

情報可視化ソフトウェア Hidden によるデータ分析課題

1955 年から現在までの映画に関連するデータ比較分析

映画が好きなため、映画に関するデータとその背景の分析を試みた。

1. 設定した軸

1955 年から 2020 年まで各年、以下のデータを集めた

- ・映画館数
 - ・邦画公開本数（本）
 - ・洋画公開本数（本）
 - ・映画館入場者数（千人）
 - ・総興行収入（百万円）
 - ・映画料金（円）
 - ・邦画配給収入（百万円）
 - ・洋画配給収入（百万円）
 - ・洋画のシェア（%）
 - ・最高配給作品額（億円）
 - ・サラリーマン平均年収（千円）
 - ・エンゲル指数（%）
 - ・2 人以上の世帯の 1 ヶ月間の消費支出（円）
 - ・紅白歌合戦平均視聴率（%）
- （ただし、1995-1961 年までのデータはなかったため、全て 1962 年と同じにしている）
- ・外国人入国者（人）
 - ・自殺者数（人）

計 16 項目 （括弧内は単位）

<データ引用元>

日本映画制作者連盟: <http://www.eiren.org/toukei/index.html>

興行収入ランキング: https://entamedata.web.fc2.com/movie/top_jmovie.html

会社員平均年収: <https://gentosha-go.com/articles/-/35815>

総務省統計局 家計調査: <https://www.stat.go.jp/data/kakei/rank/singleyear.html>

年次統計: <https://nenji-toukei.com/n/kiji/10081>

出入国在留管理庁: https://www.moj.go.jp/isa/policies/statistics/toukei_ichiran_nyukan.html

警察庁 自殺者数: <https://www.npa.go.jp/publications/statistics/safetylife/jisatsu.html>

ビデオリサーチ NHK 紅白歌合戦世帯視聴率:

https://www.videor.co.jp/tvrating/past_tvrating/music/02/nhk-2.html

2. 作成した CSV データ

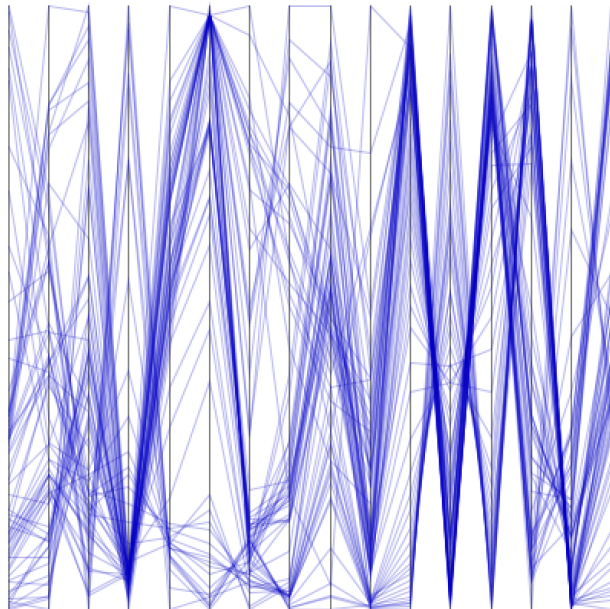
Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
5184	423	193	868912	54657	133	20993	10923	34.2	3.1	208	46.92	23211	80.4	55638	22477	
6123	514	177	993875	61899	143	23520	11317	32.5	4	231	45.02	23958	80.4	66356	22107	
6865	443	194	1098882	68153	100	25988	11600	30.9	5	249	44.39	25608	80.4	80361	22136	
7067	504	169	1127452	72346	109	29971	9435	23.9	4	251	43.79	27171	80.4	96514	23641	
7400	493	210	1088111	71141	108	30258	8865	22.7	4	274	42.42	28902	80.4	118020	21090	
7457	547	216	1014364	72798	119	31125	8606	21.7	6	300	41.57	31276	80.4	146881	20143	
7231	535	229	863430	73003	134	30020	8849	22.8	5	341	40.32	34329	80.4	180048	18446	
6742	375	228	662279	75983	160	27834	10262	26.9	7	381	39.04	38587	80.4	202181	16724	
5696	357	267	511121	77734	199	25259	11456	31.2	9	421	38.69	40246	81.4	227289	15490	
4927	344	259	431454	76937	221	22771	11553	33.7	12	467	38.07	44481	72	273551	14707	
4649	487	264	372676	75506	236	22528	11230	33.3	7	507	38.13	48396	78.1	291309	14444	
4296	442	250	345811	75750	247	20221	11771	36.8	10	549	37.31	52516	74	338584	15050	
4110	410	220	235067	78042	250	19812	12500	32.7	8	620	36.8	57071	76.7	276020	14121	

1. で上げた軸の名前を上から A、B、C...P とし、

I と J は散布図上で見ると似ていたため、I は II としている

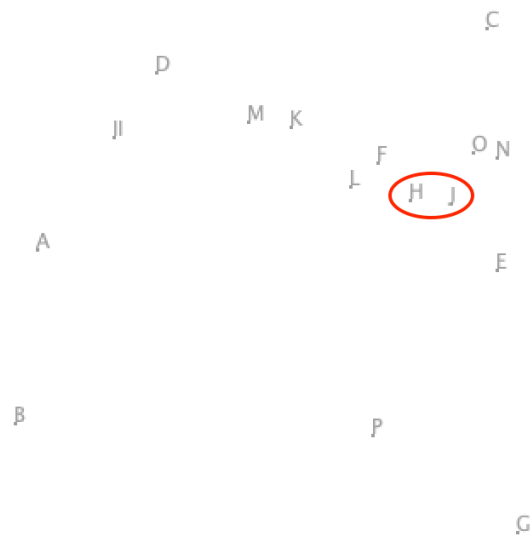
3. 実行結果とその考察

B D F H J L N P
A C E G II K M O
745 5.89 593. DD2 2652180 3.006638. 402690 1.40673.60 335 246400 187429.00



173 2.30 069. D06 5345 80.0046 85.00 073. 10208. 22.9032 37.6056 38. D01.00

散布図



散布図を見ていて気になったのが

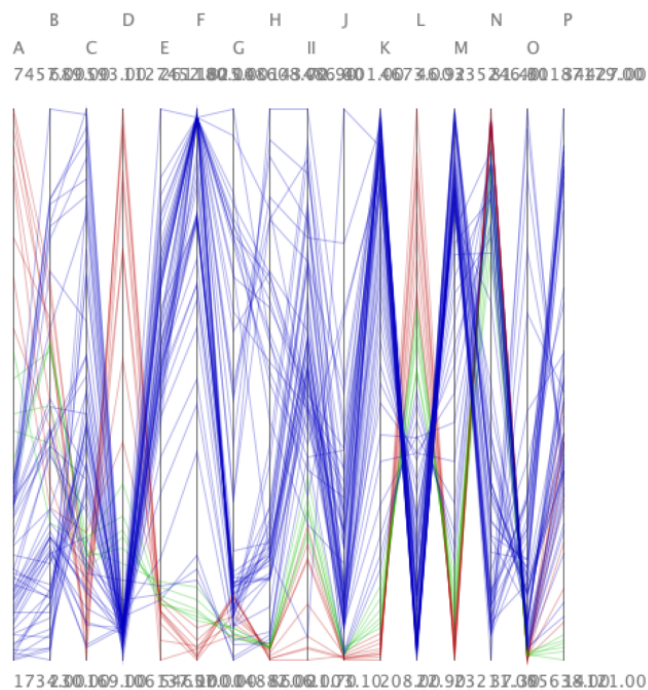
赤丸をした H「洋画配給収入」と J「最高配給作品額」の相関が強いことである。

もしこれが正の相関なのであれば、「最高配給作品額が高ければ、洋画配給収入も高くなる」

つまり、「最高配給作品は洋画が多い」という可能性も考えられる。

この後の PCP で詳細に観察しようと思う。

クラスタリング



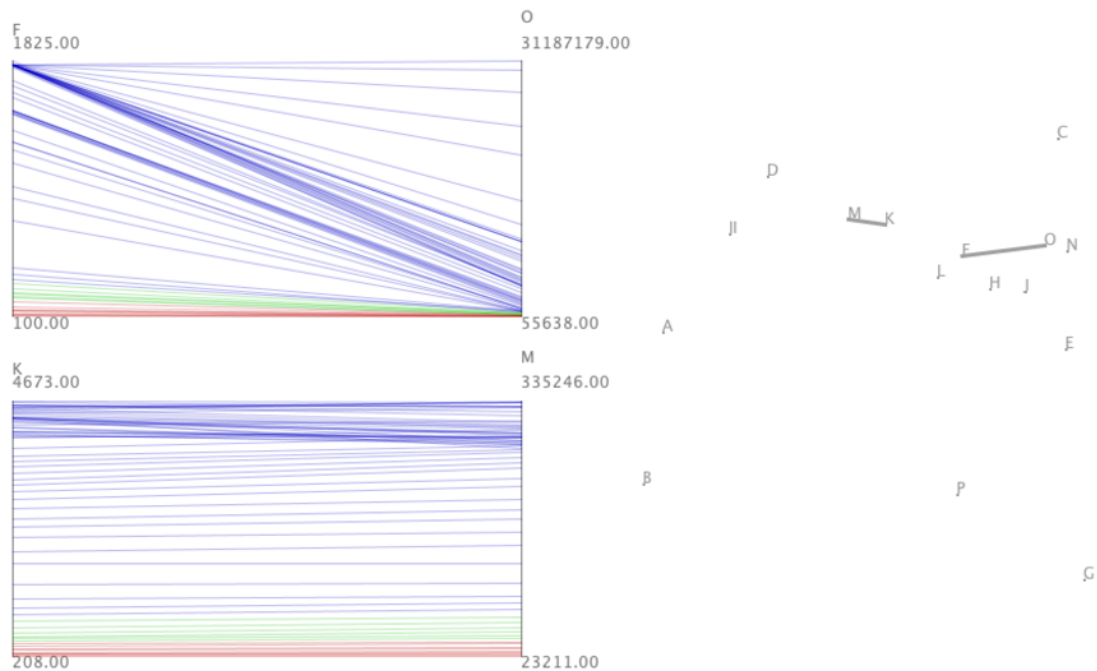
Clustering = 3 にすると、A「映画館数」軸でちょうど「上・真ん中・下」に分かれている。

このことから、青は映画館が少ない時期のデータ、緑は中くらいのデータ、赤は映画館が多い時期のデータだと分かる。

ちなみに、映画館の多さは時代と共に「多い→少ない」になっているため、青は最近のデータ、緑は中期のデータ、赤は昔のデータとして考えることもできる。

わかりやすいので、この3色で見ていこうと思う。

「映画料金」と「外国人入国者」「サラリーマン平均年収」と「消費支出」



(F「映画料金」 O「外国人入国者」 K「平均年収」 M「消費支出」)

右上のスライダーを操作し、まず出てきたのは

F「映画料金」とO「外国人入国者」、K「サラリーマン平均年収」とM「消費支出」の相関だった。

どちらも強く「正の相関」が見られる。

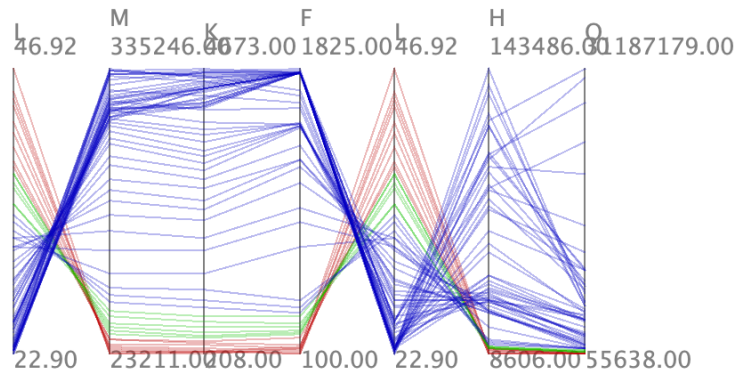
「サラリーマン平均年収」と「消費支出」

→年収が多くなるほど消費支出が多くなるということは一般的に言われているが、可視化することで再確認できた。

「映画料金」と「外国人入国者」

→映画料金は年々高くなり、外国人入国者はグローバル化により年々増えている、というだけであつた。特に面白い結果は見られなかった。

「洋画配給収入」と「外国人入国者」



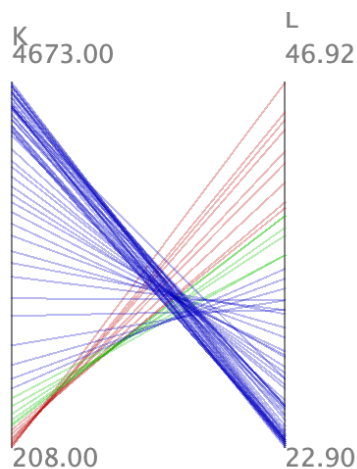
(L「エンゲル指数」 M「消費支出」 K「平均年収」 F「映画料金」
H「洋画配給収入」 O「外国人入国者」)

上のPCPでまず注目したのは、H「洋画配給収入」とO「外国人入国者」の相関である。外国人入国者が多いと日本人の海外への関心が高まり、洋画配給収入も高まるのではないかという予想から入れたデータであったが、今回の可視化から、特にそういったことは見られないことがわかった。

「映画料金」と「エンゲル指数」

上の図の中のF「映画料金」とL「エンゲル指数」に注目すると、強い「負の相関」が見られる。このことから、「エンゲル指数が高い時は映画料金が安い」ということが分かる。そのため、今後エンゲル指数に注目していれば、映画料金がなんとなく推測できるようになるかもしれない。

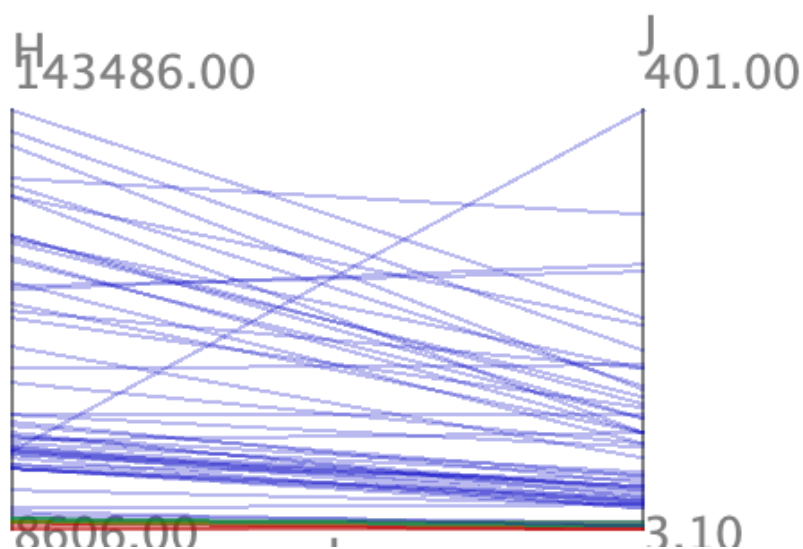
「サラリーマン平均年収」と「エンゲル指数」



(K「平均年収」 L「エンゲル指数」)

K「サラリーマン平均年収」と L「エンゲル指数」では、強い「負の相関」が見られた。
エンゲル指数は、消費にしめる飲食費の比率を指すもので、一般的には年収が上がるほど低くなると言われている。言われている通り綺麗な負の相関が確認でき、そのことを再認識した。

「洋画配給収入」と「最高配給作品額」

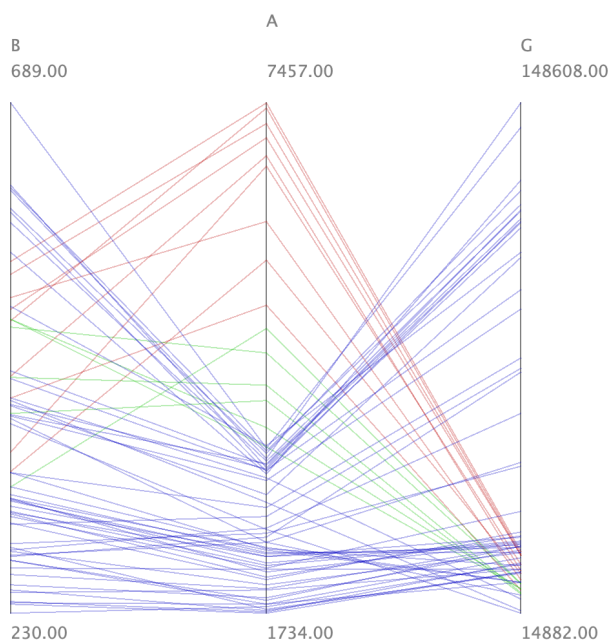


(H「洋画配給収入」 J「最高配給作品額」)

初めの散布図で気になっていた H「洋画配給収入」と J「最高配給作品額」の PCP が出てきた。
見ると、全体的には正の相関があることが分かる。
そこで、先述した「最高配給作品は洋画が多い」という可能性を調べるため「歴代興行収入ベスト100」を見たところ、なんと **64/100 作品** が洋画であることがわかった。
これは少し悲しい結果であるが、日本で公開される洋画はすでに世界で話題になっているものがほとんどなため、仕方がないこととも思える。

(参考文献: <http://www.kogyotsushin.com/archives/alltime/>)

「映画館数」と「邦画公開本数」と「邦画配給収入」



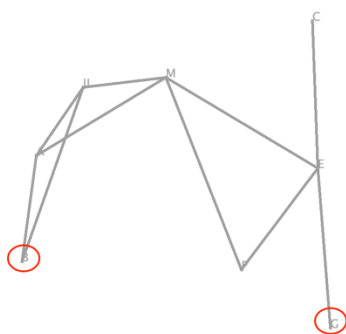
(A「映画館数」 B「邦画公開本数」 G「邦画配給収入」)

次に、Culling スライダーを右に動かし、PCP の軸の数を減らした状態で右のスライダーを操作してみた。すると、上の図のような結果が得られた。

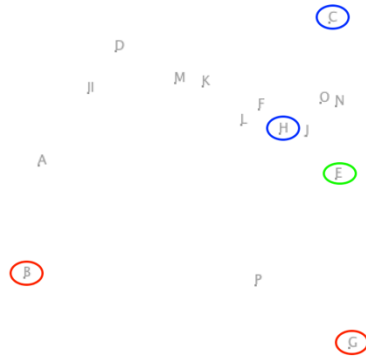
赤を一つの塊、青を一つの塊、緑を一つの塊として見ると、

A「映画館数」とB「邦画公開本数」間の上の方に負の相関がみられる。

このことから、昔は映画館数のわりに上映されている本数は少なかったということが分かる。



またこの時の散布図を見ると、赤丸をつけたB「邦画公開本数」とG「邦画配給収入」の2点は意外にも離れている。



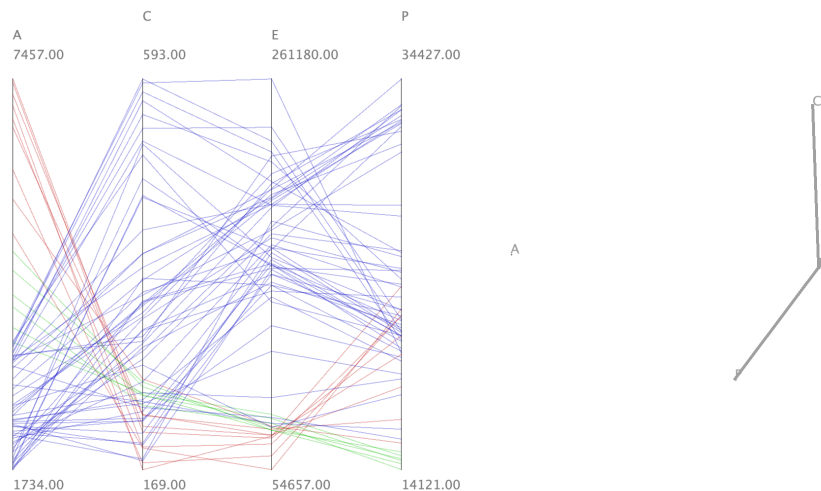
全体の散布図を見ても、

- ・B「邦画公開本数」とG「邦画配給収入」
- ・C「洋画公開本数」とH「洋画配給本数」
- ・B「邦画公開本数」とC「洋画公開本数」とE「総興行収入」

は全て離れている。

公開された本数が多いほど配給収入も多くなり正の相関が出そうだが、映画は「数打てば当たる」ということはないのだと分かる。

「総興行収入」と「自殺者数」



(A「映画館数」 C「洋画公開本数」 E「総興業収入」 P「自殺者数」)

次も同じく Culling スライダーを右に動かし、PCP の軸の数を減らした状態で右のスライダーを操作してみた。

今回軸の中に P「自殺者数」を入れたのは、「映画が人気な年は自殺者が減る」とよく行く映画館の人が言っていたからである。

しかし上の PCP を見ると、残念ながら「総興行収入」と「自殺者数」に相関は見られなかった。

ただ、E「総興業収入」の上 5 つくらいに注目すると、確かにこの年は自殺者が少ないため、全く関係がないとも言い切れないと思った。

(ちなみに E で最も興行収入が高くなっているのは、2019 年「天気の子」が大ヒットした年である)

4.感想

本来はもっと多くの年のデータを集めたかったが、1955 年からしか残っておらず 65 個分しか集まらなかったのが、残念であった。

データを可視化してみて、予想通りの結果から少し意外なものまで見ることができ、面白かった。特に、年収とエンゲル指数の関係や、年収と消費支出の関係など、社会科の教科書で見ているものが実際に折れ線グラフや散布図として見ることができてよかった。

個人的には、映画の公開本数と興行収入/配給収入に相関がなかったことが面白く、映画は数より質なのだと思う。

また、現在と 1 円の価値が変わっていたり外れ値があったりで、確認したかった情報が得られないことがあったため、情報を選ぶ難しさを実感した。