Rugby Match Predictions

Trevor Ceniviva

The Project

- Based on previous matches in a season, can we predict who will win a match with known opponents

Inspiration

- Dan Gillick, UC Berkeley PhD, Research Analyst at Google
- "How to Predict College Football Games"
  - Aimed to reduce minimum score difference between prediction and actual
    - week 1: x vs. y result: 21-3
    - week 2: y vs. z result: 15-18
    - week 3: x vs. z result: 10-42

$$[(Sx - Sy) - (21 - 3)]^2 + [(Sy - Sz) - (15 - 18)]^2 + [(Sx - Sz) - (10 - 42)]^2$$

Raw Data

- 5 seasons of matches played in the Aviva Premiership (England's top league)
- Gathered from espnscrum.com
- 23 columns, 660 games

AVIVA PREMIERSHIP - Salford, 25 April 2015, 14:00 local, 13:00 GMT

Like 0

Sale Sharks (6) 23 - 25 (22) Harlequins (FT)

Match coverage: Report

TIMELINE | **NOTES**

Time	Sale Sharks	Score	Harlequins
0	start of first half		start of first half
4		0 - 5	Walker - try
19	Cipriani - penalty goal	3 - 5	
24		3 - 10	Robshaw - try
25		3 - 12	Evans - conversion
27		3 - 17	Yarde - try
32	Cipriani - penalty goal	6 - 17	
35			Matthews - sub on Robson - sub off
39		6 - 22	Care - try
40+1	end of first half		end of first half
40	start of second half		start of second half
43		6 - 25	Evans - penalty goal
49	Taylor - sub on Jones - sub off Lewis-Roberts - sub off Harrison - sub on Seymour - sub on Ioane - sub off - sub off Arscott - sub on		
50			Marfo - sub on - sub off
51			Ward - yellow card
52	Arscott - try	11 - 25	
53	Cipriani - conversion - sub off Mills - sub on	13 - 25	
54			Botica - sub on Evans - sub off
55	Cusiter - sub off Cliff - sub on		Wallace - sub off Buchanan - sub on

TEAMS | **MATCH STATS** | **SALE STATS** | **HQUIN STATS** | **OTHER SCORES**

Sale Sharks		Harlequins
3	Tries	4
1 from 3	Conversion goals	1 from 4
2 from 2	Penalty goals	1 from 1
60.0%	Kick at goal success	40.0%
0	Dropped goals	0
	Kick/pass/run	
19	Kicks from hand	23
157	Passes	185
115	Runs	140
613	Metres run with ball	507
	Attacking	
43% (0%/100%)	Possession (1H/2H)	57% (100%/0%)
15	Clean breaks	7
21	Defenders beaten	20
11	Offloads	22
77 from 82 (93.9%)	Rucks won	87 from 94 (92.6%)
0 from 0	Mauls won	0 from 0
17	Turnovers conceded	14
	Defensive	
136/20	Tackles made/missed	122/21
87.0%	Tackling success rate	85.0%
	Set pieces	
7 won, 1 lost (87.5%)	Scrums on own feed	8 won, 1 lost (88.9%)
12 won, 2 lost (85.7%)	Lineouts on own throw	10 won, 0 lost (100.0%)
	Discipline	
9	Penalties conceded	15
0/0	Yellow/red cards	1/0

Sale Sharks		Harlequins
3	Tries	4
1 from 3	Conversion goals	1 from 4
2 from 2	Penalty goals	1 from 1
60.0%	Kick at goal success	40.0%
0	Dropped goals	0
	Kick/pass/run	
19	Kicks from hand	23
157	Passes	185
115	Runs	140
613	Metres run with ball	507
	Attacking	
43% (0%/100%)	Possession (1H/2H)	57% (100%/0%)
15	Clean breaks	7
21	Defenders beaten	20
11	Offloads	22
77 from 82 (93.9%)	Rucks won	87 from 94 (92.6%)
0 from 0	Mauls won	0 from 0
17	Turnovers conceded	14
	Defensive	
136/20	Tackles made/missed	122/21
87.0%	Tackling success rate	85.0%
	Set pieces	
7 won, 1 lost (87.5%)	Scrums on own feed	8 won, 1 lost (88.9%)
12 won, 2 lost (85.7%)	Lineouts on own throw	10 won, 0 lost (100.0%)
	Discipline	
9	Penalties conceded	15
0/0	Yellow/red cards	1/0

- Pretty messy tables here
  - “# of #”
  - “# from #”
  - “won”
  - “lost”
  - (%)

Cleaning

- Tackles made/missed
  - 136/20
  - Split into ‘tackles_made’ and ‘tackles_missed’
- Conversion goals
  - 1 from 3
  - Split into ‘conversions_made’ and ‘conversions_missed’

Cleaning

- Mauls & Rucks
  - 1 from 2 (50.0%)
  - Split into mauls_started, mauls_won, and maul_success; rucks_started, rucks_won, and ruck_success
- Scrums & Lineouts
  - 7 won, 1 lost (87.5%)
  - Split into scrums_won, scrums_lost, scrum_success; lo_won, lo_lost, lo_success

Cleaning

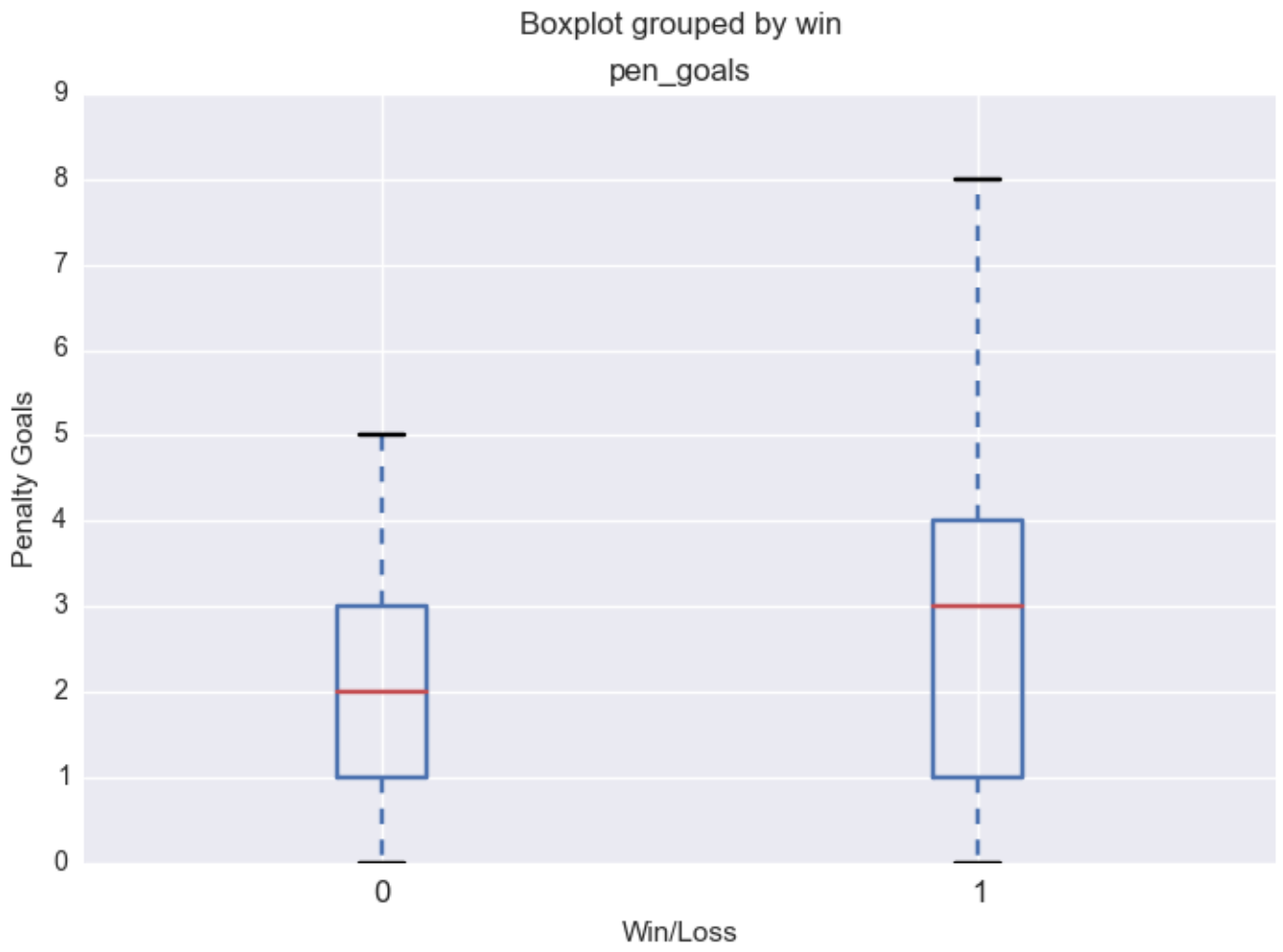
- Penalties
  - 10 (1)
  - Split into penkicks and freekicks
- Penalty goals
  - 2 from 2
  - Split into pen_goals, pen_attempt
- Drop goals
  - Split into drop_goals, drop_missed
- Tries
  - Some of these had a note of (1 penalty try) that had to be removed

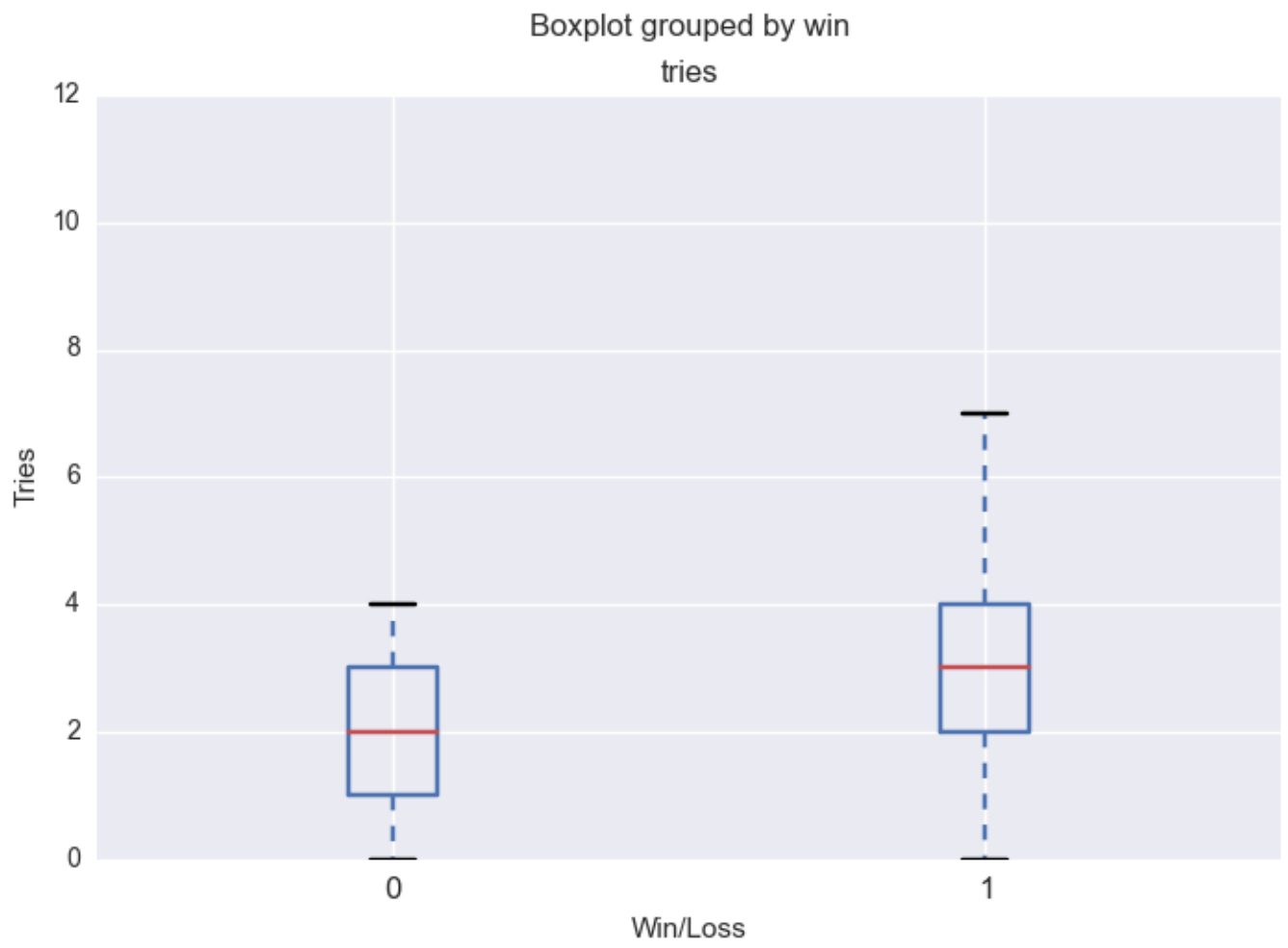
New Columns

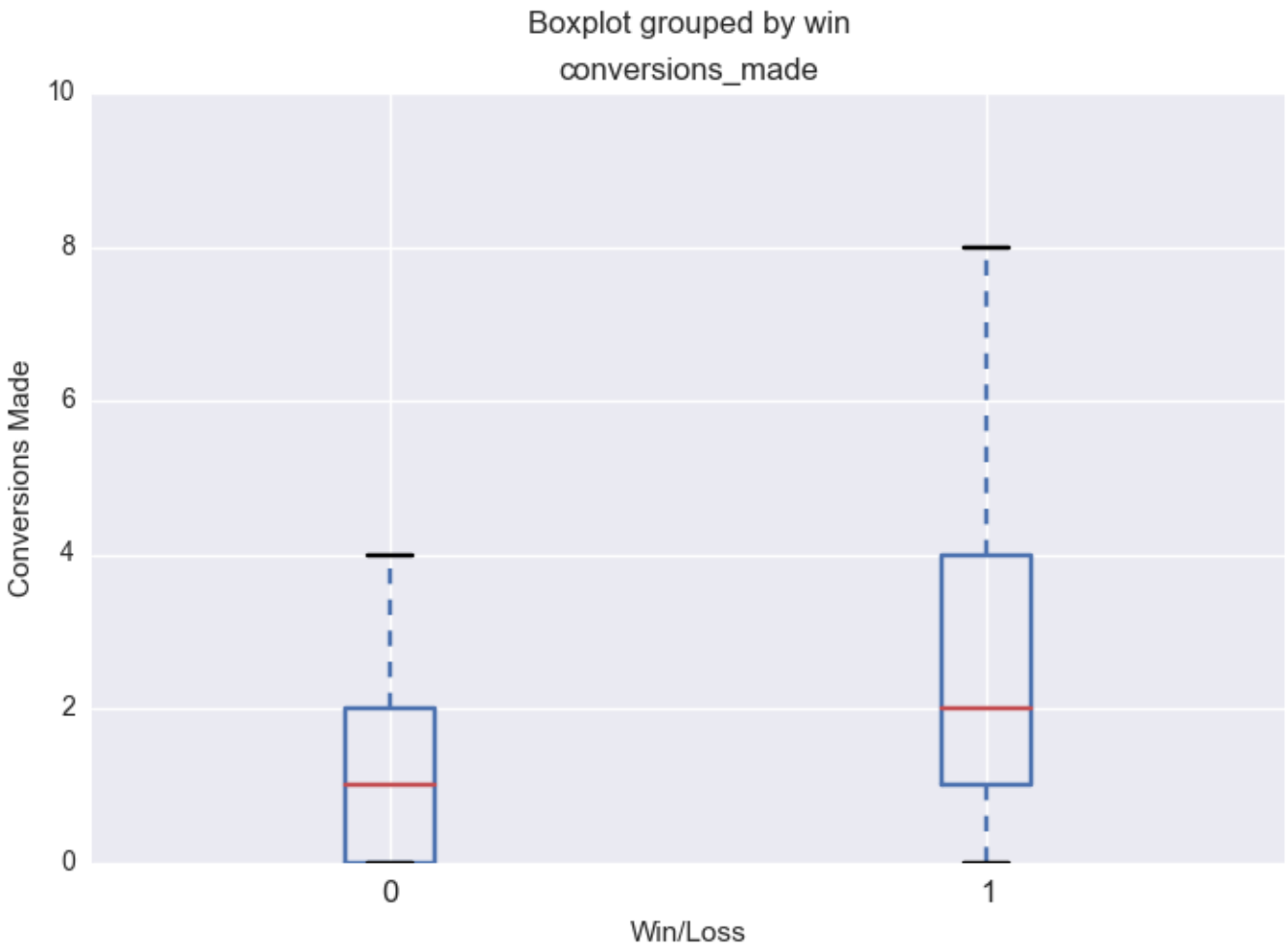
- points
  - Made up of new columns: try_points, pen_points, con_points
- meters_per_run
  - Made by dividing meters_run / runs

Visualizations

- Tons of variables to compare
- Do I want to predict Win T/F?


Boxplot grouped by win
pen_goals
Penalty Goals
9
8
7
6
5
4
3
2
1
0
0
1
Win/Loss


Boxplot grouped by win
tries
12
10
8
6
4
2
0
Tries
0
1
Win/Loss

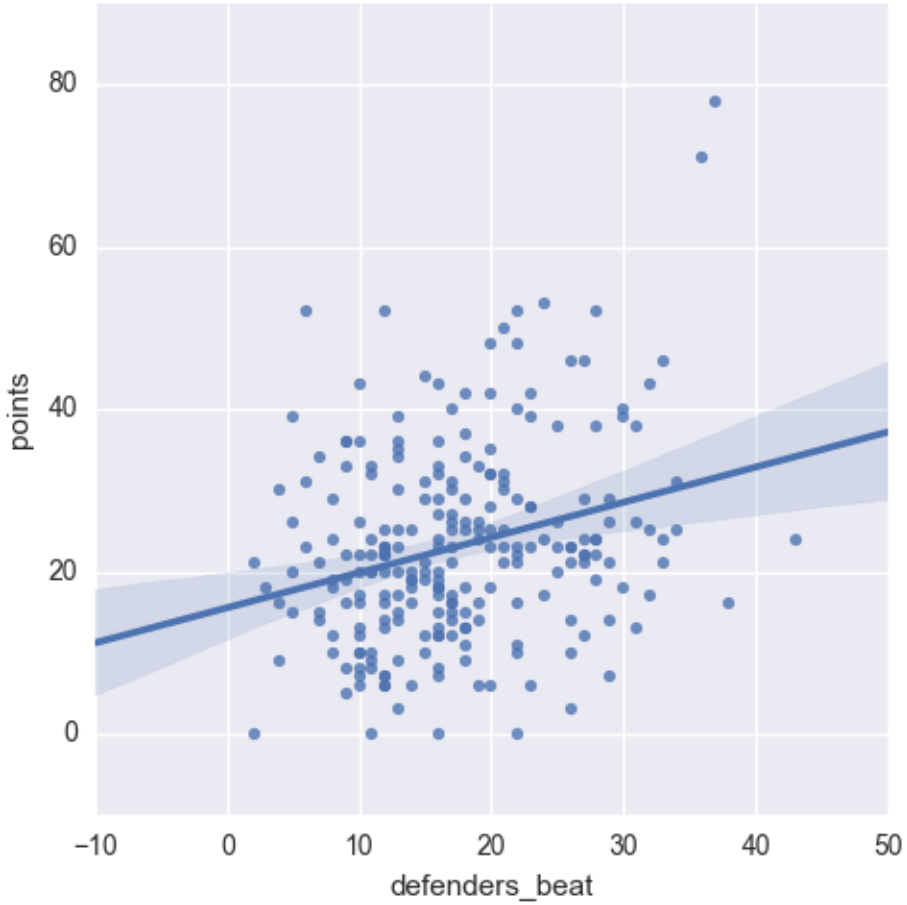

Boxplot grouped by win
conversions_made
Conversions Made
10
8
6
4
2
0
0
1
Win/Loss

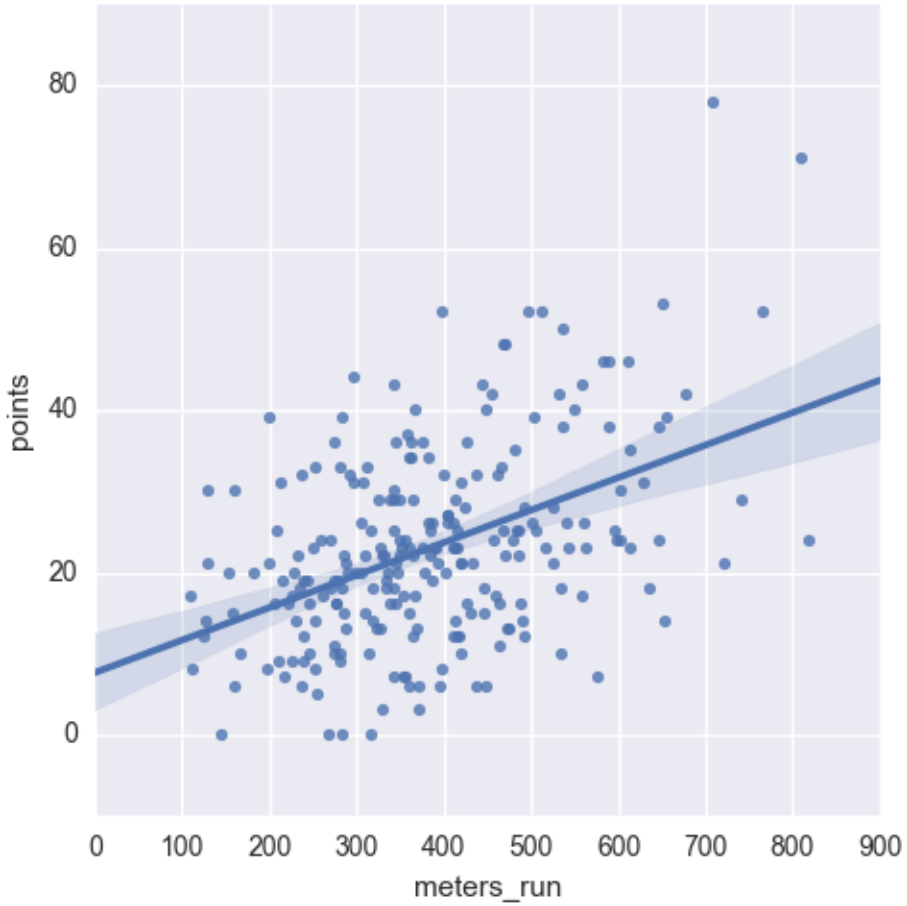
Visualizations

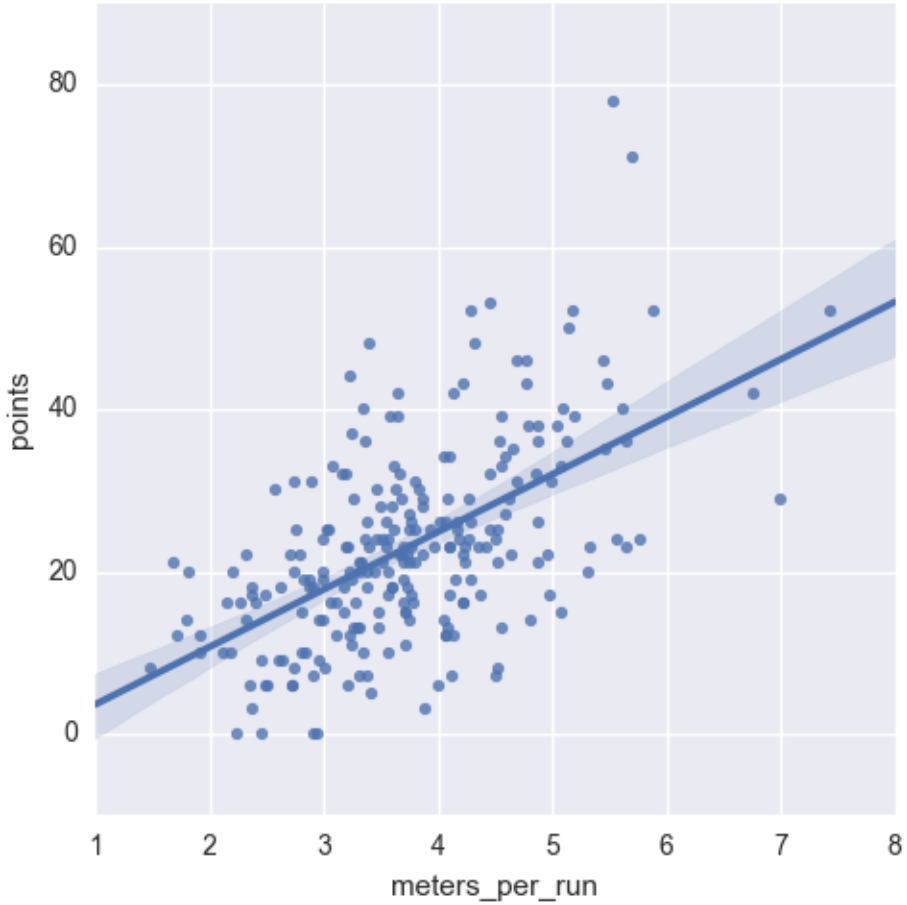
- Do I want to look at “team strength”?

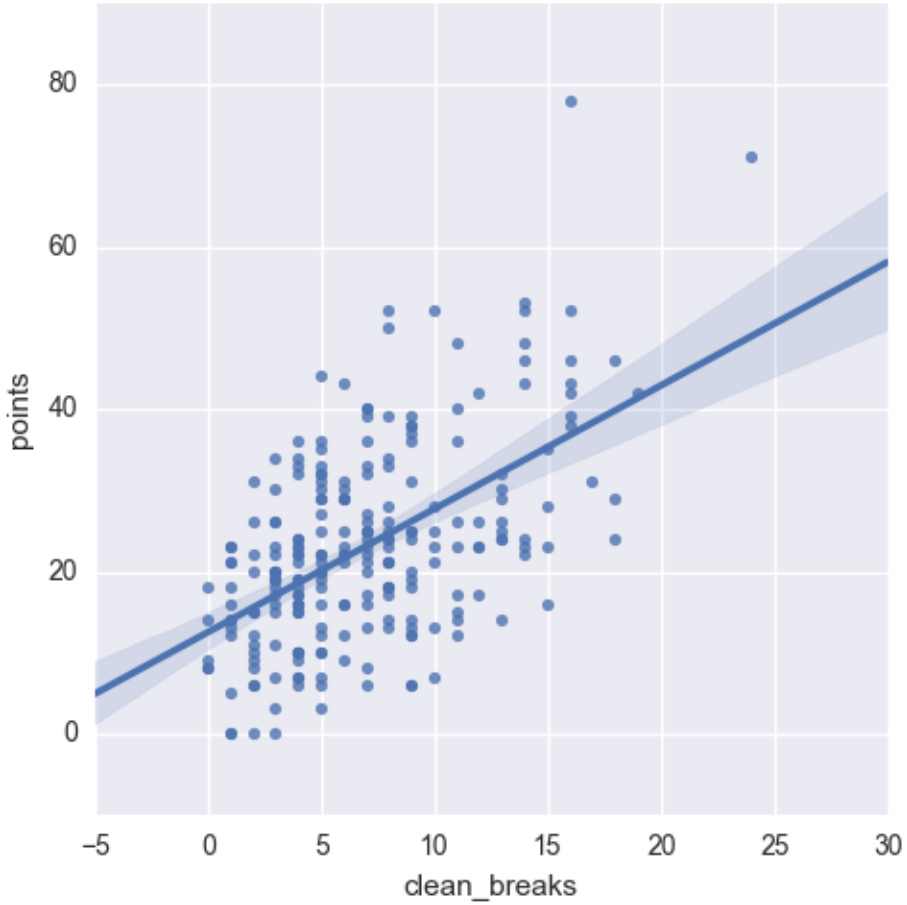
“Strength”

- This will simplify everything down to a single number, intended to predict the points scored by that team
- By comparing strengths, we can predict which of two teams will win


80
60
40
20
0
points
−10
0
10
20
30
40
50
defenders_beat


80
60
40
20
0
points
0
100
200
300
400
500
600
700
800
900
meters_run


80
60
40
20
0
points
1
2
3
4
5
6
7
8
meters_per_run


80
60
40
20
0
points
−5
0
5
10
15
20
25
30
clean_breaks

Where we're going

- "Strength" prediction model
  - Strength = predicted number of points scored in next game
- Use results of first 8 games to predict game 9
  - Use results of first 9 games to predict game 10, and so on
  - Do this for all 5 seasons
- Compare this against using a blanket "win-loss" model
  - Team with more wins predicted to win