

マラソン選手に関するデータ分析

概要

今回私はマラソン選手のデータを用いたデータ可視化を行なった。日本の代表選手や海外の選手、日本の歴代記録を残している選手のデータを用いて、マラソン選手に関する分析を行う。

データについて

今回、日本陸上競技連盟公式サイトของทีม JAPAN のページに名前の載っている男子 22 人、女子 20 人の選手。マラソン世界歴代 50 傑に載っている男子 20 人、女子 17 人の選手。マラソン日本歴代 50 傑に載っている男子 11 人、女子 10 人の選手。合計 100 人を本課題の対象選手とした。

以下表 1 に本課題で使用した、12 個のパラメーターを示す。

表 1

名称	内容
性別	男：1 女：2
生年月日	生まれた年
所属	海外の選手：0 日本の選手 右の表 2 の番号を使用
出身地	海外の選手 ケニア：50 エチオピア：51 バーレン王国：52 イングランド：53 ドイツ：54 アメリカ：55 日本の選手

表 2

Nike	1
西鉄	2
旭化成	3
安川電機	4
ホンダ	5
九電工	6
三菱重工	7
AD損保	8
中電工	9
GMOインター	10
富士通	11
トヨタ	12
JR	14
JP日本郵政	15
鹿児島銀行	16
日本郵政グル	17
大塚製薬	18
ワコール	19
第一生命グル	20
天満屋	21
ダイハツ	22
しまむら	23
スズキ	24
岩谷産業	25
積水化学	26
資生堂	27
三井住友	28
カネボウ	29
中国電力	30
エスビー	31
日清食品	32
ノーリツ	33

	都道府県番号を使用
マラソン*	フルマラソンの自己ベスト
ハーフ*	ハーフの自己ベスト
10000*	10000m の自己ベスト
5000*	5000m の自己ベスト
PB 出した年	フルマラソンの自己ベストを出した年
PB 出した年齢	[PB 出した年]-[生年月日] 自己ベストを出した月や生まれた月を考慮していないため、最大1のズレがある。
身長	身長[cm]
体重	体重[kg]

*のついたパラメータについて

マラソンや陸上競技のトラック種目の記録は時間で表されているため、今回のデータとしてはそのまま使用することができない。そのため、エクセルで最も簡単にできる時間のデータを1=1日=24時間という変換の仕方 で数値化した。エクセルに時間表記で入力すると、番号書式が勝手に時刻になるが、そこを標準に直すことで数値化される。

結果・考察

結果は以下の図4の通りである。

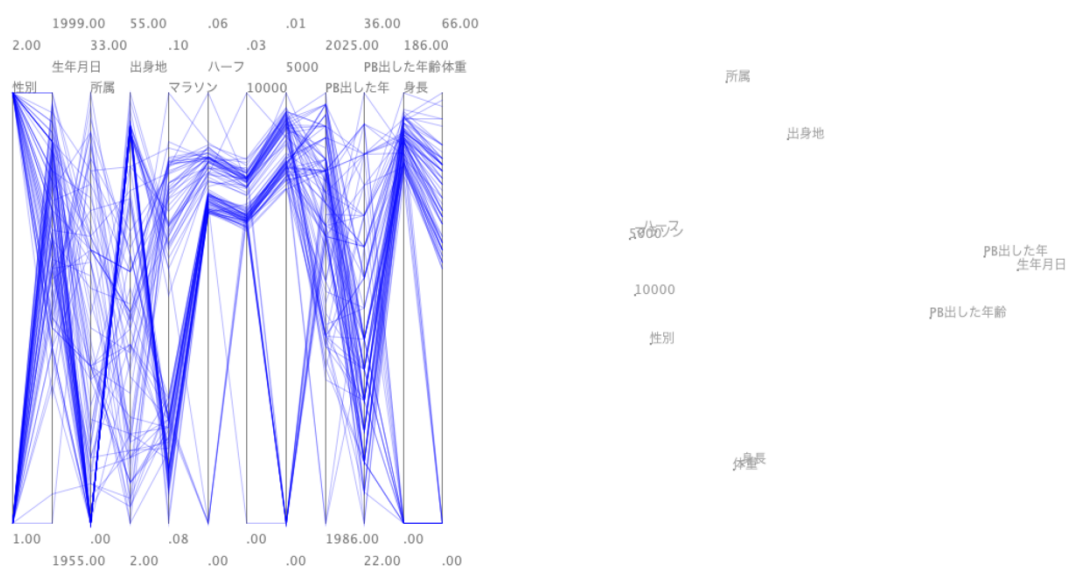


図4

散布図に着目すると、関連性の高いパラメーターが近くに位置しているのがよくわかる。一点気になることとしては、フルマラソン、ハーフマラソン、5000 のパラメーターがほとんど同じところに位置しているのに対して、10000 だけ離れている。次に右上のスライダーを散布図や PCP に変化があるまで動かしたところ、図 5 のようになった。

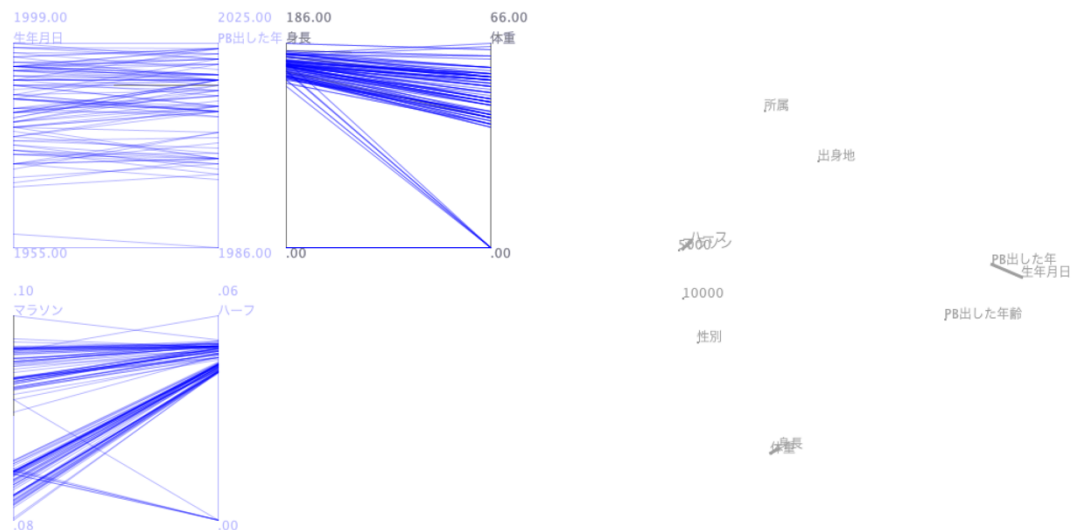


図 5

散布図を見てみると、PB を出した年と生年月日（赤丸）、身長体重（青丸）、マラソンとハーフ（緑丸）が線で繋がれた。PB を出す年齢はおよそ体の動ける 20 代から 40 代であり、PB を出した年は、生まれた年に 20~40 をたした値となるため、相関が強いのは容易に想像できる。残りの身長と体重、マラソンとハーフの記録も相関が強いのは容易に想像できる。そこで別のパラメーター同士の相関を見たいため、近いものは削除したい。左の Culling のスライダーを動かし、散布図中で近い位置にある点を削除した。図 6 のような結果になった。

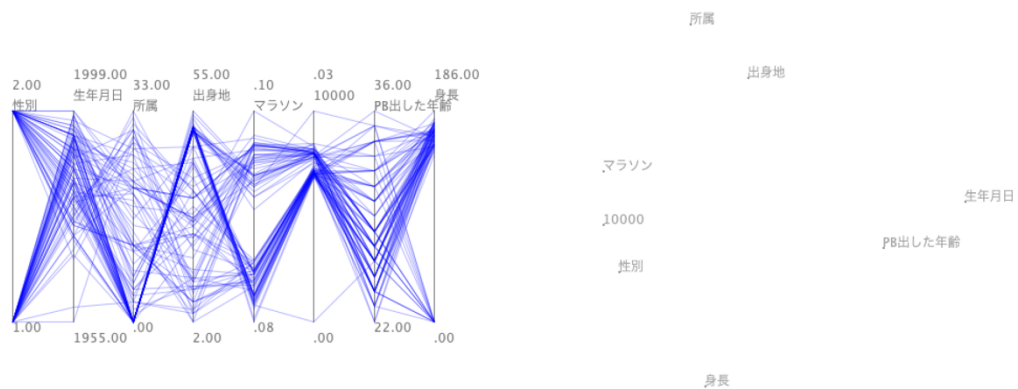


図 6

削除された点は以下の 5 点である。

- ・ ハーフと 5000：マラソンと近い
- ・ 体重：身長と近い
- ・ PB 出した年：PB 出した年と近い

削除後のパラメータは 8 つ。この 8 つのパラメータで右のスライダーを変化があるまで動かしてみる。結果は図 7 のようになった。

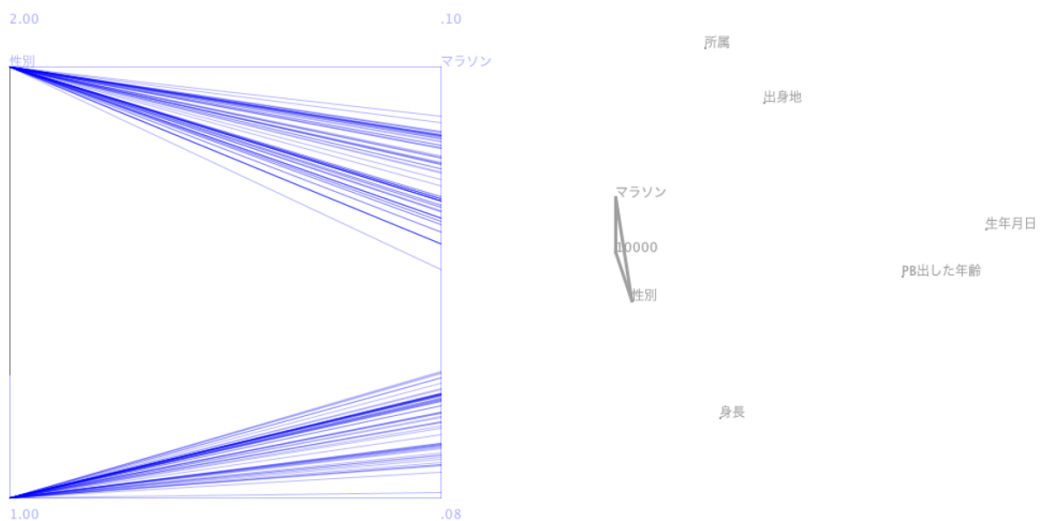


図 7

10000 とマラソンと性別が繋がれた。男女でフルマラソンと 10000m の記録に差があり、マラソンの自己ベストと 10000 の自己ベストが強い相関を持つのは容易に想像できる。そこでさらに変化があるまで、右のパラメーターを動かした。結果は図 8 のようになった。

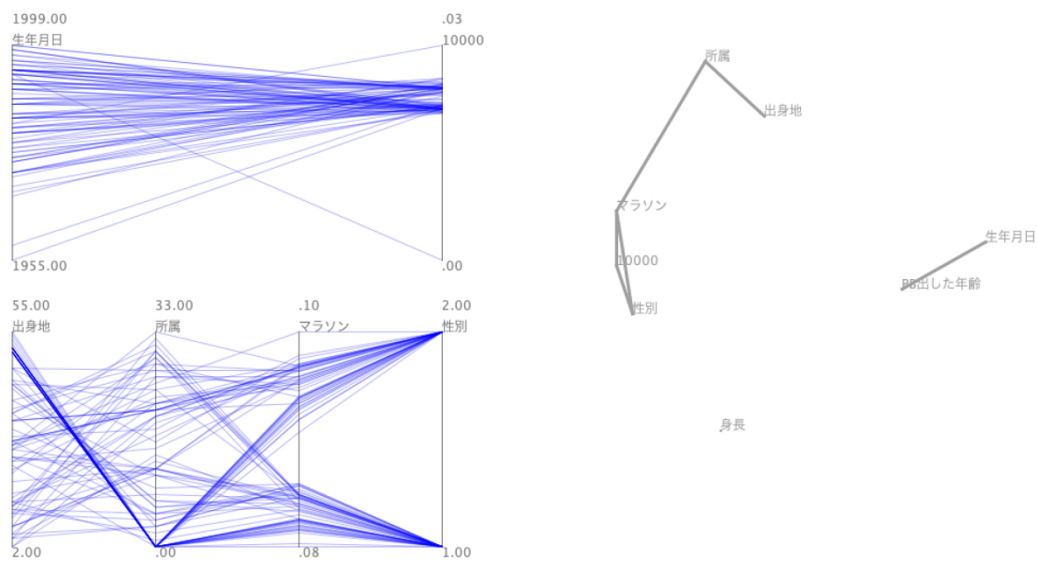


図 8

PB 出した年齢と生年月日が繋がれた。また、所属が出身地とマラソンに繋がれた。PB 出した年齢と生年月日が結ばれるのは、PB を出した年齢を生年月日と PB を出した年から計算しているからであると考えられる。所属が出身地とマラソンに繋がれるのは、データ収集の対象に偏りがあること、外国人選手の番号の付け方に関係していると考えられる。今回は所属の番号の付け方を外国人は全員 0 とした。今回外国人選手の対象は世界歴代 50 傑に載っている選手のみなため、出身国に偏りが出てしまった。女子はまだいろんな国の選手が 50 傑に載っていたが、男子はケニアとエチオピアの選手がほとんどである。そのため、所属と出身地の相関が強くなったと考えられる。また、海外の選手を世界歴代 50 傑に名前の載っている選手を対象としたため、記録も日本の選手と差がある。日本歴代 50 傑に名前の載っている選手とも差があるため、相関が強くなってしまったのだと思う。この偏りを解消するため、データは少なくなってしまうが、海外の選手のデータを一度削除して可視化してみようと思う。

外国人選手を削除したデータで可視化した結果が図 9。

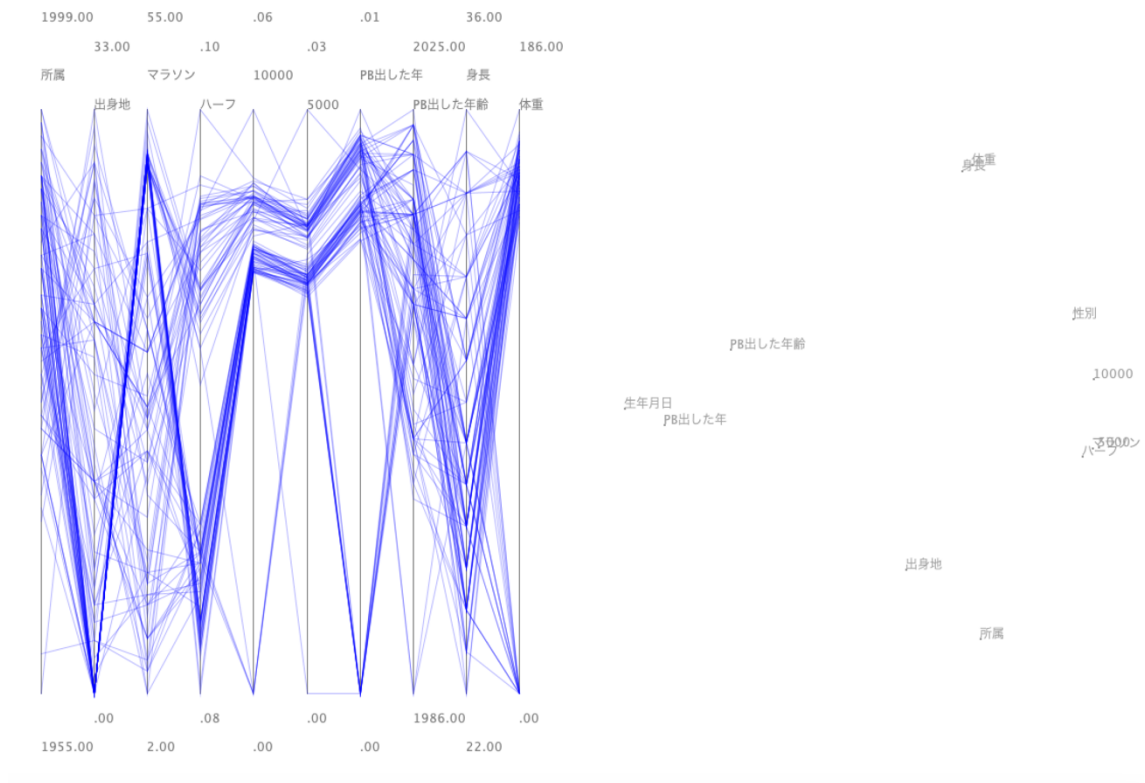


図 9

外国人選手を含んだ場合とそんなに大きくは変わらなそうである。近い点を先に削除してみる。結果は図 10 である。

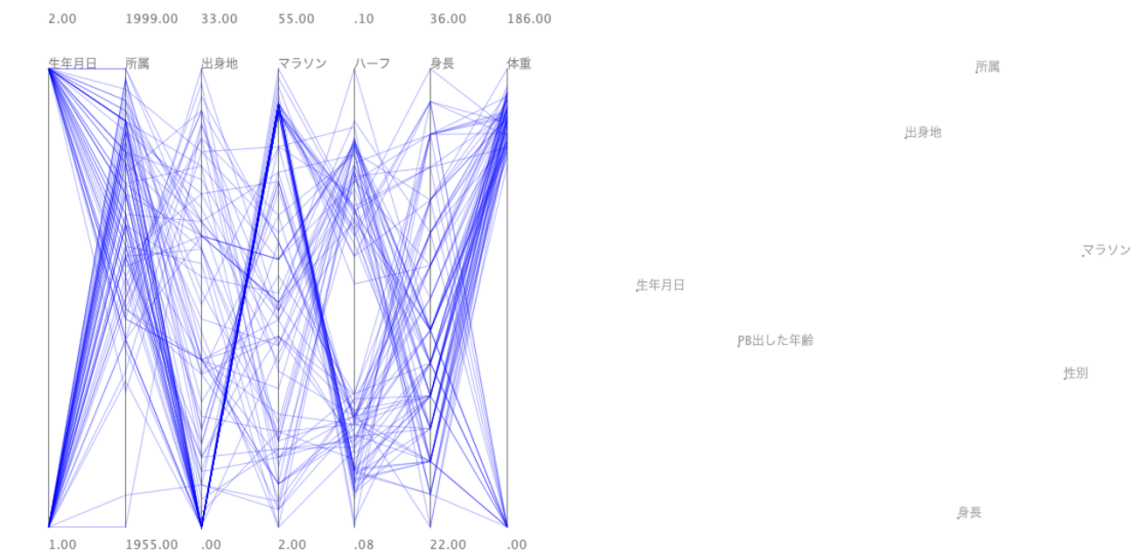


図 10

削除された点は、以下の点である。

- ・ PB 出した年：生年月日と近い

- ・ 体重：身長に近い
- ・ ハーフ：マラソンに近い
- ・ 10000：マラソンに近い
- ・ 5000：マラソンに近い

この削除後の 7 点で可視化してみる。変化があるまで右のスライダーを動かす。結果は図 11 である。

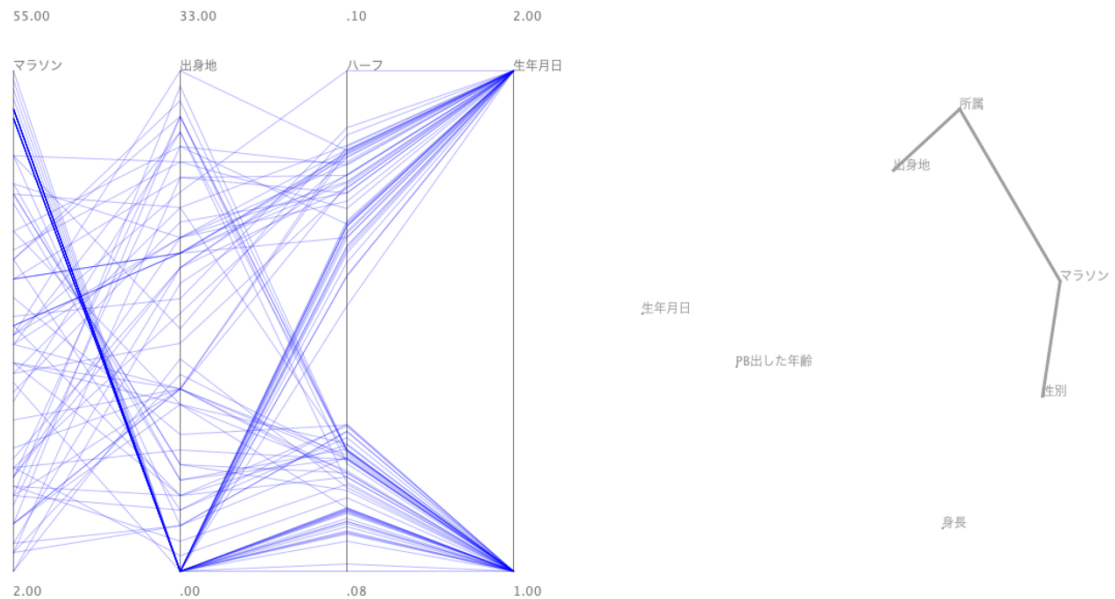


図 11

マラソンと性別、所属と出身地と、マラソンと所属が結ばれた。マラソンと性別、所属と出身地の相関が強いのは容易に想像できるが、マラソンと所属が結ばれたのは意外だった。また、外国人の選手を含めたデータで行なった時とは違う結果になった。強い選手が強いチームに参加することが多いのか、あるチームの監督が名監督で選手を強い選手に育てるのか、わからないが、ある程度チームによって強い選手が集まるのだということがわかった。もう一点予想と違ったことは身長がマラソンとつながらないこと。競技の種類によっては、身長によってだいぶ結果に差が出てくる。ただこの可視化によって、マラソン競技はあまり身長は関係ないことがわかった。

次に二つのクラスタにクラスタリングしてみた。結果が図 12 の通りである。

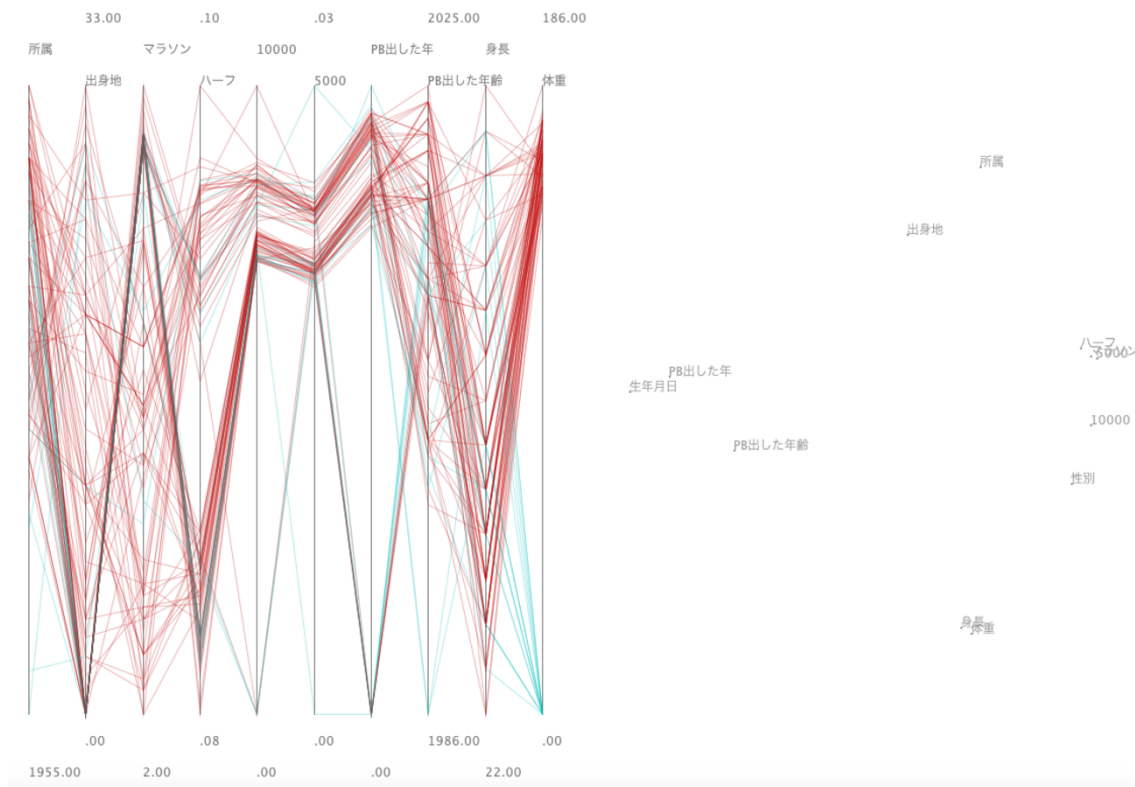


図 12

PCP をみると、体重のデータがあるかないかでクラスタリングされてしまっている。これでは意味のないため、体重のデータがないデータに関しては削除する。体重のデータがない選手を削除したデータを可視化した結果が図 13 である。

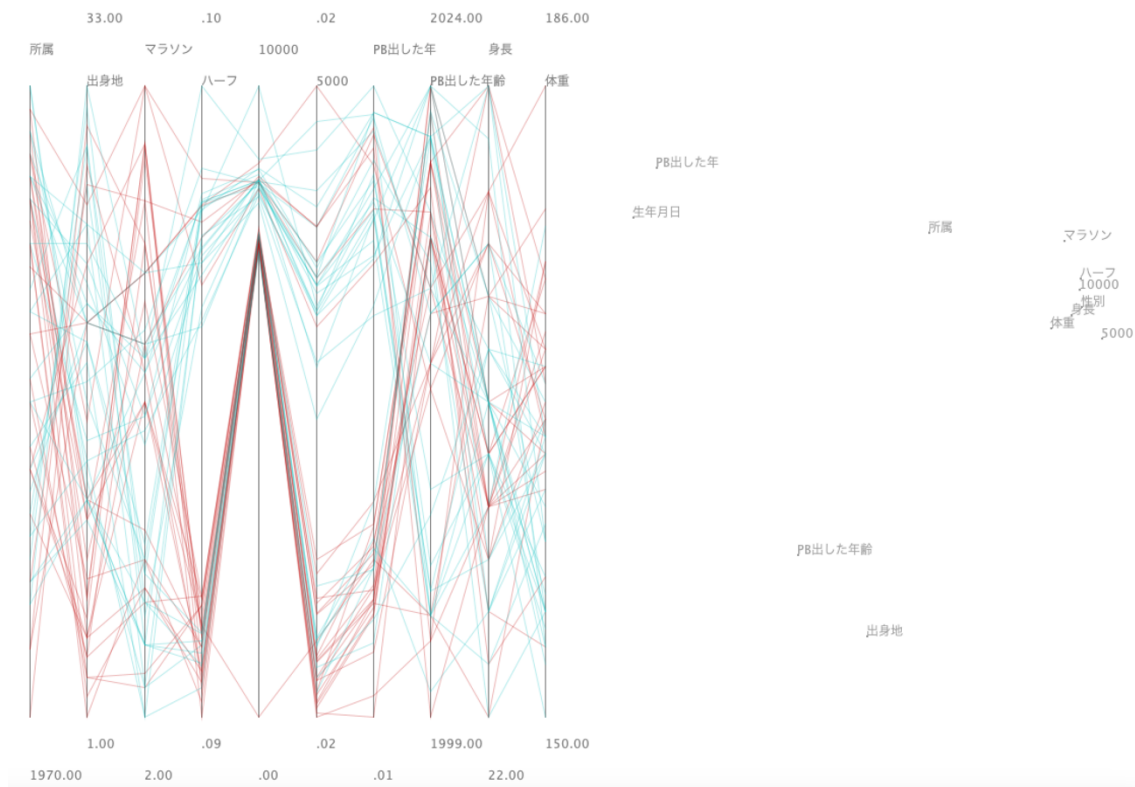


図 13

データを絞ったところ、散布図の方もかなり変化した。ただ、どうクラスタリングされているのかわからなかったため、左の Culling のスライダーを動かしてみた。結果は図 14 である。

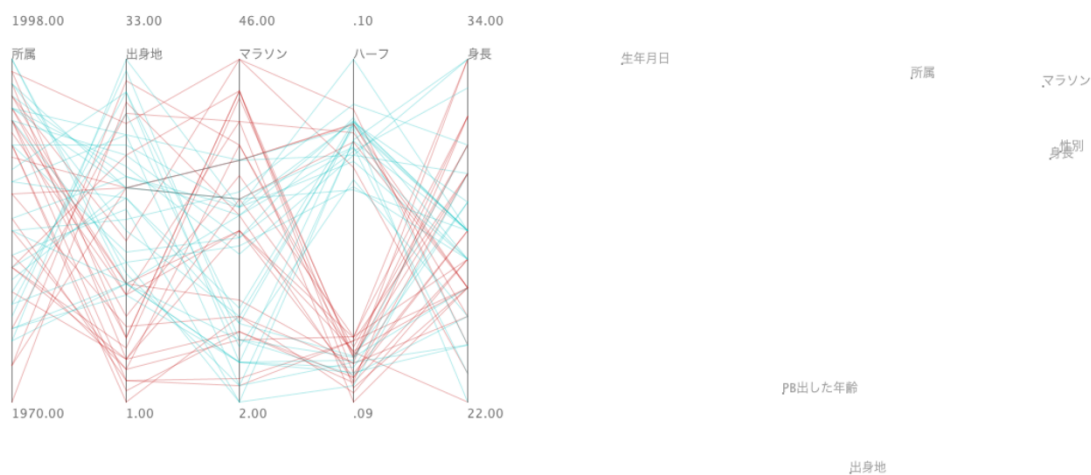


図 14

図 13 からの PCP の結果に着目して、マラソンとハーフの傾きをみると、右肩下がりになっている線は赤色になり、右肩上がりになっている線は緑色になっていることがわかる。軸

の上が値が大きいいためタイムが遅いことに注意して考えると、右肩上がりになっている選手は、フルマラソンがハーフマラソンより得意で、右肩下がりになっている選手はハーフマラソンがフルマラソンよりも得意な選手であることがわかる。また赤い選手は出身地の都道府県番号が小さい、つまり東の方出身の選手が多いことがわかり、緑色の選手は出身地の都道府県番号が大きい、つまり西の方の出身の選手が多いことがわかる。これは全く予想がつかなかった。

感想

様々な角度から可視化し分析してみて、可視化しないと気づかないようなマラソン競技の特性を知ることができて面白いと感じた。データの収取がかなり大変だったこともあり、幅広い選手のデータを収集することができず、データに偏りが出てしまったのはよくなかった。日本の選手だけに絞るか、各国の代表選手に絞ってデータを収取すればよかったと思った。身長がマラソン競技の結果に関係ないと知って、確かに箱根駅伝の選手などをみても、身長の高い選手もいれば、低めの選手もいるなと思いました。これは関係のない話ですが、私はランニングサークルに入っていて、フルマラソンの大会に出たりしていて、最近タイムも目指し始めたのですが、マラソン選手の体重を見て、やっぱりこのくらい軽くないと速く走れないよと、少しダイエットを頑張ってみようと刺激を受けました。

参考文献

- ・ JAFF 日本陸上競技連盟公式サイト
<https://www.jaaf.or.jp/athletes/?event=8>
- ・ ベースボールマガジン社「マラソン世界歴代50傑」
https://www.sportsclick.jp/user_data/event/rikumaga_teisei201907.pdf
- ・ 厚生労働省「都道府県番号」
<https://www.mhlw.go.jp/topics/2007/07/dl/tp0727-1d.pdf>
- ・ ウィキペディア
<https://ja.wikipedia.org/wiki/>