情報可視化ソフトウェア Hidden によるデータ分析課題
1 都 3 県の偏差値 45 以上の大学比較

1. はじめに

アルバイトをしている塾の生徒さんから、国立大学の良さを聞かれた際に、「学生数に対して教員数が多く、面倒見が良い」ことなどを思いついたが、あくまで自分の感覚に過ぎなかったため、実際にデータとして分析した上で説明できるようになりたいと思ったこと、自分が受験する大学を選ぶ際に考慮した情報にどのような相関関係があるのか気になったことから、1 都 3 県（「神奈川」「埼玉」「千葉」「東京」）の大学についてさまざまな変数で比較してみることにした。

2. データの出典・背景

2-1. 対象大学の選出方法について

河合塾が提供する大学検索システムKei-Net(https://search.keinet.ne.jp/search/option/)において、

1.「各種条件から探す」

2.「大学情報」の「設置区分」において「女子のみの大学（学部・学科）」→「含む」、「二部・夜間主」→「含まない」（二項目ともデフォルトのままにした）にする

3.「所在地」の都道府県選択において「神奈川」「埼玉」「千葉」「東京」の1都3県を選択

4.「入試情報（一般入試）」の「入試難易度」において「2次試験・個別学力検査ボーダー偏差値帯」を「45.0以上74.9以下」に設定（上限値はデフォルト）、「BF（ボーダー・フリー）の大学を」→「含まない」とする

1.〜4.の操作を行い、文部科学省管轄外の大学校等は除いた上で、127校を対象に選んだ。（なお、検索システムの機能上、一度に100大学までしか選択できなかったため、「所在地」の都道府県選択において一度「東京」、「神奈川」「埼玉」「千葉」の順に県別に情報を取得する形をとった。）

2-2. 使用した変数について

変数名	内容
A	学部数
B	学科数
C	1学部あたり学科数（平均）
D	学生総数（人）
E	学生女子率（%）
F	偏差値
G	1学科あたり入試日程数（平均）
H	学費（円）
I	教授数（人）
J	准教授数（人）
K	教員1人あたり学生数（人）
L	キャンパス数
M	最寄駅からのアクセス（分）
N	歴史（年）
O	国公立か私立か（0か1）

以下、各変数データの取得方法について記述する。

＊なお、以下のデータ取得の中で、夜間学部の数値も別に記載されている場合は、それを除いて集計した。

＜A（学部数）・B（学科数）＞

上述した大学検索システム Kei-Net での検索結果のそれぞれの大学ページにおいて、「大学情報メニュー」＞「大学概要」＞「学部・学科」を選択して開かれるページに載っている「設置学部・学科」の表においてそれぞれの行数をカウントした。

＜C（1 学部あたり学科数（平均））＞

Excel で(変数 B の値) / (変数 A の値)という数式で、小数第 3 位まで求めた結果。

＜D（学生総数（人））＞

国公立大学→大学改革支援・学位授与機構「大学基本情報」

https://portal.niad.ac.jp/ptrt/table.html の「2022（R4）」の「(7-A）学生数」ファイル

の値から集計。

私立大学→「大学受験パスナビ:旺文社」の「探す・調べる」
https://passnavi.evidus.com/search_univ/index.html の検索結果のそれぞれの大学ページにおいて、「学生総数」の値を取得。

＜E（学生女子率（%））＞

変数 D の値の取得場所から、男女別の学生数も取得し、
((女子学生数) / (変数 D の値))*100 で小数第 3 位まで Excel で計算。
なお、男女別の学生数が載っていなかった私立大学については、大学ポートレート
https://portraits.niad.ac.jp/ において該当大学のページの学生数を見ることで集計。

＜F（偏差値）＞

上述の「大学受験パスナビ:旺文社」の「探す・調べる」の検索結果のそれぞれの大学ページに記載されている学部ごとの偏差値（幅がある）の中央値を取ってからそれらの平均値を Excel で小数第 2 位まで計算。
なお、下限が BF（ボーダーフリー）である学部については、BF の値を仮に 25（大体偏差値の下限になりそうな値）として計算。
偏差値が載っていない学部については、計算に入れないことにした。

＜G（1 学科あたり入試日程数（平均）＞

上述した大学検索システム Kei-Net で、大学ごとの検索結果を「入試難易度・日程」で表示させた際の行数を入試日程数と捉えることにした上で、(入試日程数) / (変数 B の値）という数式で、Excel で小数第 2 位まで計算。

＜H（学費（円））＞

上述した大学検索システム Kei-Net での検索結果のそれぞれの大学ページにおいて、「大学概要」＞「学費」ページに記載されている「左記合計額」と、「大学概要」＞「学部・学科」の表に「入学定員」として記載されている、該当する学科ごとの定員の積をとった上で、定員数の合計で割ることにより、1 人あたりの学費を Excel で小数第 2 位まで算出した。
なお、公立大学のために、県内者かどうかで学費が変わる場合には、その平均に定員数を

かけることにした。
入学定員が書かれていない学科については、その下の「学生数」の表を見た上で、
「学生数も書かれていない」→計算に入れない
「学生数は書かれている」→他の学部の人数から定員数を推測して計算に入れる
形をとった。

＜I（教授数（人））・J（准教授数（人））＞
上述の「大学受験パスナビ:旺文社」の「探す・調べる」の検索結果のそれぞれの大学ページに記載されている「教員数」項目の中の「教授」「准教授」の値をそれぞれ集計。

＜K（教員1人あたり学生数（人））＞
(変数Dの値) / ((変数Iの値)+(変数Jの値)）で、Excelで小数第2位まで計算。

＜L（キャンパス数）＞
上述した大学検索システムKei-Netでの検索結果のそれぞれの大学ページにおいて、「キャンパス」項目に記載されているキャンパス数を数えた。

＜M（最寄駅からのアクセス（分））＞
上述した大学検索システム Kei-Net での検索結果のそれぞれの大学ページにおいて、次の手順でデータを集計
1. 「大学概要」＞「キャンパス所在地」ページに記載されている一番上に書かれているキャンパスを見る
（ただし、明らかに1都3県以外のキャンパスが一番上の場合（東海大学）のみ、1都3県にあるキャンパスで一番上のものを選択）
2-1. 電車の駅からの徒歩のみ or 電車の駅からの徒歩とバスでの行き方の両方が書かれている場合
→電車の駅からの徒歩のうち、一番上の分数をとる
（「直結」とある場合は0分とカウント。幅がある場合は中央値を小数点以下切り捨て）
2-2. バスでの行き方のみ書かれている場合
→5分とカウント（実際は「下車」なのですぐなのであろうが、バスだと一般に電車からの乗り換え時間があると考えたため）

<N（歴史（年））>

上述した大学検索システム Kei-Net での検索結果のそれぞれの大学ページにおいて、「沿革」項目に記載されている一番古い年を見て、2022 – (一番古い年）で計算。

記載がなかった大学に関しては、それぞれ以下の方法で集計。

・共立女子大学

上述の「大学受験パスナビ:旺文社」の「探す・調べる」の検索結果の大学ページに記載されている「沿革」文章内の創立年を一番古い年として計算。

・産業能率大学

https://www.sanno.ac.jp/admin/organiza/nenpyo.html

(最終閲覧: 2023/02/14)において、「学校」と入っている出来事が起きた最初の年を一番古い年として計算。

・デジタルハリウッド大学

https://ja.wikipedia.org/wiki/%E3%83%87%E3%82%B8%E3%82%BF%E3%83%AB%E3%83%8F%E3%83%AA%E3%82%A6%E3%83%83%E3%83%89%E5%A4%A7%E5%AD%A6

(最終閲覧: 2023/02/14)に記載されている「創立」年を一番古い年とする。

・立正大学はKei-Netでの検索結果のそれぞれの大学ページにおいて、「沿革」項目に記載されている 2 番目に古い年にした（＊立正大学の源流ではなく、＊立正大学開校の起点となるもの）

<O（国公立か私立か（0 か 1））>

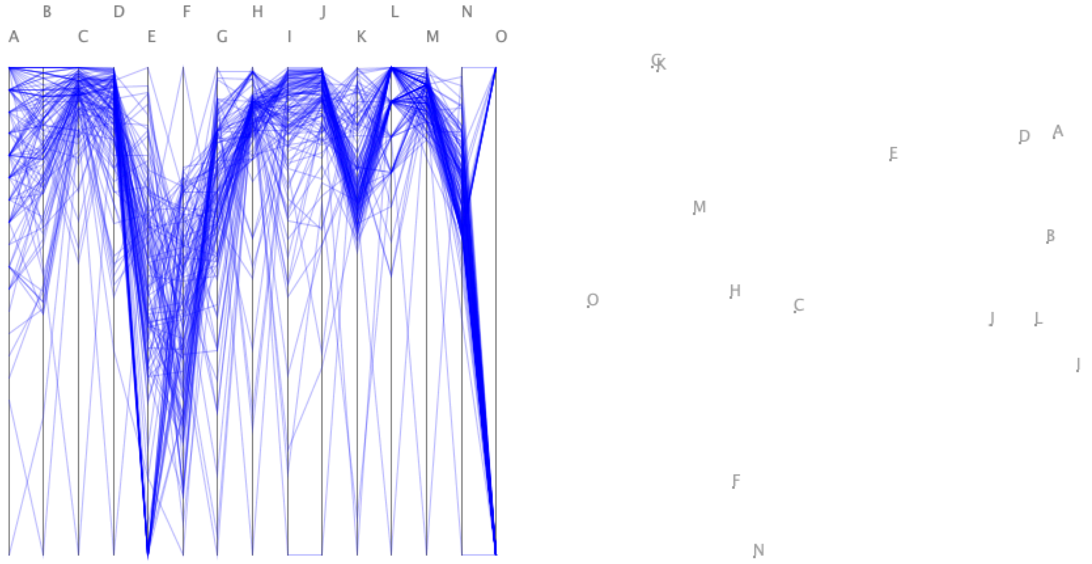
上述した大学検索システム Kei-Net での検索結果のそれぞれの大学ページにおいて、国立か公立か私立かを見て、国公立なら 0、私立なら 1 を割り当てた。

2-3. 実際に作成したデータ

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
2	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
3	3	14	4.667	2010	100.000	60.00	1.71	817800.00	83	66	13.49	1	7	147	0
4	1	3	3.000	3245	11.587	56.25	2.00	817800.00	132	139	11.97	1	5	104	0
5	10	44	4.400	13962	20.069	67.75	0.14	817800.00	1304	984	6.10	3	8	145	0
6	2	6	3.000	1483	59.407	60.00	1.50	1034930.47	167	141	4.81	2	2	94	0
7	3	29	9.667	3606	64.836	62.50	1.38	834800.00	104	79	19.70	1	5	149	0
8	3	8	2.667	1926	31.205	52.50	2.00	817800.00	125	78	9.49	2	10	147	0
9	1	36	36.000	4375	57.074	55.00	1.47	817800.00	145	97	18.08	1	25	149	0
10	2	15	7.500	2004	64.072	61.25	1.00	1340000.89	122	91	9.41	3		135	0
11	6	17	2.833	4803	12.929	65.00	0.35	947400.00	364	335	6.87	2	1	141	0

3. 可視化結果

3-1. 初期状態の可視化結果



最初に何も設定しなかった場合、上記のような可視化結果が得られた。

この時点で読み取れることとして、以下がある。

【散布図から読み取れること】

以下の変数同士が散布図上で近くに位置することから、相関がありそうだと予想できる。

○G（1学科あたり入試日程数（平均））と K（教員1人当たり学生数（人））…①

○A（学部数）と D（学生総数（人））…②

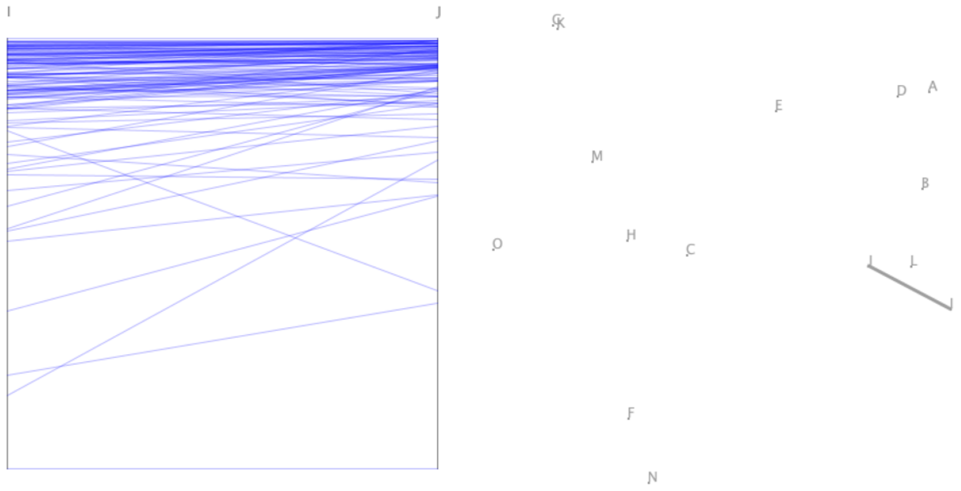
○L（キャンパス数）と J（准教授数（人））…③

以下、上記で予想したことも念頭に置きながら、Culling スライダーや画面右上のスライダーを動かしたり、Clustering を活用したりすることで、各変数の相関について見ていく。

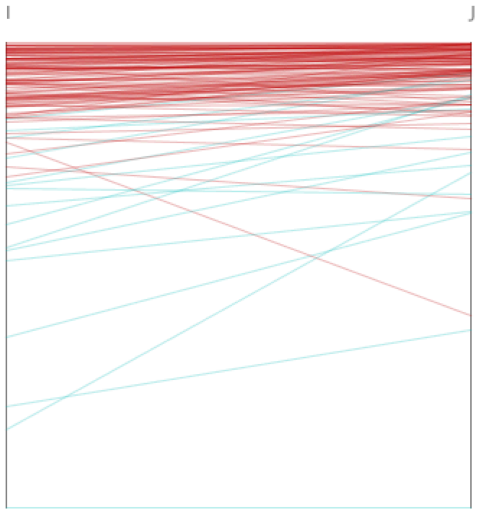
3-2. 右上のスライダーのみを変化させる（Culling スライダーは動かさない）

【 I. 教授数と准教授数の相関】

右上スライダーを上げていくと、最初に I（教授数（人））と J（准教授数（人））の 2 変数のみを軸とする折れ線群が得られた。

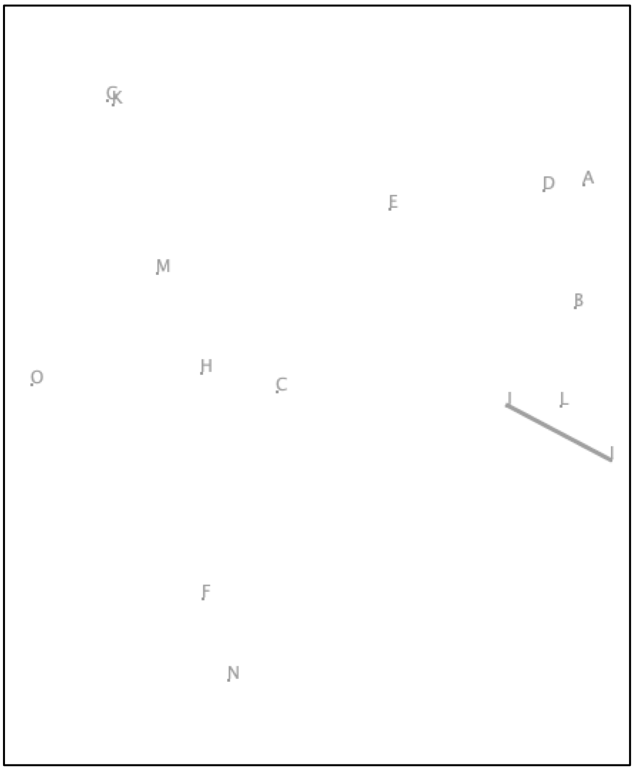


Clustering=2 としてみると、以下の図が得られる。



上図において I の軸において上の方にある赤色の折れ線群は J の軸においても上の方にあり、I の軸において下の方にある緑色の折れ線群は J の軸においても赤色の折れ線群よりは下にあることがほとんどであるため、教授数と准教授数に正の相関があると読み取れる。この結果は、I（教授数（人））（あるいは J（准教授数（人）））と A（学部数）、B（学科数）、D（学生総数）との間の相関よりも、教授数と准教授数の間の相関が強いという点で意外であった。

実際、ここで右側の散布図に着目すると、



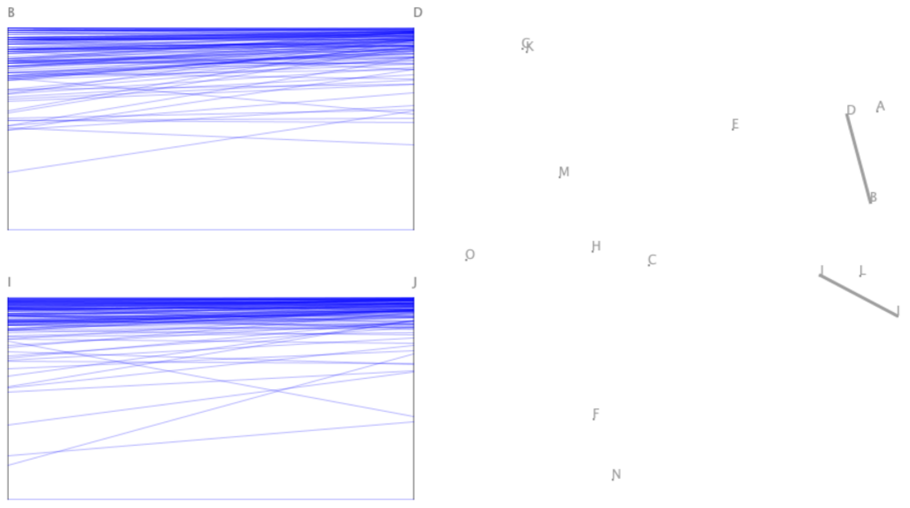
I（教授数（人））とJ（准教授数（人））の距離よりも、I, J それぞれと A（学部数）、B（学科数）、D（学生総数（人））との距離の方が遠いことが確認でき、ここからも A（学部数）、B（学科数）、D（学生総数）との間の相関よりも、教授数と准教授数の間の相関が強いと言える。

この結果の考察としては、文部科学省の「2. 大学教員の職の在り方について」
https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/houkoku/attach/1342439.htm
の「3. 助教授について　(2)准教授の新設　1. 准教授の職名・職務内容」の項目の中で、
「大学制度上、この准教授のうち、更に研鑽を積み、大学や学部等全体における教学面の運営全体についての第一次的な責務を担う（最終的な権限と責任は学長にある。）に相応しい業績や資質能力が認められた者が昇進して教授になると考えられ、」
と書かれており、大学制度上、大学内で最終的には学長が権限と責任をもって准教授から教授に昇進させるプロセスがある、つまり准教授から教授への昇進は同一大学内で起こるのが一般的だと考えられるため、おおよそどの大学でも教授数と准教授数は正の相関を持つのだと考えられる。

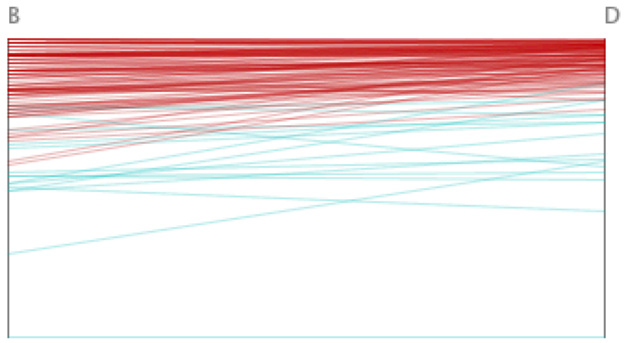
【 II. 学科数と学生総数の相関】

さらに右上スライダーを上げていくと、B（学科数）と D（学生総数（人））の 2 変数の

みを軸とする折れ線群が追加で得られた。



こちらも、先ほどの【 I. 教授数と准教授数の相関】での操作と同じように、Clustering=2 としてみると、以下の図が得られる。



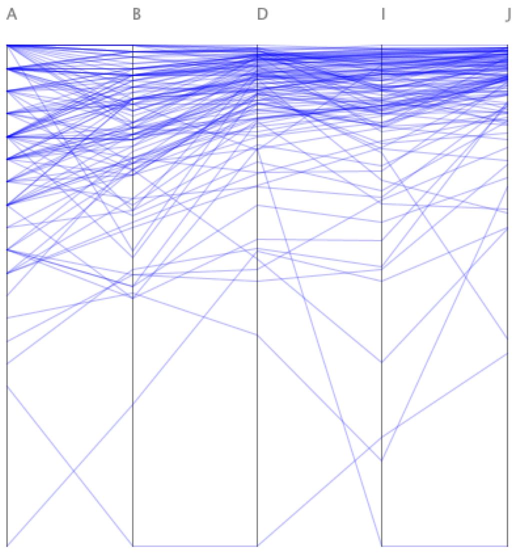
こちらも、先ほどの【 I. 教授数と准教授数の相関】と同じように、上図において B の軸において上の方にある赤色の折れ線群は D の軸においても上の方にあり、B の軸において下の方にある緑色の折れ線群は D の軸においても赤色の折れ線群よりは下にあることがほとんどであるため、学科数と学生総数に正の相関があると読み取れる。

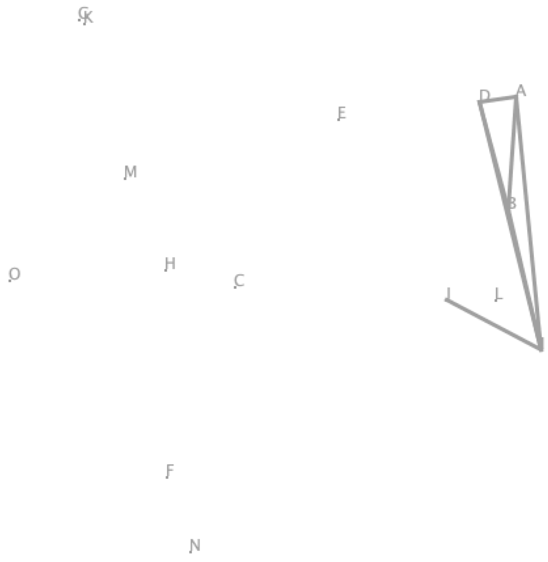
これは学科数が多ければ、学生数も多いという点で、イメージとも一致する理解しやすい結果であった。

【 III. 学部数・学科数・学生総数・教授数・准教授数の相関】

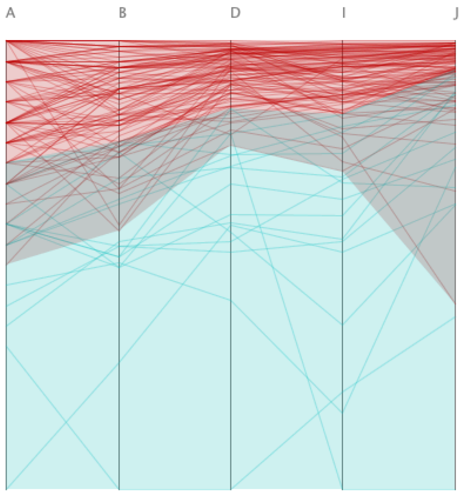
さらに右上スライダーを上げていくと、A（学部数）、B（学科数）、D（学生総数

（人））、I（教授数（人））、J（准教授数（人））の 5 変数を軸とする折れ線群が得られた。なお、このうち、A（学部数）と D（学生総数（人））の相関は、「3-1.初期状態の可視化結果」において【散布図から読み取れること】の一つとして、②と名前をつけていた相関にあたる。





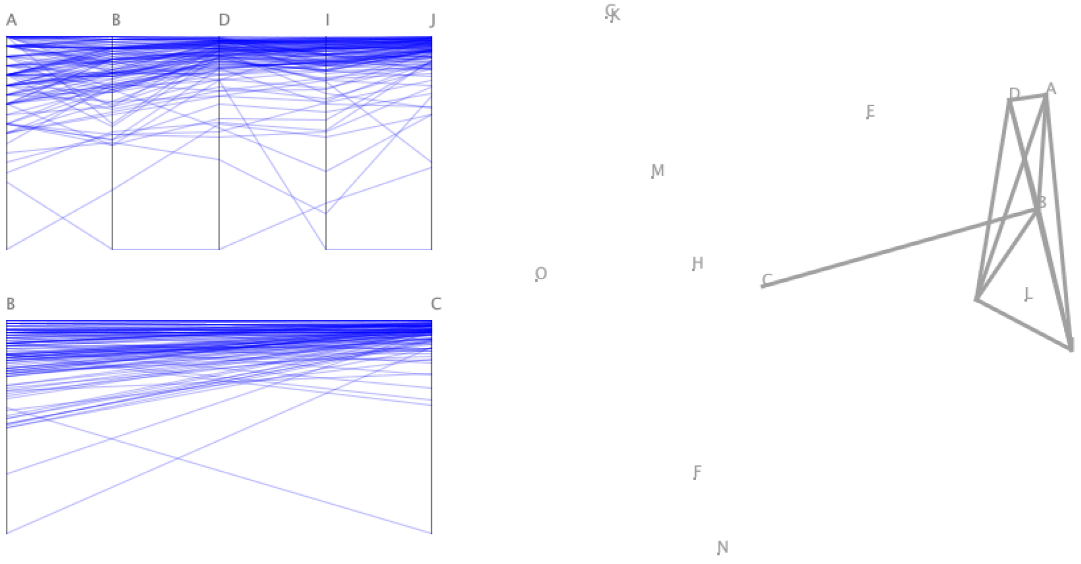
Clustering=2 として、Paint で塗りつぶしてみると、以下の図が得られる。



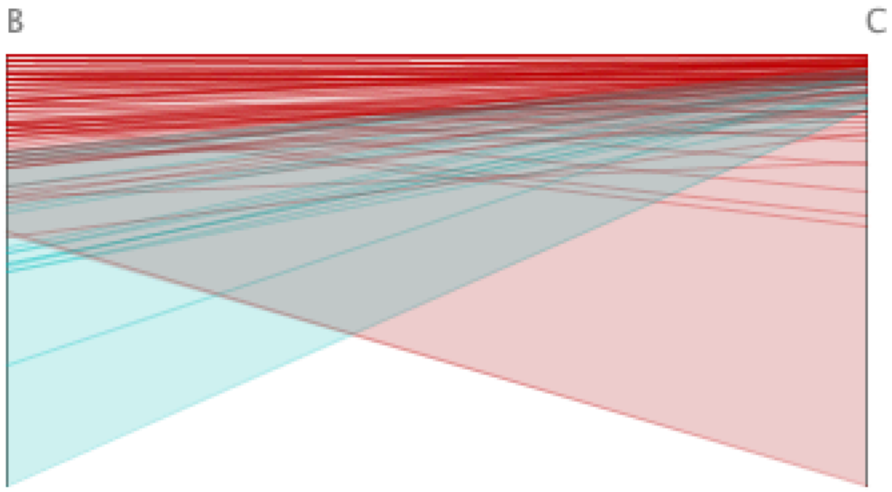
どの 5 変数の軸でも上の方にある赤色の折れ線群とどの 5 変数の軸でも下の方にある緑色の折れ線群に分かれていることがわかるので、これら 5 変数には正の相関があるとわかる。

つまり、学部数、学科数、学生総数（人）、教授数（人）、准教授数（人）の 5 変数には正の相関があるとわかる。
このことは、学部数や学科数が多ければ、それだけ所属する構成員が多く、学生も教授も准教授も多いというイメージに一致する。

【 IV. 学科数と 1 学部あたり学科数（平均）の相関】
さらに、スライダーを上げてみると、B（学科数）、C（1 学部あたり学科数（平均））の 2 変数を軸とする折れ線群が得られる。



こちらも Clustering=2 として、Paint で塗りつぶしてみると、以下の図が得られる。



B の軸において、赤色の折れ線群は上の方に位置し、緑色の折れ線群が下の方に位置するが、

C の軸においては、確かに上の方に赤色の折れ線が一定数あるものの、緑色の折れ線が赤色の折れ線より上の方に位置していたり、逆に赤色の折れ線が緑色の折れ線より下の方に位置していたりする様子が見られる。

ここから、B（学科数）と C（1 学部あたり学科数（平均））には負の相関があると読み取れる。

この結果は、学科数が多いほど、1学部あたり学科数が少なく、

学科数が少ないほど、1学部あたり学科数が多いということを表している。

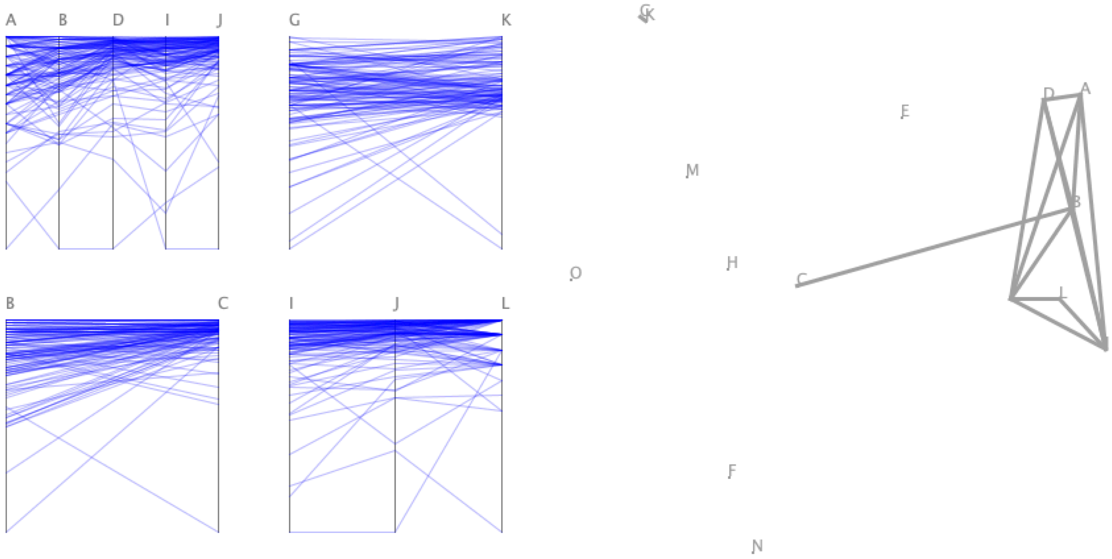
学科数が多いほど、C（1 学部あたり学科数（平均））が少なくなるということは、B（学科数）が多い大学ほど、学科の数に対して、A（学部数）もかなり多く設置しているということになりそうである。つまり、ただ単に絶対的な学部数と学科数に正の相関がある（【 III. 学部数・学科数・学生総数・教授数・准教授数の相関】より）だけではなく、学科を多く設置する大学の方が、学科をそこまで設置しない大学に比べて、相対的にも多くの学部を設置することがわかる。

この結果の考察としては、文部科学省の「国立大学の学生定員管理の柔軟化について」

https://www.mext.go.jp/content/20200528-mxt_hojinka-000007541_4.pdf

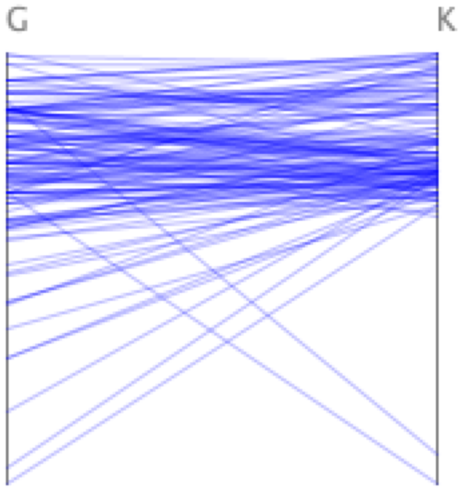
という資料 p.3の「定員管理の仕組み（学部）」の表から私立大学では「収容定員の範囲内の学部・学科等の整備」が「届出」となっていることから、届出に慣れている私立大学は学部も学科もたくさん整備するのではないかと考えられる。

さらに右上スライダーを上げていくと、G（1学科あたり入試日程数（平均））とK（教員1人あたり学生数（人））の2変数を軸とする折れ線群、I（教授数（人））、J（准教授数（人））、L（キャンパス数）の3変数を軸とする折れ線群の2つの折れ線群が追加で得られる。

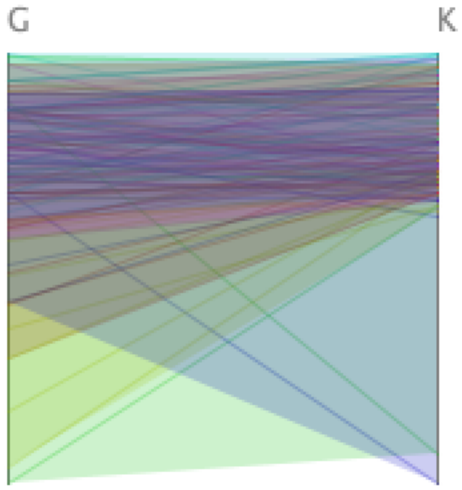


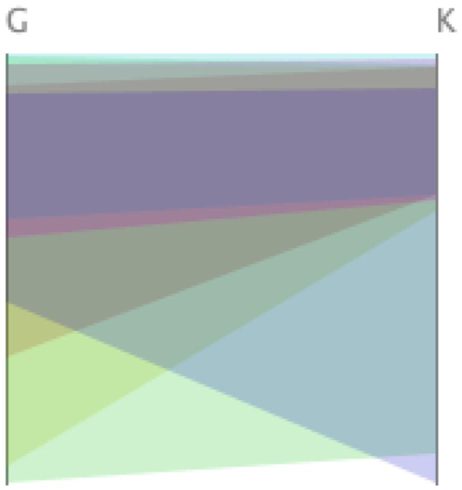
【 V. 1 学科あたり入試日程数（平均）と教員 1 人あたり学生数の相関】

まず G（1 学科あたり入試日程数（平均））と K（教員 1 人あたり学生数（人））の 2 変数間の相関を詳しく見てみる。なお、この相関は、「3-1.初期状態の可視化結果」において【散布図から読み取れること】の一つとして、①と名前をつけていた相関にあたる。



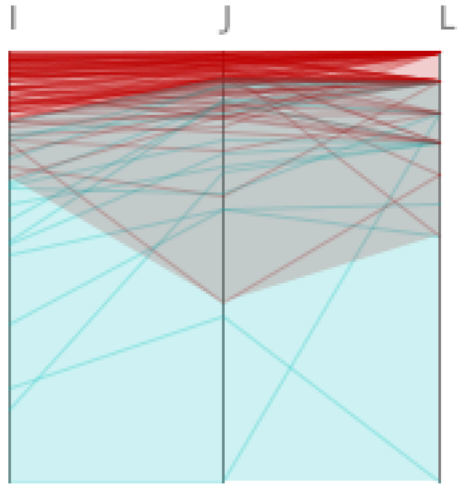
すると、やや見にくくはあるが、2つの軸どちらも上の方に折れ線群が密集しているものの、その中で折れ線群が交差している様子が見られる。
つまり、密集している折れ線群の中で、
Gの軸において上の方にあった折れ線がKの軸では下の方に行き、
Gの軸において下の方にあった折れ線がKの軸では上の方に行っている様子が見られる。
また、このように折れ線群が交差している様子は、あまり折れ線群が密集していない下の方でも見ることができる。
ここから、G（1学科あたり入試日程数（平均））とK（教員1人あたり学生数（人））には負の相関があるのではないかと推測できる。
つまり、この結果は
1学科あたり入試日程数（平均）が多いほど、教員1人あたり学生数（人）が少なく、
1学科あたり入試日程数（平均）が少ないほど、教員1人あたり学生数（人）が多い
ことを示していると考えられる。
ただ、私は、「1学科あたり入試日程数（平均）」は多いほど、たくさんの受験生に受けてもらいたい大学、つまりあまり人気のない大学、というイメージであり、「教員1人あたり学生数（人）」は多いほど、面倒見の良い大学ということで人気があるというイメージを持っていた。そのため、この2変数は負の相関になるのではないかと予想していたため、この結果はかなり意外であった。
そこで、少し表示方法を変えて違う角度から考察してみる。
この2変数を軸とする折れ線群をClustering=6として、Paintで塗りつぶしてみると、以下のようになる。
（左図は、折れ線群も表示されているもの。右図は、折れ線群の表示をなくしたもの。）





上図を見ると、Gの軸において、上半分に幅広く広がっている折れ線群がKの軸において全体に広がっている様子を示す紫色の折れ線群以外は、色分けからもGの軸での上下の位置関係とKの軸での上下の位置関係がそこまで逆転していないことがわかる。
ここから、G（1学科あたり入試日程数（平均））とK（教員1人あたり学生数（人））には必ずしも負の相関があるとは言えないのではないかと推測できる。
この2変数の相関については、表示方法によって考察が変わってきてしまい、今回はあまりはっきりとした相関を読み取れなかったが、機会があれば別の手法で試してみたい。

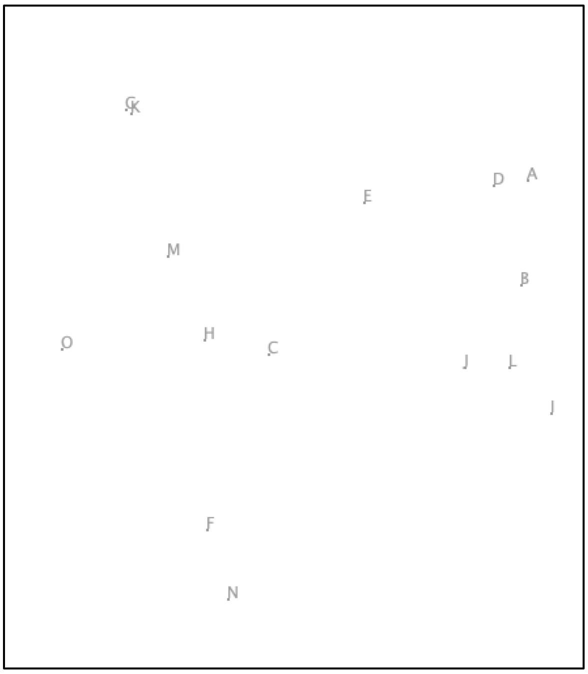
【 VI. 教授数、准教授数、キャンパス数の相関】
追加で得られた2つの折れ線群のうち、二つ目のI（教授数（人））、J（准教授数（人））、L（キャンパス数）の3変数の相関を詳しく見てみる。
I（教授数（人））、J（准教授数（人））、L（キャンパス数）の3変数を軸とする折れ線群をClustering=2として、Paintで塗りつぶしてみると、以下のようになる。
なお、このうち、L（キャンパス数）とJ（准教授数（人））の相関は、「3-1.初期状態の可視化結果」において【散布図から読み取れること】の一つとして、③と名前をつけていた相関にあたる。



上図から、どの3変数の軸でも上の方にある折れ線群とどの3変数の軸でも下の方にある折れ線群に分かれていることがわかるので、これら3変数には正の相関があるとわかる。つまり、教授数（人）、准教授数（人）、キャンパス数の3変数には正の相関があるとわかる。

ここまでの操作では、一度もキャンパス数の変数が出てこなかったため、ここでキャンパス数と一緒に出てきた変数は最もキャンパス数と強い相関を持つ変数であると考えられ、この結果は、キャンパス数が学生総数や学部数、学科数よりも教授数や准教授数とより強い正の相関を持つことを示していると考えられる。

実際、ここで散布図を確認すると、



I（教授数（人））、J（准教授数（人））、L（キャンパス数）間の距離よりも、I, J, L それぞれと A（学部数）、B（学科数）、D（学生総数（人））との距離の方が遠いことが確認でき、ここからも A（学部数）、B（学科数）、D（学生総数）との間の相関よりも、教授数（人）、准教授数（人）、キャンパス数の間の相関が強いと言える。

この点が、私にとっては意外なものであった。

この結果の原因としては、以下のようなことがあるのではないかと考えた。

教授や准教授といった教員に対しては、教員が居室を大学に持つなどするためその分の大学の場所を用意する必要がある。

一方で、学生に対しては場合によっては学部や学科共通の教室で授業を受けられるようにしさえすれば、それ以外の場所は極端に言えば用意しなくても良い。

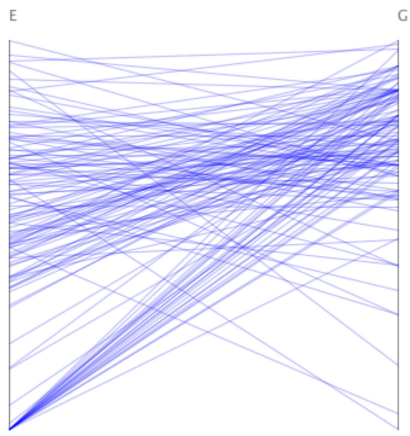
ここで、大学で用意する必要のある場所が多ければキャンパス数が多く必要になると考えれば、学生総数や学部数、学科数よりも教授数や准教授数とキャンパス数がより強い正の相関を持つことは自然なのではないかと考えられる。

3-3. 右上のスライダーと Culling スライダーを同時に動かす

次に、Cullingスライダーも右に動かしながら、右上のスライダーも上げていく。

【 VII. 学生女子率と1学科あたり入試日程数（平均）の相関】

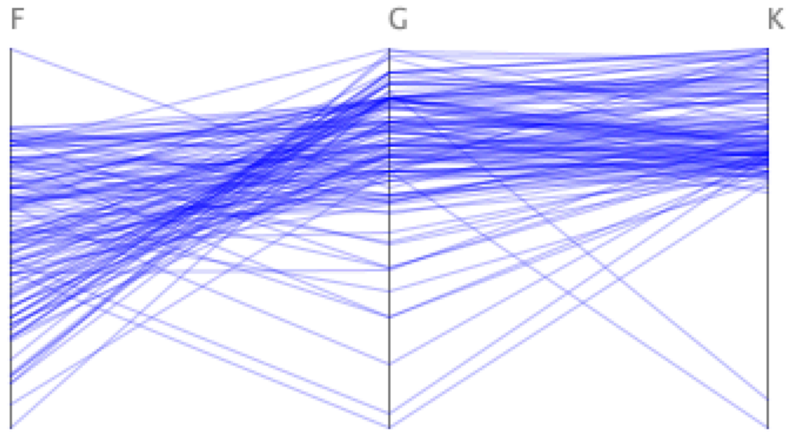
すると、E（学生女子率（%））と G（1 学科あたり入試日程数（平均））の 2 変数を軸とする折れ線群が得られる。



上図では、折れ線群が交差している様子が見られることから、E（学生女子率（%））と G（1 学科あたり入試日程数（平均））には負の相関があると考えられる。

特に、上図をよく見ると、E（学生女子率（%））の値が最大値になっている折れ線群は G（1 学科あたり入試日程数（平均））の軸において、かなり上の方にあるとわかる。
ここから、E（学生女子率（%））の値が最大値になっている大学、つまり女子大は 1 学科あたり入試日程数（平均）が少ないと考えられるが、その原因として、「女子大を受ける人は女子大をあえて希望する」ことがあるという話を何度か聞いたことがあるため、女子大は各入試で受かった人は入学してくれる割合が高いと見て、そこまで入試日程を多く用意していないのではないかと考えた。

【 VIII. 偏差値と 1 学科あたり入試日程数（平均）の相関】
さらに右上スライダーをあげると、F（偏差値）と G（1 学科あたり入試日程数（平均））、K（教員 1 人あたり学生数（人））の 3 変数を軸とする折れ線群が得られる。



G, K 間の相関については、【 V. 1 学科あたり入試日程数（平均）と教員 1 人あたり学生数（人）の相関】で考察したため、ここでは、F（偏差値）と G（1 学科あたり入試日程数（平均））の間の相関を見る。
上図では、折れ線群が交差している様子が見られることから、F（偏差値）と G（1 学科あたり入試日程数（平均））の間には負の相関があると考えられる。
この結果は、【 V. 1 学科あたり入試日程数（平均）と教員 1 人あたり学生数（人）の相関】の考察でも書いたように、「1 学科あたり入試日程数（平均）」は多いほど、たくさんの受験生に受けてもらいたい大学、つまりあまり人気のない大学、というイメージであることを踏まえると、G（1 学科あたり入試日程数（平均））が多いほど、F（偏差値）が低くなるという負の相関はイメージに一致する結果であった。

4. まとめ・感想

1都3県の大学について、さまざまな変数から比較し、イメージ通りであったこと、意外であったことの両方の結果を得ることができた。

ただ、「1.はじめに」でこのテーマを選んだ理由として書いた「国立大学の良さ」については、今回の結果からは特に考察することができなかったため、機会があれば今後別の可視化手法や、別のデータ取得方法を用いて考察したい。

また、データを集める際に、どのようなデータが適切か（「絶対的な数にするのか、比にするのか」や「ただ単に平均を取るので良いのか」など）を吟味する必要を強く感じたため、今回あまり相関が見られなかった変数が本当にその取得方法でよかったのかも含めて今後の課題としたい。