データで見るプロ野球

２年　半澤映拓

目次

- １．野球をデータで見るとは？
- ２．今年の日本一は・・・
- ３．年俸と成績の関係（打者編）
- ４．年俸と成績の関係（投手編）
- ５．まとめ

１．野球をデータで見るとは？

昨今のプロ野球では選手のデータを細かく解析し、戦術や選手の評価に活かしている

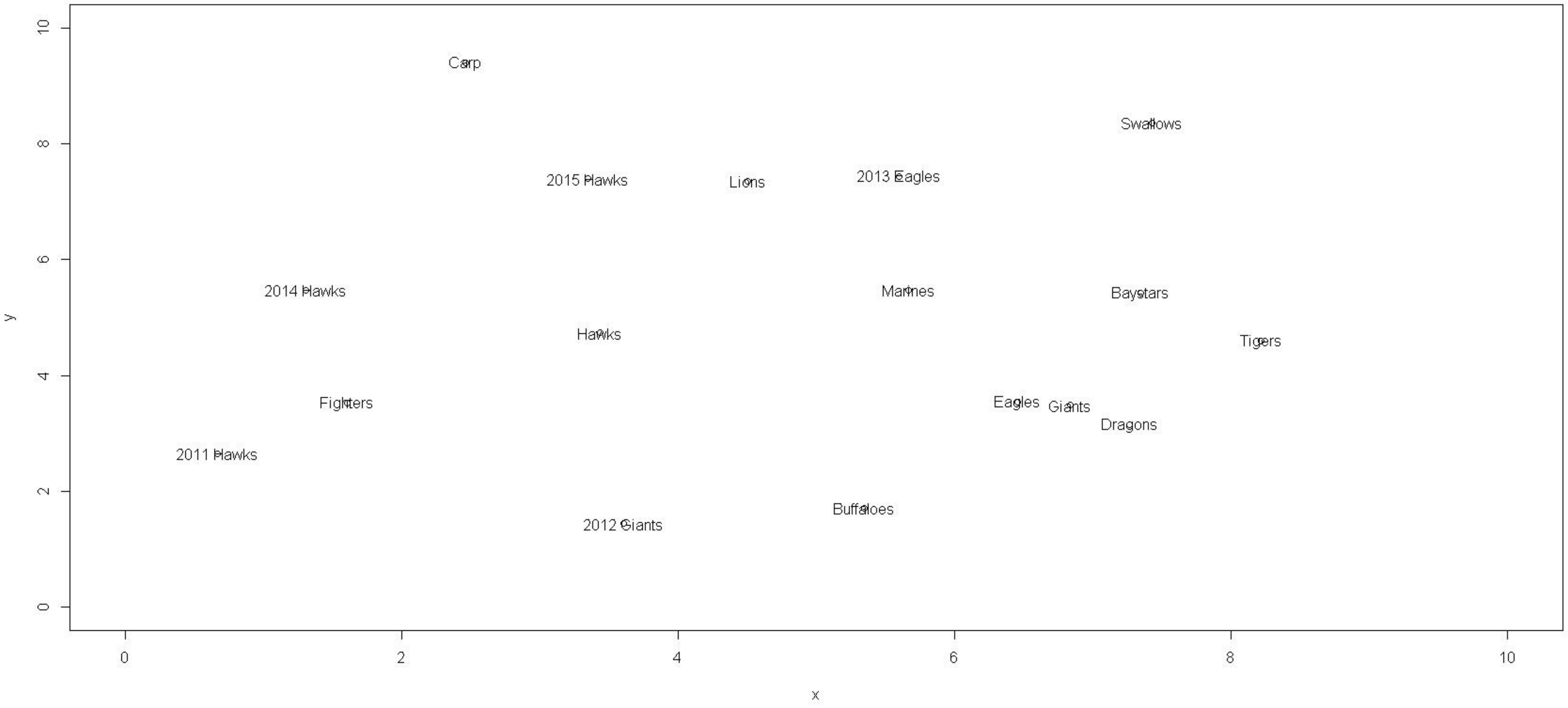
（統計学的見地から野球のデータを分析する手法をセイバーメトリクスと呼ぶ。その中にはビッグデータ分析も）

プロ野球のデータをマイニングすると見えてくることもあるのでは？

２．今年の日本一は･･･

	Team batting average	Team ERA	Average score	Average base steal	Average sacrifice	Team on-base
Carp	0.273	3.38	4.85	0.847	0.695	0.344
Giants	0.25	3.38	3.65	0.409	0.809	0.309
Baystars	0.247	3.77	3.79	0.508	0.576	0.306
Tigers	0.245	3.53	3.67	0.445	0.546	0.316
Swallows	0.258	4.9	4.29	0.555	0.605	0.334
Dragons	0.245	3.69	3.55	0.429	0.773	0.356
Fighters	0.265	2.88	4.35	0.983	1.25	0.34
Hawks	0.255	2.99	4.19	0.737	0.965	0.335
Marines	0.255	3.53	4.17	0.595	0.819	0.328
Eagles	0.253	3.94	3.64	0.396	0.937	0.32
Lions	0.266	3.81	4.27	0.681	0.664	0.339
Buffaloes	0.249	4.03	3.25	0.768	1.09	0.315
2015 Hawks	0.267	3.16	4.55	0.657	0.762	0.34
2014 Hawks	0.28	3.25	4.22	0.861	1.01	0.344
2013 Eagles	0.267	3.51	4.36	0.43	0.764	0.338
2012 Giants	0.256	2.16	3.71	0.708	0.931	0.326
2011 Hawks	0.267	2.32	3.82	1.25	1.02	0.323

チーム打率、チーム防御率、1試合平均得点、1試合平均盗塁数、1試合平均犠打数、チーム出塁率から
今年の１２球団と2０1１～２０1５の日本一のチームの自己組織化マップを作成

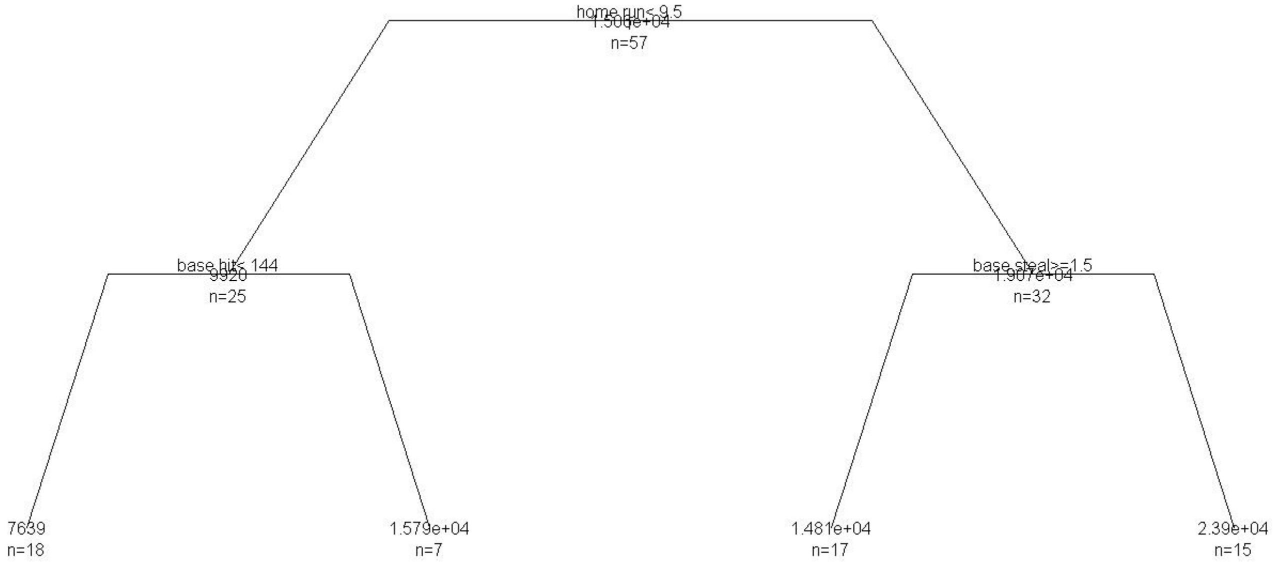


左下は防御率の低さが特徴のチーム、上には得点力の高いチームが集まっている
２０１４、２０１１ホークスと似ているファイターズが順当か。もしかしたらライオンズにも可能性が？

３．年俸と成績の関係（打者編）

annual salary	position	base hit	home run	run batted in	base steal	batting average	on-base percentage
41000	infielder	145	37	82	1	0.278	0.367
40000	infielder	155	6	42	9	0.281	0.38
35000	infielder	100	10	46	1	0.24	0.342
33000	outfielder	150	11	82	1	0.284	0.34
32600	infielder	83	15	47	0	0.242	0.37
30000	infielder	78	12	39	1	0.236	0.31
28000	outfielder	127	17	68	11	0.262	0.366
27000	outfielder	182	34	99	32	0.363	0.469
25000	infielder	129	12	68	10	0.269	0.353
25000	outfielder	91	18	62	0	0.258	0.352
25000	infielder	111	27	89	0	0.235	0.307
24500	infielder	142	30	102	1	0.263	0.339
24400	infielder	128	16	73	0	0.255	0.289
22000	infielder	153	35	94	8	0.287	0.357
22000	infielder	183	38	100	34	0.329	0.416
20000	outfielder	139	20	76	1	0.281	0.361
20000	infielder	141	17	72	0	0.271	0.364
20000	outfielder	143	10	42	3	0.268	0.358
17500	outfielder	109	15	52	3	0.251	0.31
16000	outfielder	114	10	48	14	0.256	0.324
16000	infielder	151	4	66	9	0.284	0.354
15000	infielder	151	4	66	9	0.284	0.354

今年の年俸を基準変数としてポジション、安打数、ホームラン数、打点、盗塁数、打率、出塁率
との関係を回帰木で分析

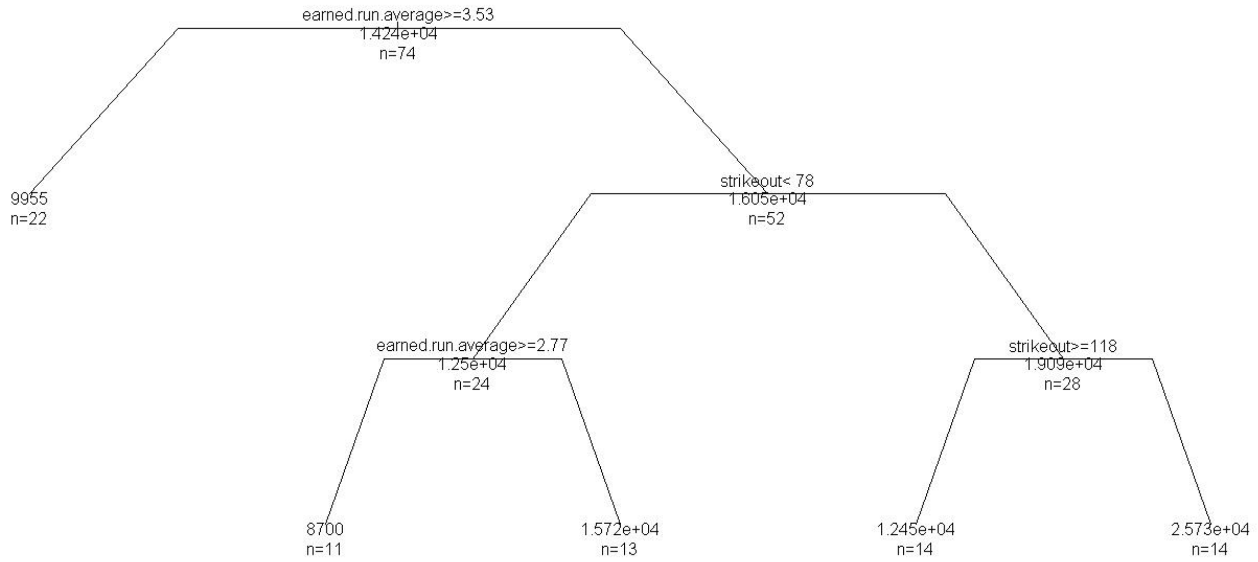


主にホームラン数、安打数、盗塁数が評価の対象？

４．年俸と成績の関係（投手編）

annual salary	pitting appearance	win	lose	save	complete games	shutout	strikeout	earned run average
60000	26	11	8	0	1	0	106	2.55
50000	16	7	6	0	0	0	79	3.19
50000	65	5	1	41	0	0	102	1.11
40000	20	10	7	0	2	0	92	3.22
35000	54	3	1	2	0	0	59	1.38
32000	60	4	5	2	0	0	31	2.73
30000	29	9	12	0	0	0	194	2.97
30000	33	0	3	12	0	0	39	4.06
24000	21	13	3	0	4	2	107	1.92
22500	16	5	6	0	5	0	91	3.02
22000	56	0	1	39	0	0	71	1.5
22000	28	15	9	0	1	0	117	3.39
22000	72	5	2	0	0	0	62	2.05
20000	25	13	9	0	1	1	90	3.31
20000	26	10	8	0	1	0	85	3.51
19000	50	3	3	0	0	0	30	2.7
18500	63	3	8	2	0	0	55	2.62
18000	20	9	9	0	1	0	111	2.48
17000	28	14	7	0	1	1	221	2.4
15600	24	8	8	0	0	0	101	2.94
15000	28	10	11	0	3	1	215	2.91
15000	15	9	0	0	0	0	120	2.52

今年の基準変数として登板数、勝数、負数、セーブ数、完投試合数、完封試合数、奪三振数、
防御率との関係を回帰木で分析



防御率と奪三振数が主な評価基準？

５．まとめ

- 複数年契約などの複雑な事例を取り除く、解析対象を細分化し条件を揃えるなどの必要がある
- 野球にデータマイニングを取り入れるのは有用そうだ
- セイバーメトリクスの有用性を示すのにも使えそうであると感じた