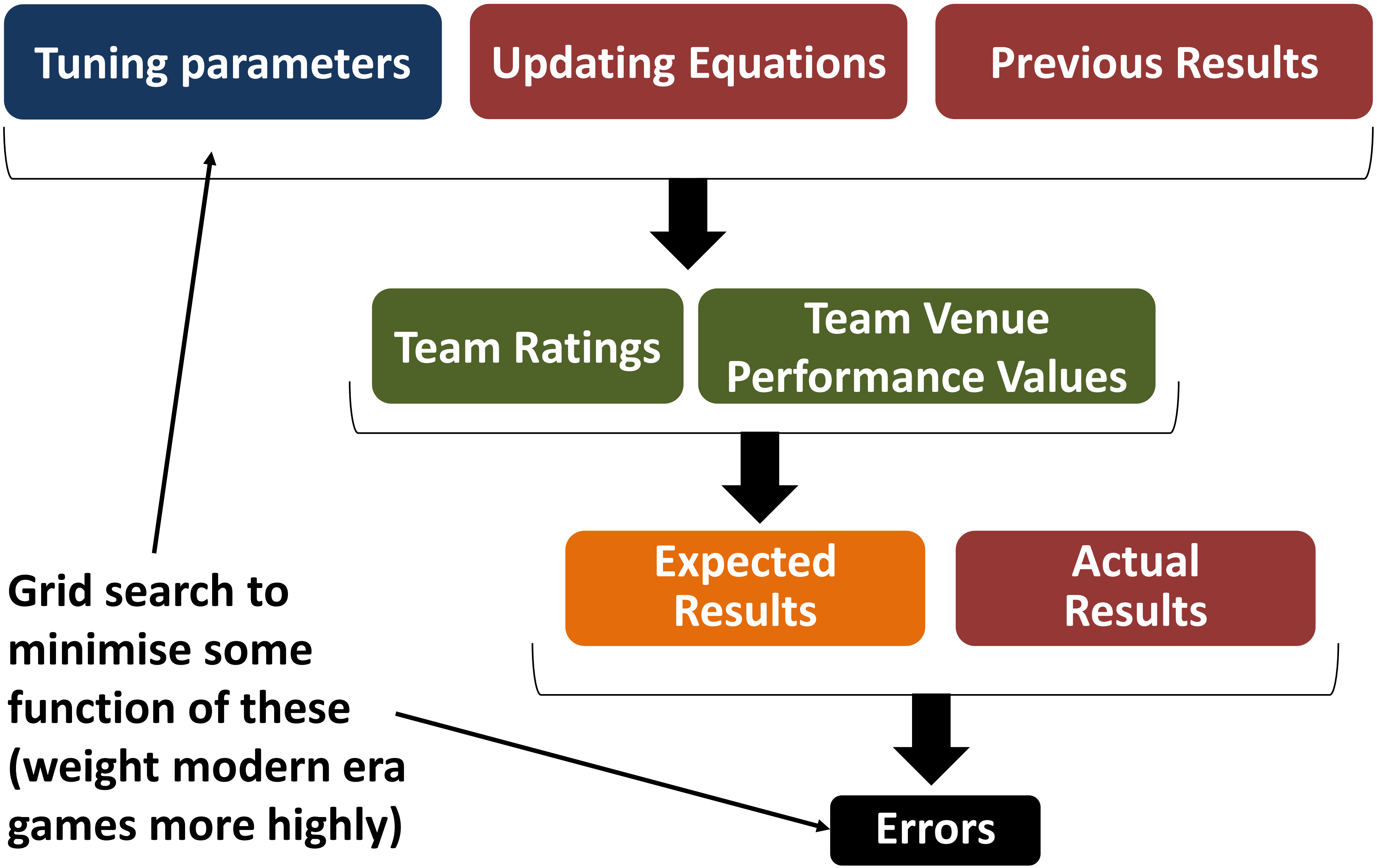
Note

Similar equations are used to update Offensive Ratings
Five different k values are used at different points in the season
Ratings are set to 0 initially

Historical Ratings: Offensive & Defensive



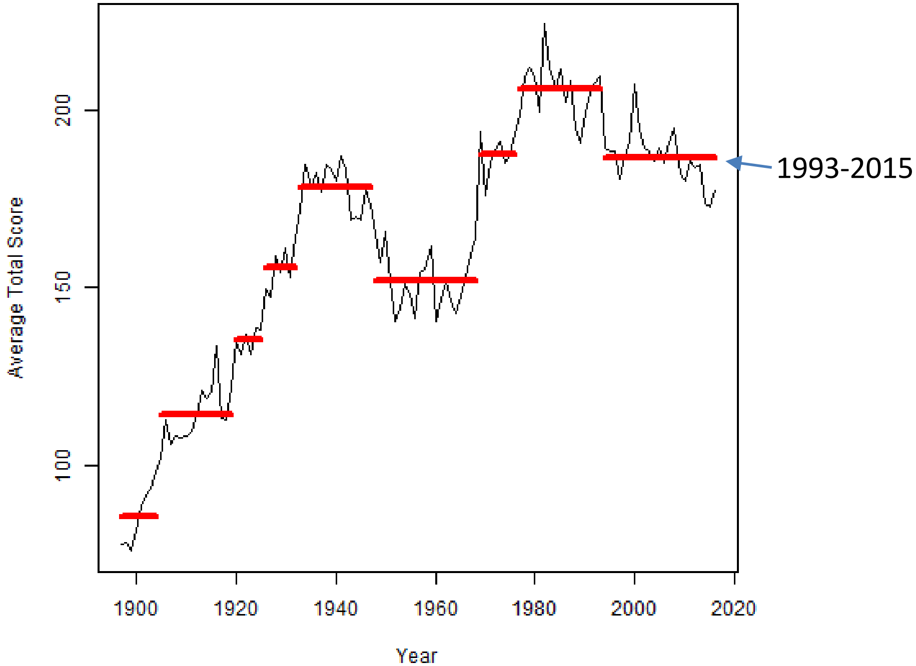
The Optimisation Process



Changepoint Analysis: Eras in V/AFL

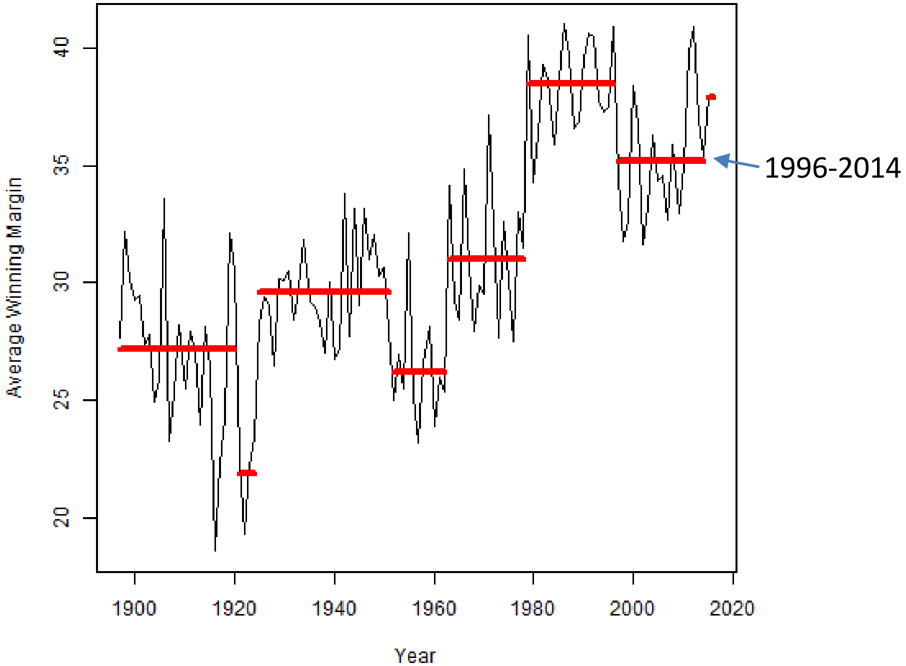
We want to weight errors in the "modern era" more heavily.

But when, approximately, did the modern era commence?



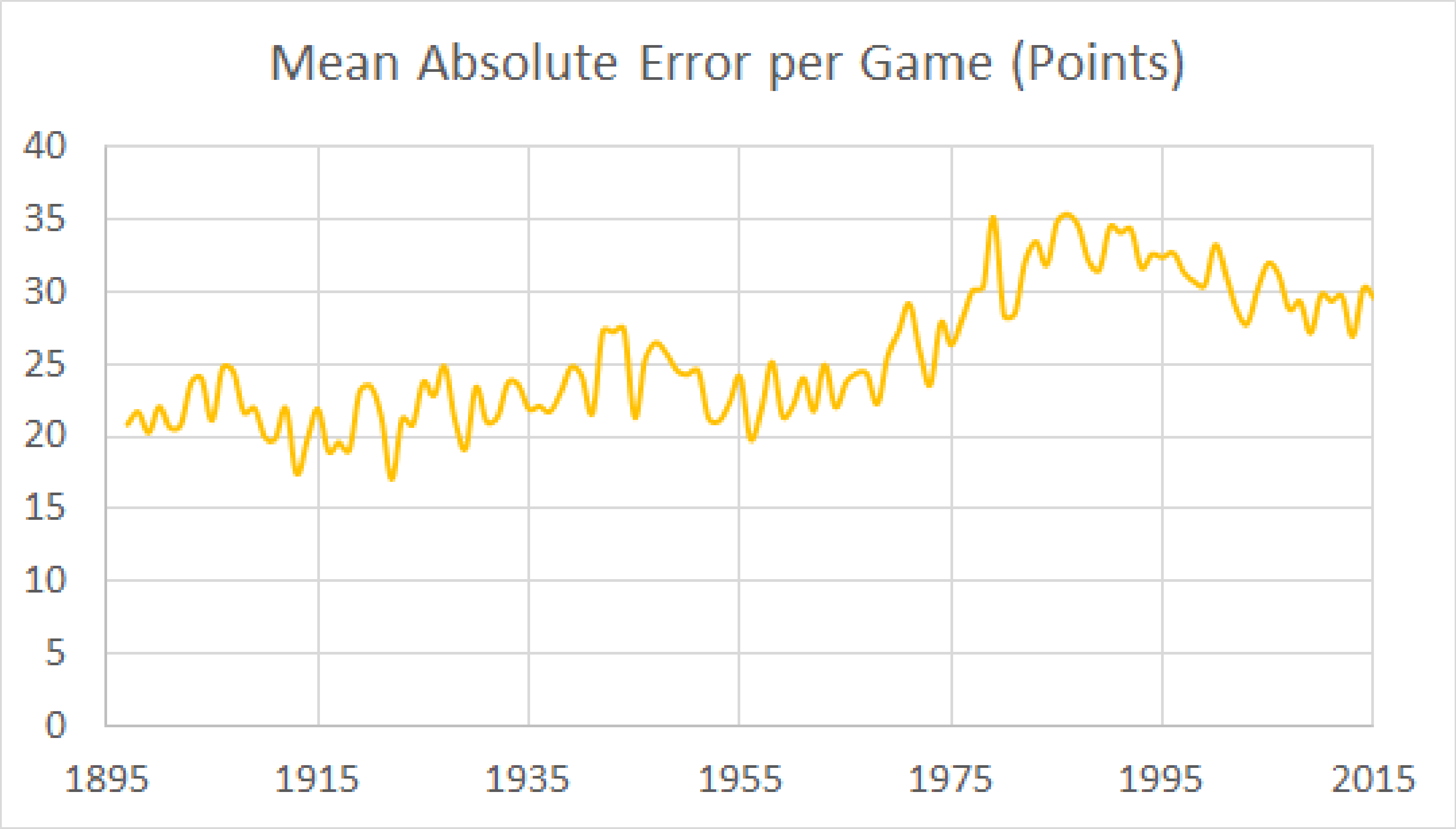
Based on the average total score per game in each season

Based on the average victory margin per game in each season



Here we search, simultaneously, for changes in the mean and variance of the timeseries

Performance



MAEs for entire seasons have generally finished in the 20 to 35 points per game range.

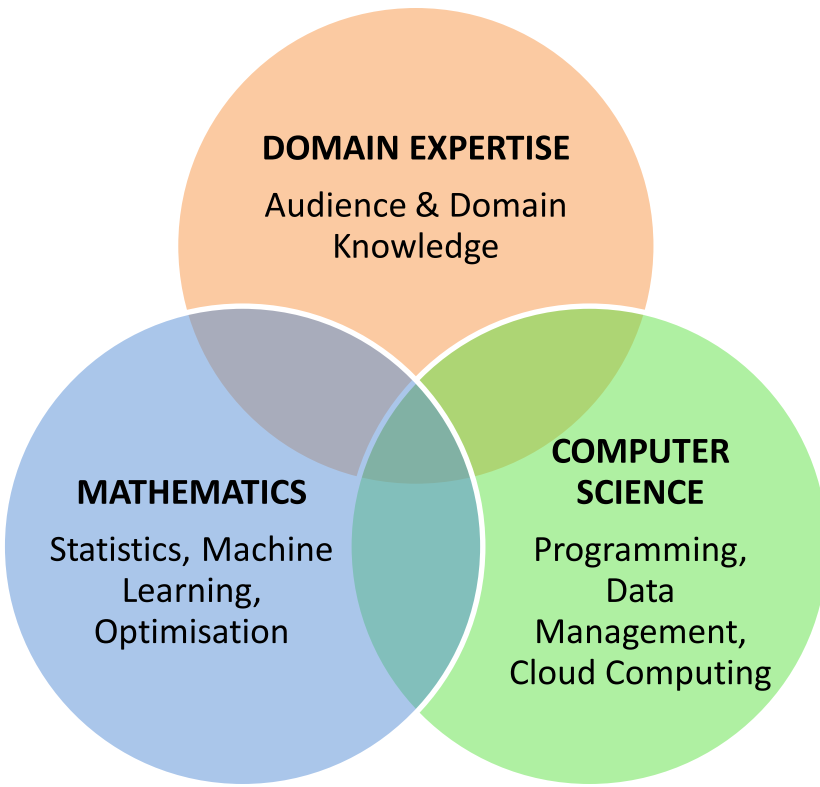
That's comparable with the performance of expert forecasters.

Communicating Results

Communication / Storytelling

Data Visualisation / Graphic Design

The most important skills for conveying your results to a non-technical audience



Finding Encouragement

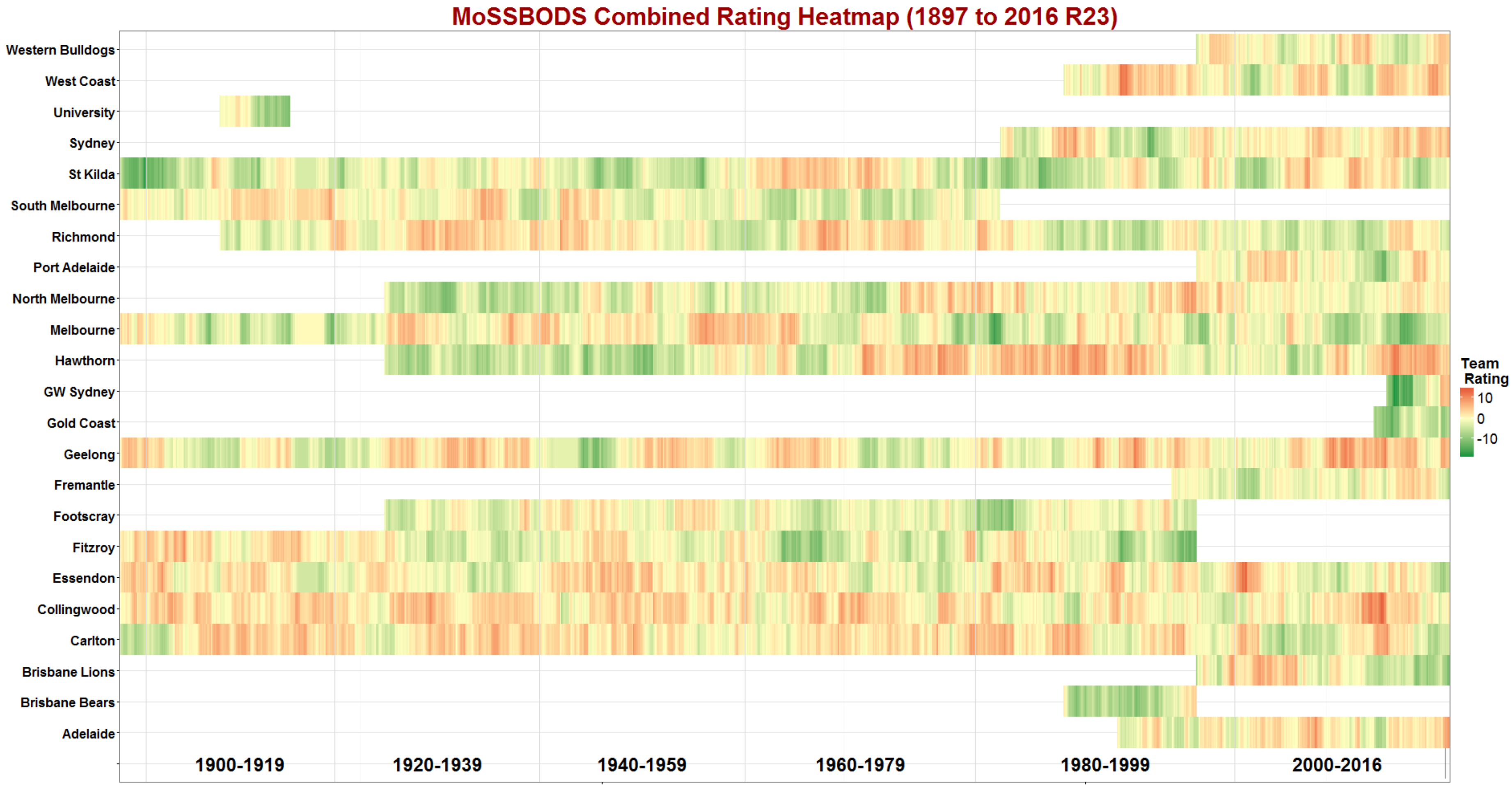
Actual comment from a Big Footy forum member in 2015 …

"It's shaping as a pretty difficult season to predict this year, **so I have dug up this amazing blog with lots of stats and maths talk and shit.**

Anyway, gives you a real insight into how the draw can impact a team's ultimate ladder position …"

… add one to the "compliment" bucket

Historical Ratings: Combined All-Time



This style of presentation highlights:

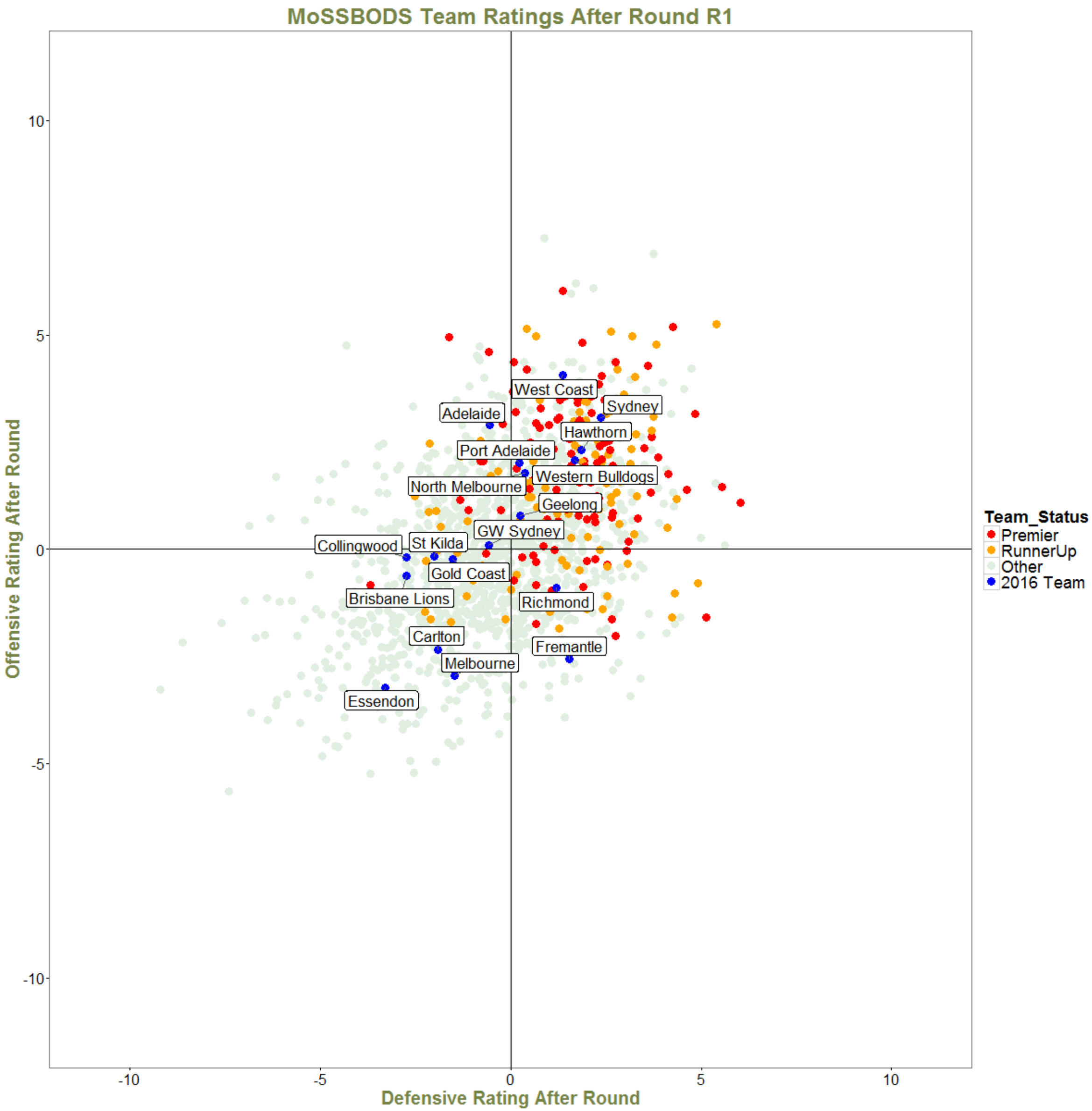
- The tendency for teams to be poor when they start (and finish)
- Team's entry and exit points in the competition
- The extent and depth of periods of above- and below-average ability

The 2016 Season

Animating your outputs is a very effective way to increase interest and understanding for your audience.

This animation highlights:

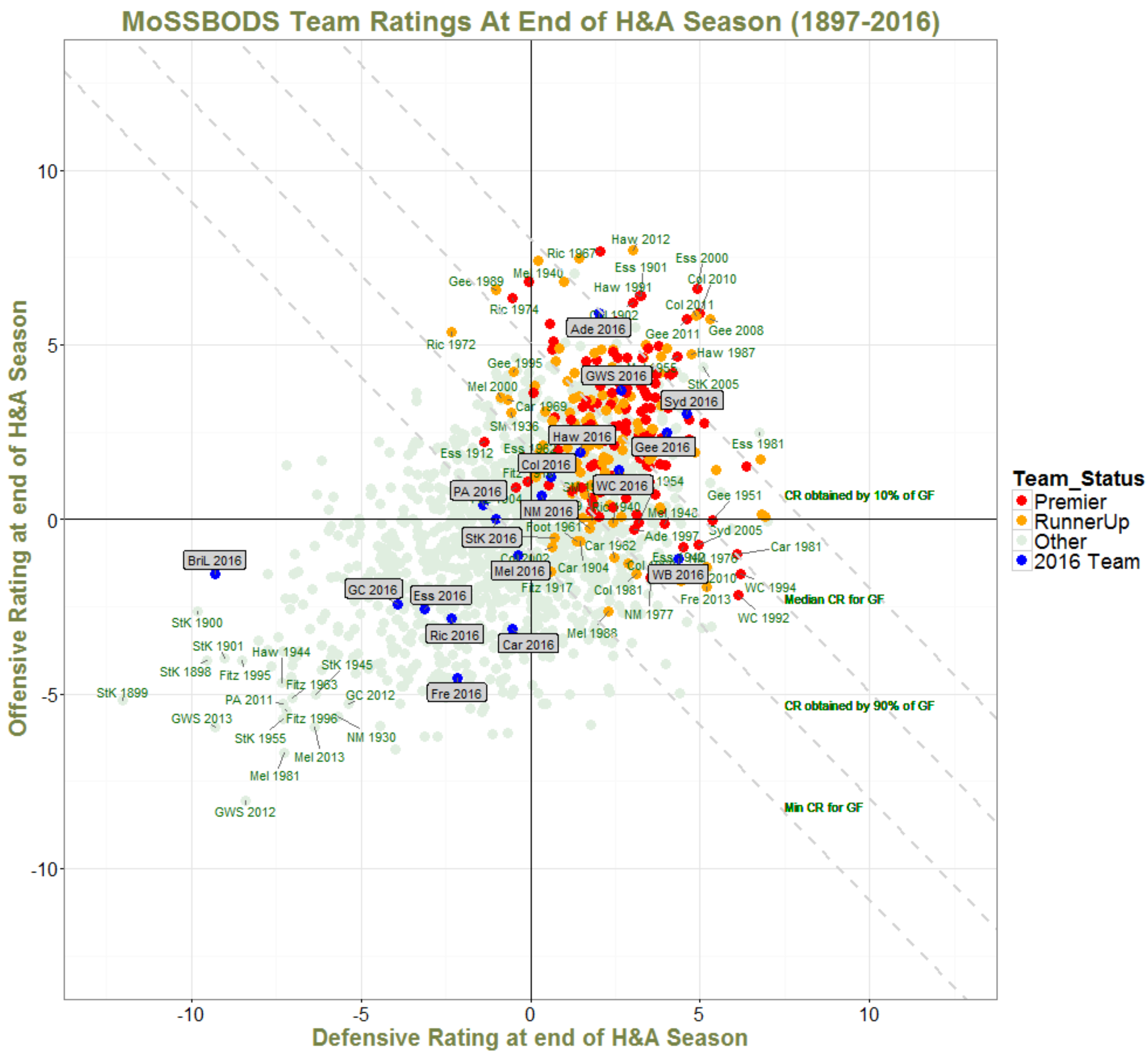
- Ratings spread as the season progresses
- Positive relationship between offensive and defensive ratings
- Wide differences in offensive and defensive abilities of the top teams



End of Home and Away Season (All)

Providing historical context also helps your audience grasp the significance of the most recent data

Here we can see how the 2016 teams compare with previous Grand Finalists at the end of the Home and Away season



Application #2: Predicting Outcomes

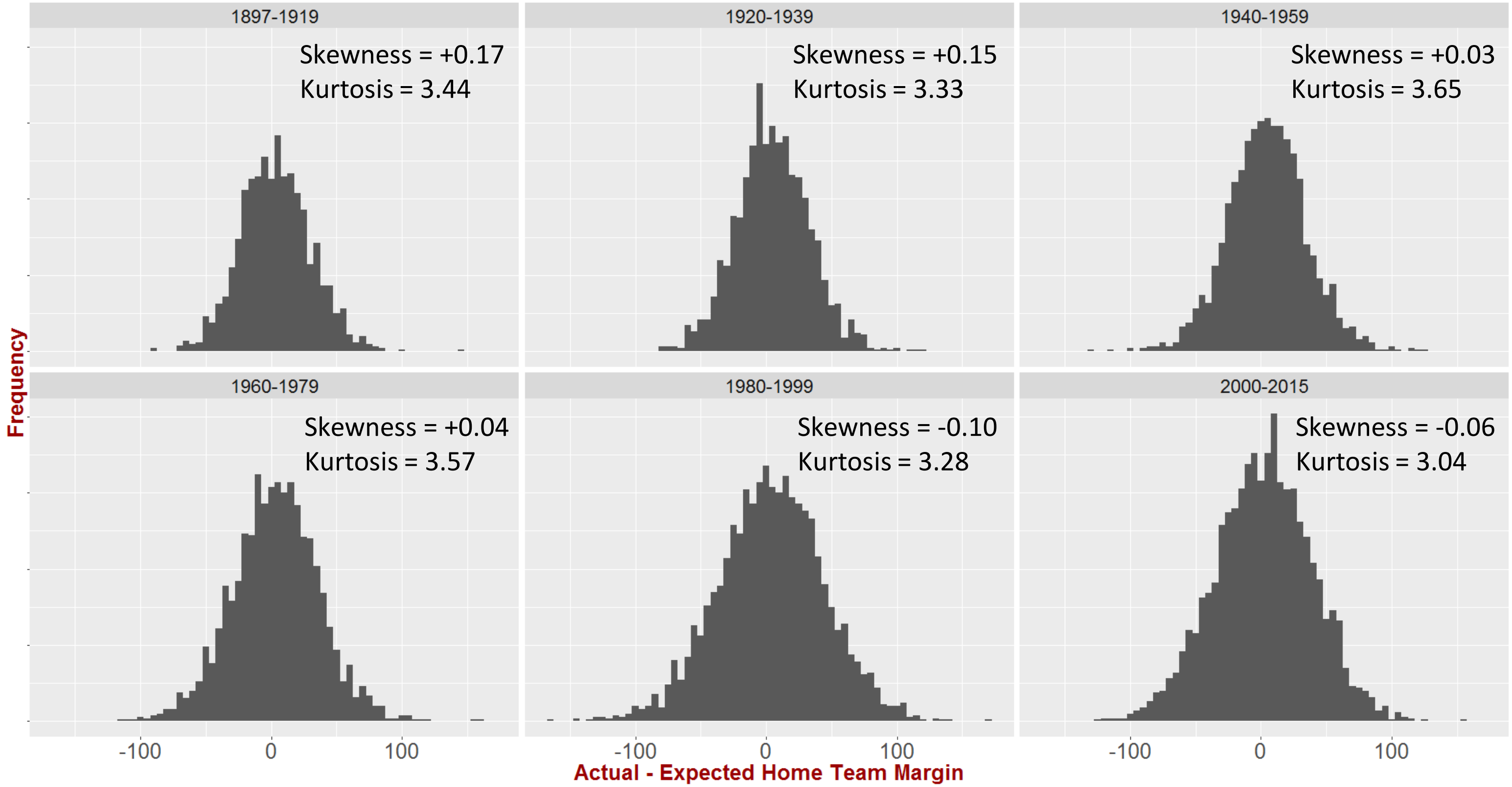
by how much is team X win likely to defeat team Y on Saturday? How confident should I be about that prediction?

will team Z make the finals?

Background

For many years, I'd observed that game margins – the difference between the home and away teams' scores – seemed to approximately follow a Normal distribution once you adjusted for the quality of the teams.

(They did, however, usually fail standard tests of Normality).



I'd always wondered why …

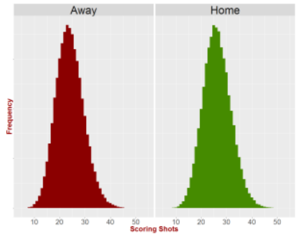
Scoring From First-Principles

We conceptualise scoring as a two-step process

Teams generate Scoring Shots

Bivariate Negative Binomial

model as function of team ratings and venue effects



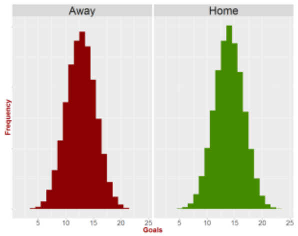
Some are converted to Goals

remainder

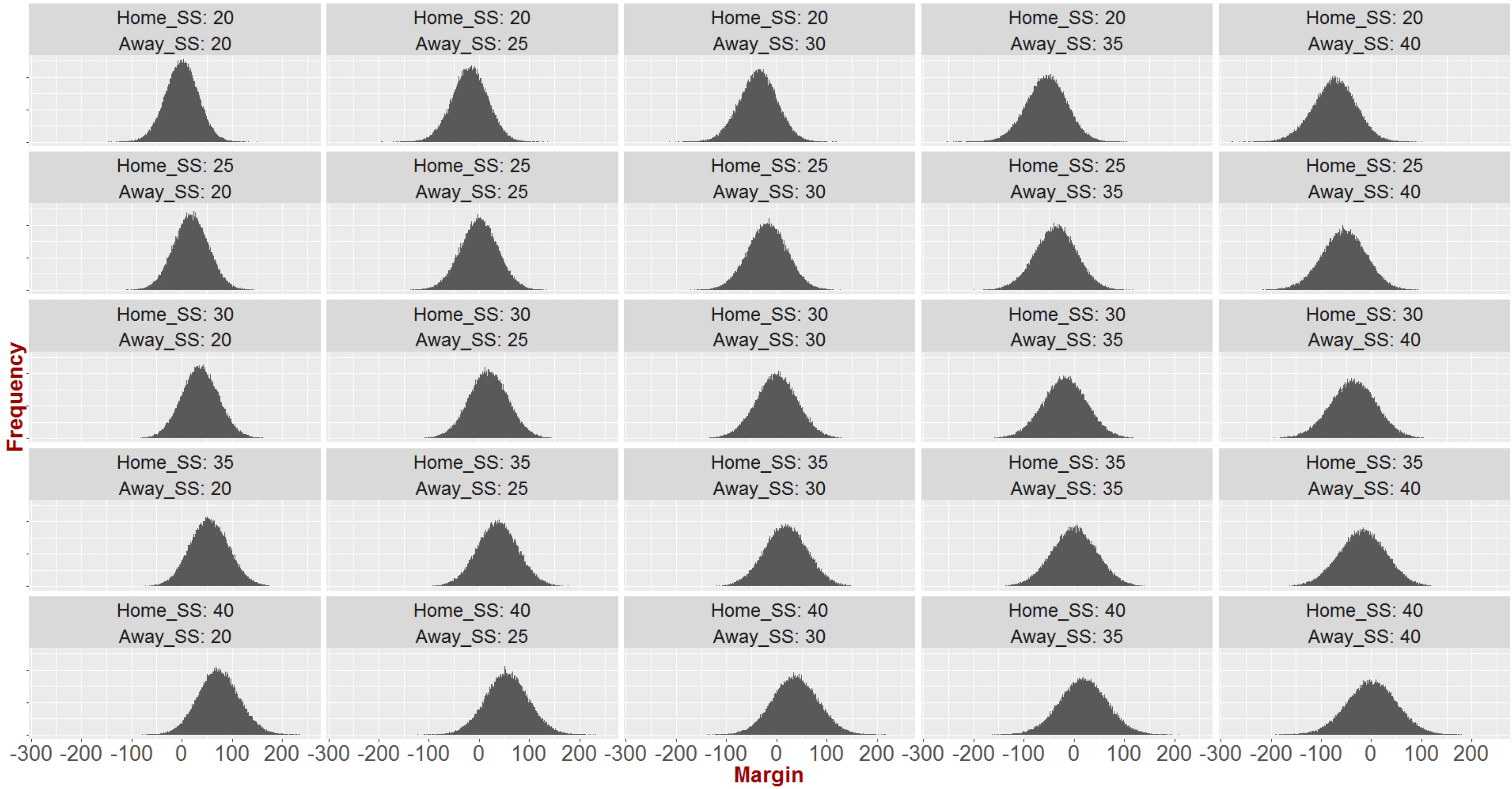
The rest are registered as Behinds

Two Beta Binomials
(Home Team/Away Team)

model using historical Scoring Shot and Goal production data



Modelling Game Margins



Modelled game margins (ie home score – away score) do, indeed, look Normal-like (though they also fail formal tests).

In most scenarios, we have skewness < 0 and we always have kurtosis > 3.

Inferring Single Game Probabilities

We can use the model to estimate victory probabilities depending on the teams' expected Scoring Shot values and assumed expected conversion rates

Estimated Home Team Victory Probability

Home Team Expected Scoring Shots	Away Team Expected Scoring Shots: 20	21	22	23	24	25	26	27	28	29	30
20	51%	46%	42%	38%	34%	30%	26%	24%	22%	18%	16%
21	55%	51%	46%	40%	38%	34%	30%	28%	24%	22%	19%
22	59%	54%	50%	46%	42%	38%	35%	30%	28%	24%	22%
23	63%	60%	55%	51%	46%	43%	38%	35%	32%	28%	25%
24	67%	62%	58%	55%	50%	46%	43%	39%	35%	32%	28%
25	69%	66%	62%	58%	54%	52%	46%	43%	39%	35%	33%
26	74%	70%	65%	62%	58%	55%	50%	47%	43%	39%	36%
27	76%	73%	69%	66%	62%	58%	54%	50%	46%	43%	38%
28	80%	76%	73%	69%	66%	62%	58%	55%	50%	47%	44%
29	83%	78%	76%	72%	68%	65%	62%	57%	54%	51%	47%
30	84%	82%	78%	75%	72%	69%	64%	61%	58%	54%	50%

A Home team expected to generate 26 Scoring Shots will beat an Away team expected to generate 21 Scoring Shots about 70% of the time

Reflections in Passing

The Normal distribution is so ubiquitous and often (at least roughly) applicable that it's easy to forget to stop and consider why it might or might not apply in your particular case.

Modelling a system from first-principles or on a multi-step basis (as we've done here) can lead to deeper insights and better models.

Before fitting your data using some statistical distribution it's worth reviewing the conditions that need to apply for the distribution to be appropriate – even if only approximately.

Application #3: In-Running Predictions

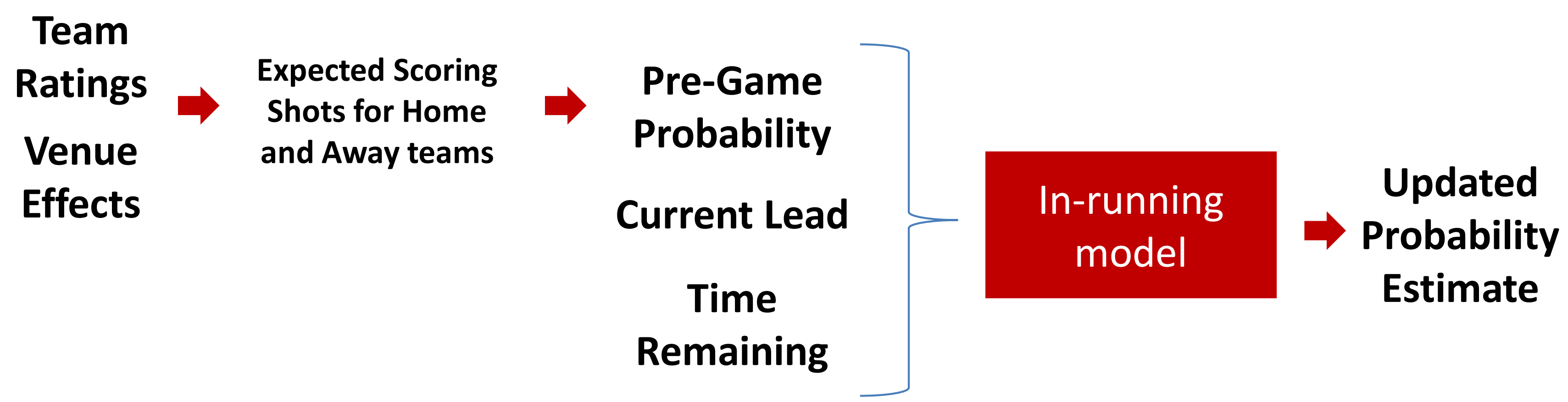
if my team is leading by X points with M minutes to be played, how confident should I be that they'll win?

Background

During the course of a game, to predict its outcome, is it relatively more valuable to know the current score or to know which team was the pre-game favourite and by how much?

How does the value of these pieces of information change as the game progresses?

To answer this, I constructed the following system of models:



Data for the In-Running Model

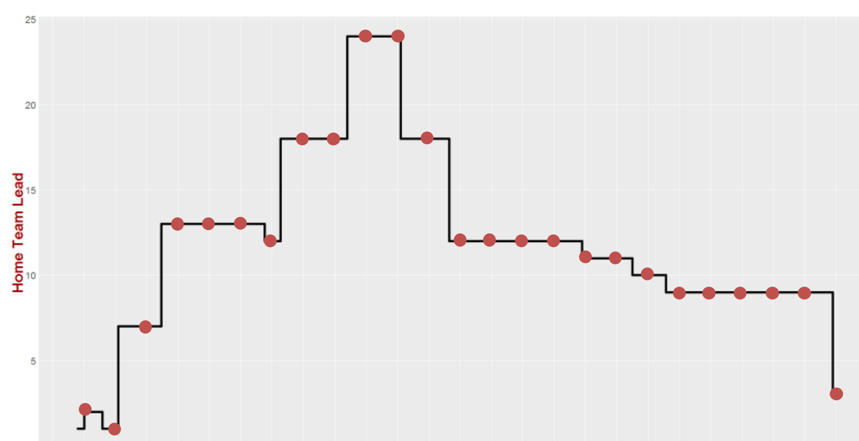
Score progression data is available for the 1,786 games played since 2008. The data is event-based, with each row representing either a score by one of the teams or the end of a quarter.

Zaine Cordy goal	19m 54s	1.0.6 - 1.2.8
Tory Dickson goal	22m 28s	2.0.12 - 1.2.8
		2nd quarter (32m 49s)
Tom Boyd goal	2m 42s	3.0.18 - 1.2.8

The number of events in a game varies from 30 to 76.

So, to ensure that each game carries equal weighting in the final model, we discretise the score progression data by sampling it at 25 equally-spaced time intervals within each quarter of each game.

This yields 100 home team lead values for every game.



In-Running Model

We fit a probit model of the form

$$\boldsymbol{P(Result = 1 \mid X) = \Phi(X'\beta)\ where\ X\ comprises}$$

$$Intercept$$

Adjusts for bias in pre-game probability estimate

$$\frac{Home\ Lead}{Fraction\ of\ Game\ Remaining^{k_1}}$$

Ensures that leads of a given size are "safer" as the remaining time reduces

$$Fraction\ of\ Game\ Remaining^{k_2}$$

Models change in importance of home team status as game progresses

$$\frac{PreGame\ Prob}{Fraction\ of\ Game\ Remaining^{k_3}}$$

Models diminishing importance of pre-game probability estimate as game progresses

$$Home\ SS\ Run^{k_4}$$

$$Away\ SS\ Run^{k_5}$$

Models "momentum" effects of the home / away team registering consecutive Scoring Shots within a quarter

Functional Form

We choose $\boldsymbol{\beta}$ and $k1$ to k_5 to minimise AIC

```
Coefficients:
                                           Estimate Std. Error z value Pr(>|z|)    
(Intercept)                               0.1937807  0.0174271  11.120  < 2e-16 ***
I(Home_Lead/(1 - Game_Fraction)^0.5)      0.0318349  0.0001957 162.631  < 2e-16 ***
I((1 - Game_Fraction)^0.5)               -1.7309305  0.0268917 -64.367  < 2e-16 ***
I(Probability_of_Victory * (1 - Game_Fraction)^0.6)  2.9886137  0.0248869 120.088  < 2e-16 ***
I(Home_Run_SS^2)                          0.0062331  0.0007083   8.800  < 2e-16 ***
I(Away_Run_SS^1.5)                       -0.0046371  0.0016301  -2.845  0.00445 ** 
```

There is some weak evidence for in-game momentum in that teams registering streaks of scoring increase their victory probability as the streak extends

The choice of functional form was inspired by this 2007 paper

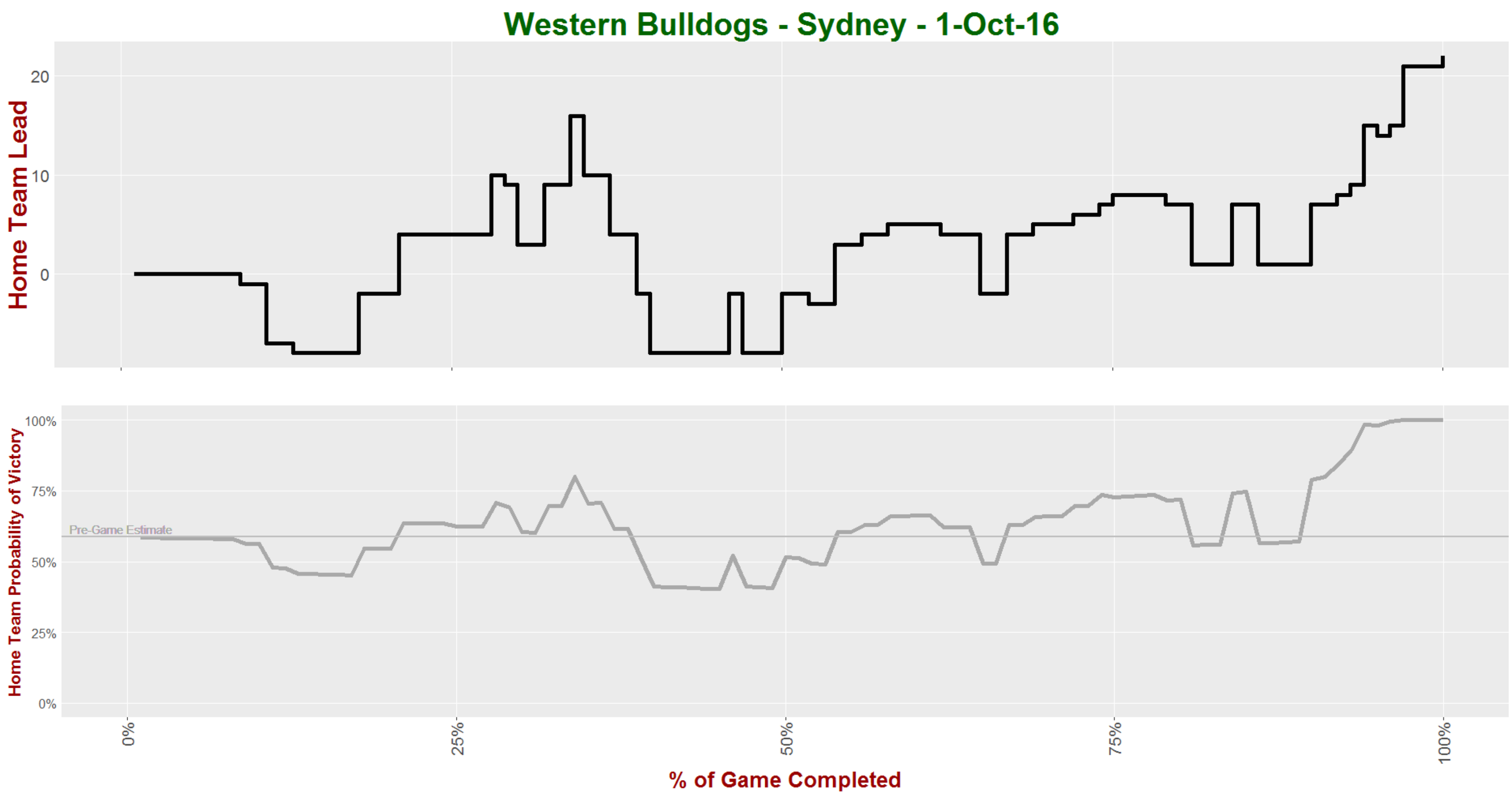
A Brownian Motion Model for the Progress of Sports Scores

Hal S. Stern

Journal of the American Statistical Association, Vol. 89, No. 427. (Sep., 1994), pp. 1128-1134.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199409%2989%3A427%3C1128%3AABMMFT%3E2.0.CO%3B2-W

In-Use Example: 2016 Grand Final



With the model, we can estimate the home team's victory probability at any point during a game.

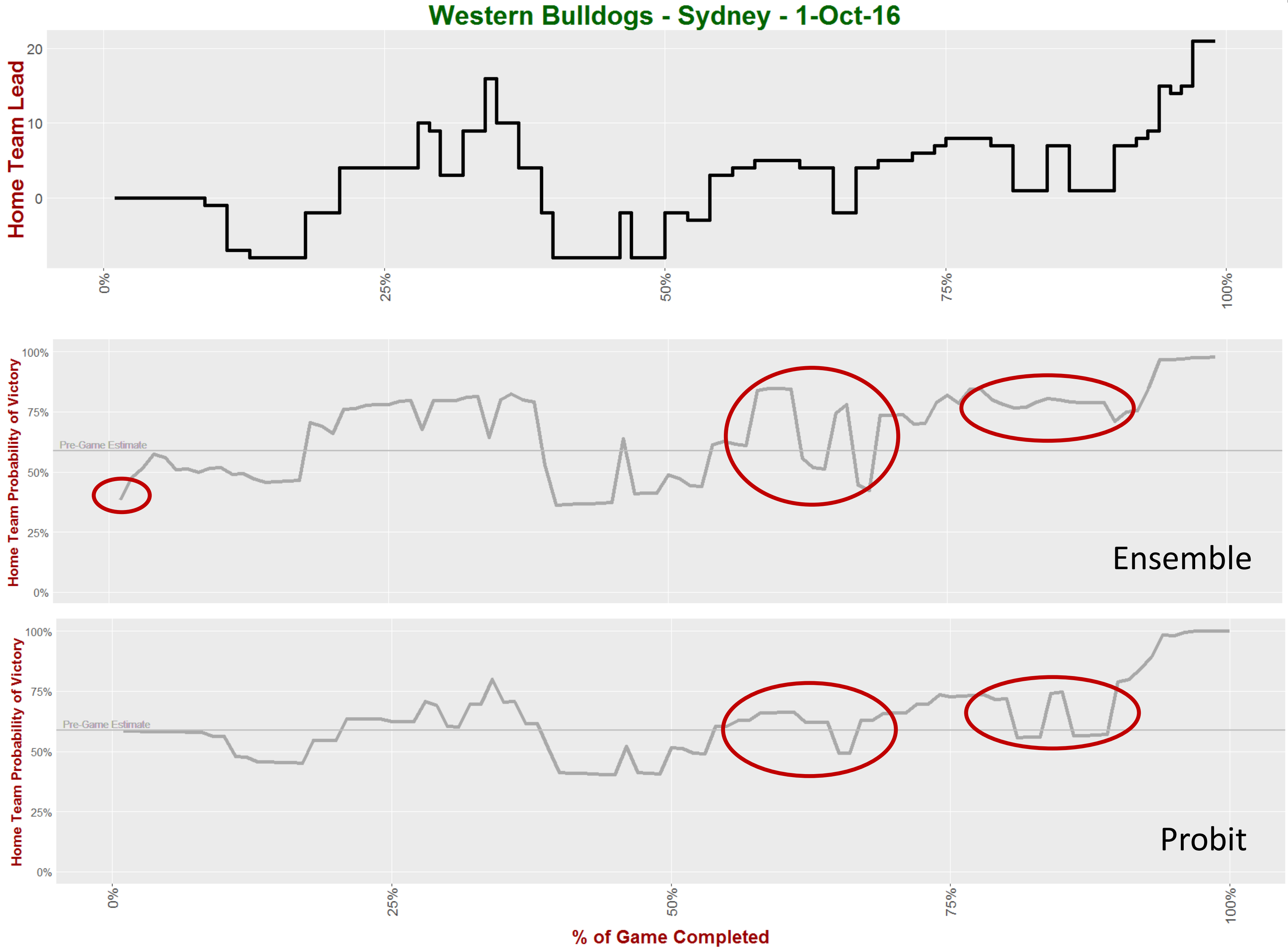
What About An Ensemble?

We might also try an ensemble comprising models from the following categories:

- k-nearest neighbour
- discriminant analysis
- tree
- generalised linear
- rule-based
- regression models

As regressors, we use the same transformed variables as we did for the probit model.

Ensemble vs Probit



A Case for Imposing a Functional Form

Getting the right functional form for the in-running model is vital to producing useful in-running predictions.

I’ve tried, without success, to build a competitor ensemble model (as you’ve just seen).

Maybe it’s just me, but they seem to suffer without the hard constraint of a “sensible” functional form.

Sometimes, not explicitly choosing a functional form isn’t ideal.

Application #4: Classifying Games

how many types of Grand Final are there?

What Do You Remember About a Game?

What are the most salient features of a sporting context?

- Who wins?
- By how much?
- How close the game was throughout?
- How often the lead changed?

So, we cluster the 117 V/AFL Grand Finals in which a result was achieved using seven variables

- Whether or not the team that eventually won was ahead at the end of the 1st, 2nd, and 3rd quarters (3 variables)
- What the winning team's lead was at the end of the 1st, 2nd, 3rd, and Final quarters (4 variables)

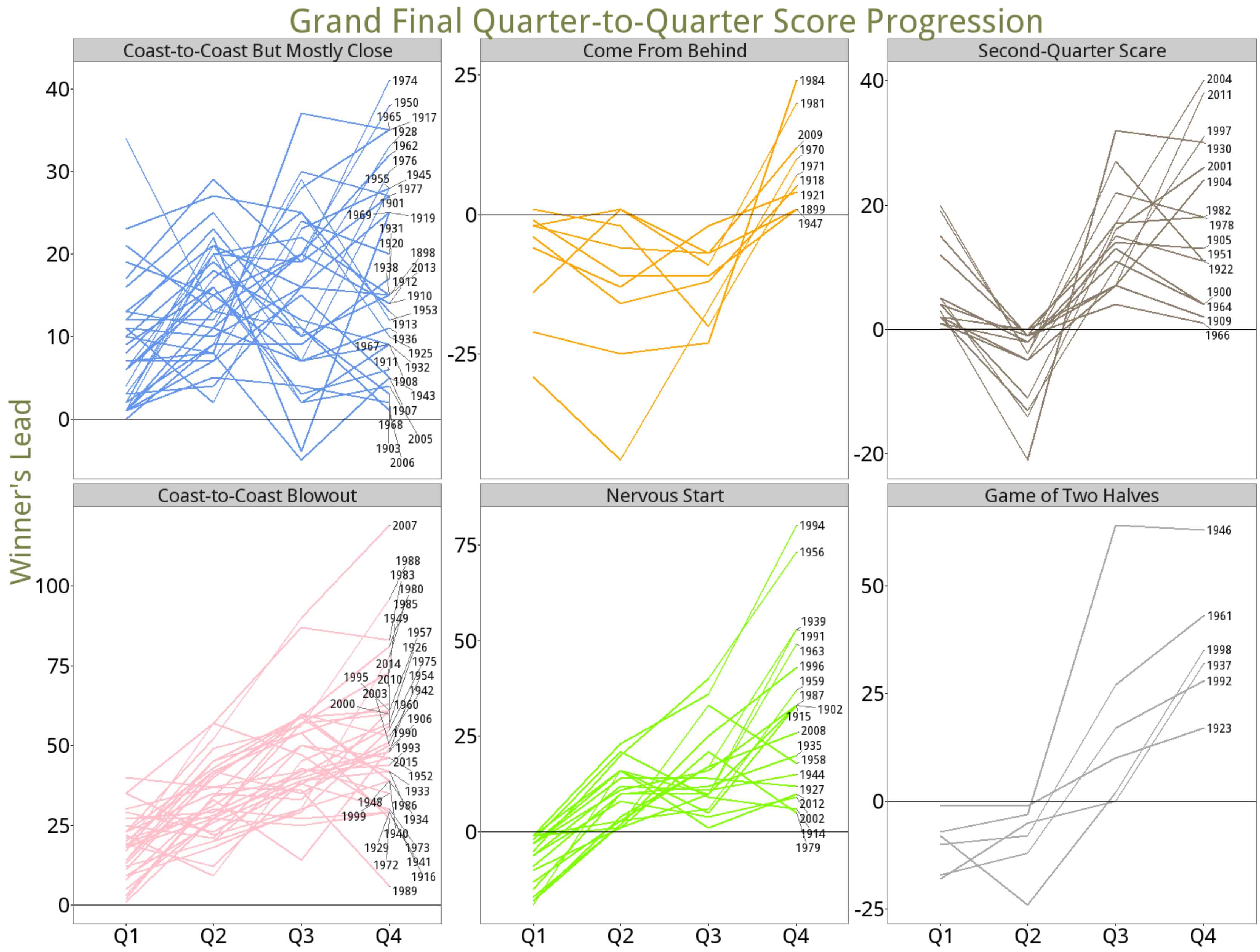
On the Naming of Clusters

Type of Grand Final	# of GFs of this type	Games of this Type in Which Winner Leads or is Tied At ...			Mean Winning Team Lead At ...				Proportion of Games from Era of Type					
		QT	HT	3QT	QT	HT	3QT	FT	1897-1919	1920-1944	1945-1969	1970-1994	1995-2015	All
Coast-to-Coast But Mostly Close	34				8.6	14.6	15.0	18.0						
Come-From-Behind	9				-8.7	-12.8	-12.0	9.3						
Second-Quarter Scare	15				6.3	-5.6	14.6	18.1						
Nervous Start	19				-7.9	10.2	15.6	32.0						
Coast-to-Coast Blowout	34				18.2	32.9	46.0	51.9						
The Game of Two Halves	6				-10.2	-8.8	20.0	36.3						
All Grand Finals	**117**	**72%**	**78%**	**91%**	**6.1**	**13.3**	**22.2**	**30.4**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Cluster naming is an important but difficult task

- Memorable but not ‘deceptive’ names
- Encapsulate the most salient features of the cluster without misrepresenting any of the cluster members to too large an extent

The Six Types of Grand Final



Some Thoughts About Clustering

Rarely, if ever, is there a truly “best” cluster solution on real-word data

- Combine some of the standard quality/validity metrics (eg gap, silhouette width, Dunn Index, connectivity, average proportion of non-overlap, average distance etc) with a practical view of what a “sensible” solution should provide

Need the solution to provide a narrative that’s

- Faithful to the data
- Interesting
- Useful

Cluster membership is rarely binary

- Always “edge cases”
- When naming clusters, consider the archetypal cases and then ensure the names don’t misrepresent the outliers too much

Needs judgement and domain expertise on the part of the modeller

Concluding Thoughts

A Bias For Action

A model’s strengths and weaknesses are only ever fully revealed with real world use.

Be biased, where possible, towards deploying and then refining a model rather than undertaking more tuning using existing data.

You can’t start meaningfully improving your model until you’ve got a real-world measure of its performance.

The Role of the Modeller/Analyst

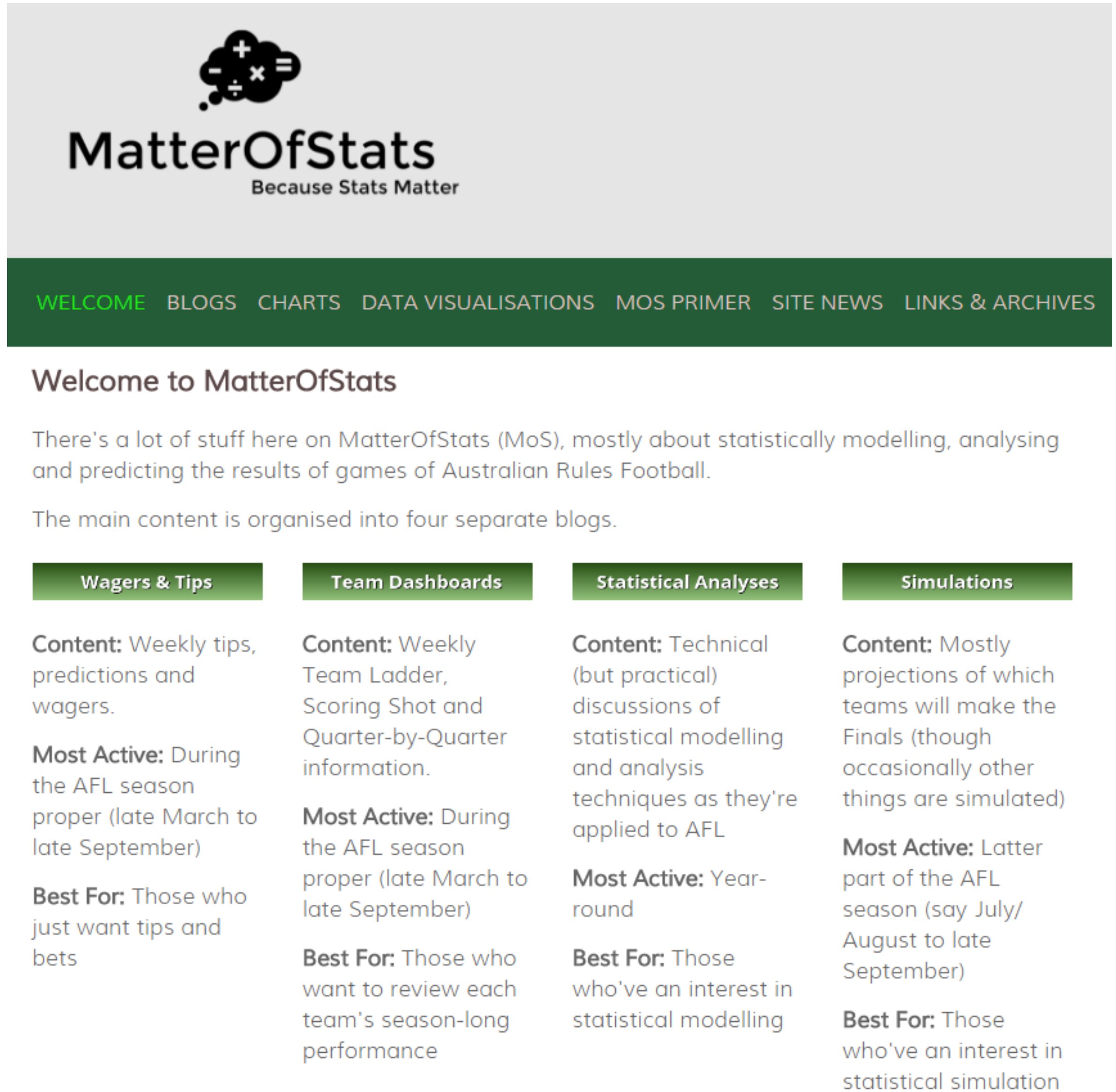
It’s hard to separate the model from the modeller who created it.

The final model summarises

- the data selected (and transformed) by the modeller
- in a manner chosen by the modeller
- given his/her domain and technical expertise
- to help answer a particular interesting question as understood by the modeller

MatterOfStats

- Operating since 2006
- Rate teams
- Predict outcomes
- Project seasons
- Analyse history and results



MatterOfStats
Because Stats Matter

WELCOME BLOGS CHARTS DATA VISUALISATIONS MOS PRIMER SITE NEWS LINKS & ARCHIVES

Welcome to MatterOfStats

There's a lot of stuff here on MatterOfStats (MoS), mostly about statistically modelling, analysing and predicting the results of games of Australian Rules Football.

The main content is organised into four separate blogs.

Wagers & Tips

Content: Weekly tips, predictions and wagers.

Most Active: During the AFL season proper (late March to late September)

Best For: Those who just want tips and bets

Team Dashboards

Content: Weekly Team Ladder, Scoring Shot and Quarter-by-Quarter information.

Most Active: During the AFL season proper (late March to late September)

Best For: Those who want to review each team's season-long performance

Statistical Analyses

Content: Technical (but practical) discussions of statistical modelling and analysis techniques as they're applied to AFL

Most Active: Year-round

Best For: Those who've an interest in statistical modelling

Simulations

Content: Mostly projections of which teams will make the Finals (though occasionally other things are simulated)

Most Active: Latter part of the AFL season (say July/ August to late September)

Best For: Those who've an interest in statistical simulation

Appendix

Scoring Shot Model

Scoring Shot Model

Historical Home team and Away team Scoring Shot Production Data (2000-2016)

Bivariate Negative Binomial

Model as a function of team ratings and venue effects

- *Coefficient estimates*
- *Estimate of variance-covariance matrix for scoring shot production*

cor(Excess Home SS, Excess Away SS) = -0.23

Parameters Relating to Home Team Scoring Shot Production

log(Home Team Mean SS)	3.248	***
log(Home Team Size Parameter)	5.161	***
Venue Effect	0.0014	
Travel Penalty	0.0124	***
Home Team Offensive Rating	0.0280	***
Home Team Defensive Rating	0.0075	***
Away Team Offensive Rating	-0.0052	***
Away Team Defensive Rating	-0.0302	***

Mean SS = exp(3.248) = 25.7

(measure of variability in SS production)

SS production most dependent on own offensive and opponent defensive ability

Model explains about 25-28% of the variability in Scoring Shots for both teams

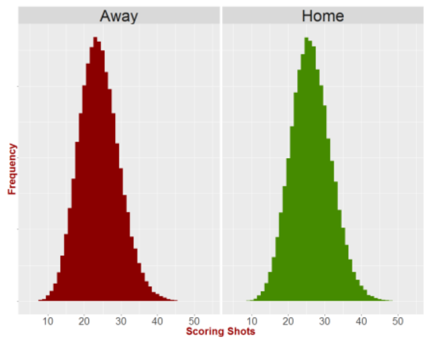
Parameters Relating to Away Team Scoring Shot Production

log(Away Team Mean SS)	3.210	***
log(Away Team Size Parameter)	4.653	***
Venue Effect	-0.016	***
Travel Penalty	-0.017	***
Home Team Offensive Rating	-0.0035	
Home Team Defensive Rating	-0.0369	***
Away Team Offensive Rating	0.0308	***
Away Team Defensive Rating	0.0055	***

Mean SS = exp(3.210) = 24.8

(measure of variability in SS production)

SS production most dependent on own offensive and opponent defensive ability



Example assuming both team's offensive and defensive ratings are 0

Goals Model

Teams convert some number of Scoring Shots into Goals

Historical Home & Away Team Goal scoring (2000-2016) → *Beta Binomial* — Model using Home/Away team Scoring Shot and Goal data →

- *Estimate of theta parameter for Home and Away teams*

Beta Binomial assumes that the conversion rate for a game is drawn from a Beta distribution. That conversion rate is then used as the probability in a standard Bernoulli distribution.

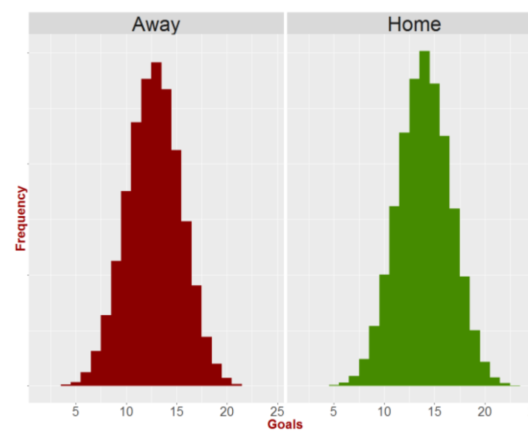
Parameters Relating to Home Team Scoring Shot Conversion

Mean Home Team Conversion Rate	53.7% ***
Theta	261.9 ***

Parameters Relating to Away Team Scoring Shot Conversion

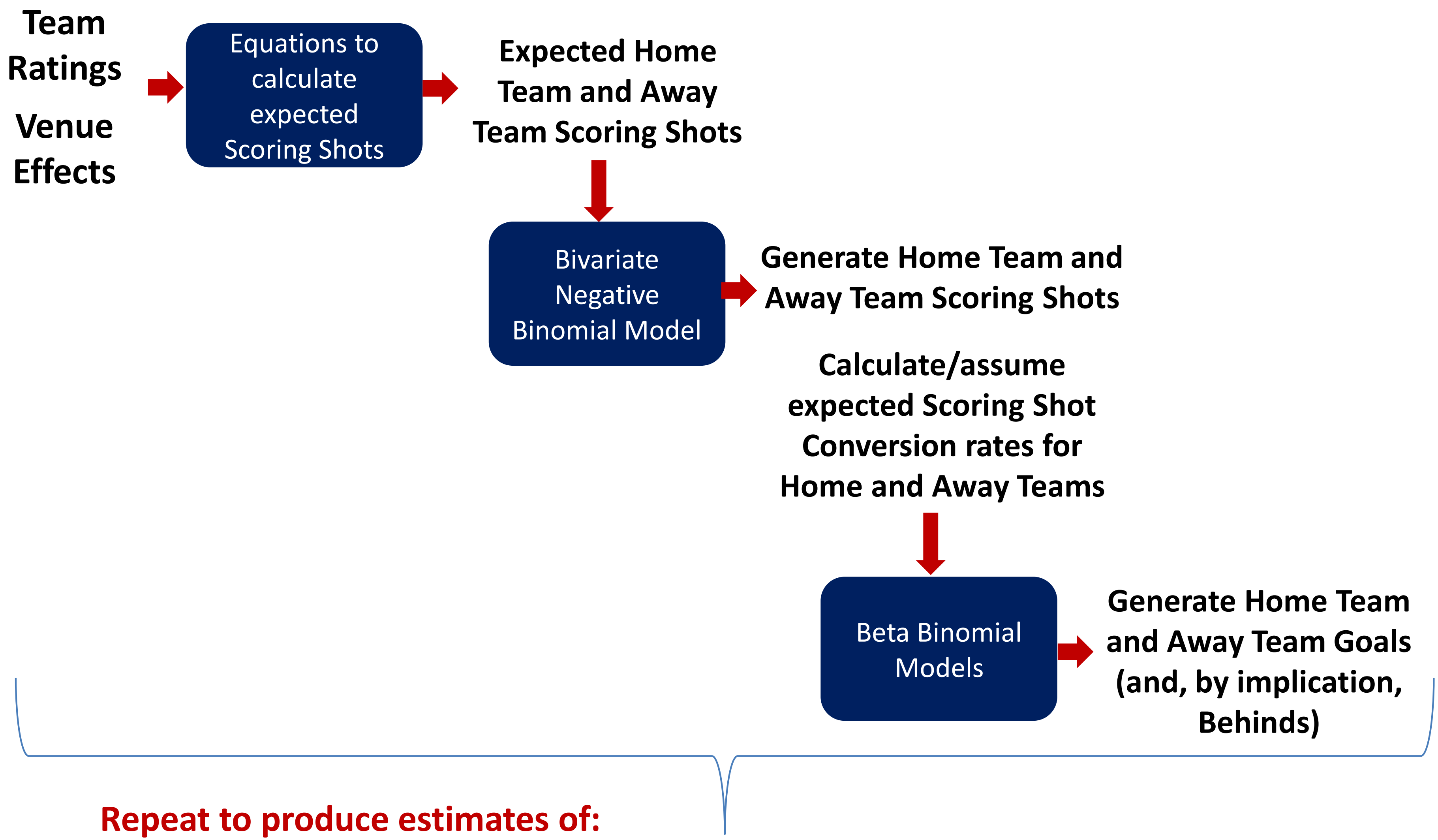
Mean Away Team Conversion Rate	53.2% ***
Theta	108.9 ***

Models explain about 66% of the variability in Scoring Shot Conversion for both teams



Example where Home Team produces 26 SS / expected Conversion of 53.7%, Away Team produces 24 SS / expected Conversion of 53.2%

Simulating a Single Game



Repeat to produce estimates of:

- **Home Team Win Probability**
- **Expected Margin**
- **Likelihood of a Draw**
- **Likelihood of a given victory margin**
- **Likelihood of a given number of Scoring Shots**

Analysing Sport vs Corporate Data

WHAT'S DIFFERENT

Analysing and modelling sporting data

- provides an opportunity to create **well-defined** problems with **clear and objective** measures of success.
- often requires answers that are **probabilities**; here such answers aren't treated as statements of ignorance and timidity.

WHAT'S SIMILAR

In analysing and modelling sporting data, as in business

- there is **accepted wisdom** that is amenable to data-driven validation (and that's often wrong).
- the **precise question you ask** and **what you intend to do with the answer** should drive the analysis and modelling process