可視化からの知識探索
― 可視化は何をみえる化しているか？ ―

Knowledge Exploration from Visualization
― What does Visualization Make Visible? ―

南 俊朗
Toshiro Minami

【要　約】

我々は多くの「可視化」図形に取り囲まれている．小学生以来お馴染みの折れ線グラフ，棒グラフ，円グラフによる統計データの表示を始め，人やものなどの関連性を線でつないで表現する関係グラフ，さらには天気図など，可視化された図形は様々である．プレゼンテーションに関する図書を開くと，良いスライド作成のコツとして，言葉ですべてを説明するのではなく内容を図解して示すことが推奨されている．それでは，なぜグラフや図解などの可視化手法を用いると，我々人間にとって理解が容易になるのであろうか？可視化の持つどのような要因が理解を助けるのであろうか？本稿の目的は，この素朴な疑問への解答への糸口を求め，模索の第 1 歩として，可視化の持つ要因や特性を分析することである．最初に，可視化とは何か，具体的な可視化手法にはどのようなものがあるのかを探索する．次に，いくつかの可視化手法を取り上げ，それがどういう理由で我々の理解を助けるかを分析する．その結果は，我々は流れに基づき図形の形状を認識し，その変化として角などを捉えていることや，長さや角などを比較により種々の特性抽出を行っているということである．可視化技術はこの認識機構を活用している．最後に，本稿の分析結果をまとめ，今後の課題や方向性について展望する．

キーワード： 可視化の原理，グラフ表現，理解容易性，視覚化のコツ

[Abstract]

We are surrounded by quite a lot of “visualized” figures. For example, we see line graphs, bar charts, pie charts for representing statistical data, relation graphs to visualize the relationship between human and objects. Further, weather maps are used in the weather forcasting programs in TVs and network sites, and we are very familiar with them. As we have a look of a book dealing with presentation techniques, the authors insist to utilize schematic representations for effective presentations instead of using texts only. Our question is, why it becomes easier to recognize and understand what are explained if we use such schematic representations? What aspects of visualized representations help us recognize and understand what are represented? In this article, we pursue our preliminary analysis in order to find a possible answer to such a naïve question about visualization. We start with searching for visualization methods, and then we analyze and discuss why these methods help us recognize the data easier. As a result, we conclude we recognize based on the stability and change of properties and visualization technologies utilize such phenomena of humans. Finally, we try to find a solution to our question by accumulating our analysis results in this article.

Keywords: Principles of Visualization, Graph Representation, Understandability, Tips for Visualization

1　はじめに

今はデータの時代である．膨大なデータが日々生成され蓄積されている．その背景には磁気ディスク（HDD）の大容量化や低価格化の進行がある．Kryder の法則と呼ばれるように指数関数的な大容量化の傾向は今後も継続する勢いである．また，有線通信だけではなく，無線通信によるネット接続環境が整備された結果，いつでもどこからでもネット利用が可能となった．しかも世代が進むにつれて接続速度も高速化されてきている．この結果，Web 検索だけではなく，ネットショッピングなどネット環境で提供される情報サービスに関する記録（ログデータ）の容量が急増した．

センサーの普及も蓄積データ量増加の一因である．多くの人達が用いているスマートフォン（スマホ）には，タッチセンサーやカメラ，マイクなどの利用者が意識的に用いる入力装置だけではなく，GPS，加速度，磁気などのセンサーが組み込まれていることが多い．これらのセンサー機器の入力データは様々な形態により保存されている．今後は，たとえば健康管理に役立つデータをスマホやウェアラブル端末などにより収集し，それを蓄積することが普及するものと考えられる．このことも，これからのデータ量の増大を後押しする．

膨大な量の蓄積データは，最近ではビッグデータと呼ばれる．この用語はデータ量の膨大さを表しているが，重要なのはデータ量そのものではない．ビッグデータが注目されるのは，これらのデータを解析することで，これまで知られていなかった，あるいは知ることが困難であった新しい知見を得ることができ，それがビジネスや社会に対して有益な情報として活用できるからである．

ビッグデータはデータ自体が膨大であるため，従来のデータ分析手法をそのまま適用するだけでは時間がかかり過ぎたり対象が膨大すぎて不十分な情報しか得られなかったりする．その解析ツールとして，データ可視化 (Data Visualization) は必要不可欠である．従来の統計解析手法や相関ルールなどのデータマイニング手法を対象データにそのまま適用するだけでは的確な結果は得られない．データの特性を踏まえた解析ツールを組み合わせることで初めて良好な解析が可能となる [10][15]．可視化ツールを用いることで全体の概要を把握したり，データの特性を直観的にすばやく把握できるようになり，それを踏まえて焦点を絞った分析を適切に行うことが容易になる．

しかし，可視化の設定によっては異なる結論が導出されてしまうこともある．たとえば，記事 [7] では，半導体の集積度に関するムーアの法則に陰りがでているとの日経エレクトロニクスの記事はグラフの y 軸の取り方の誤りによるものと指摘しており，異なる結論を導いている．本例は，可視化の利用には十分な注意を払う必要があることを示しており，忘れてはならない．

ここで生じる素朴な疑問は「そもそも可視化とは何なのか？」,「なぜ，可視化された方が見やすかったり，理解が容易であったりするのか?」である．本疑問は素朴ではあるが，可視化についての根源的な問いかけである．しかし，この疑問に答える研究は，さほどなされていない．日本の論文検索システム CiNii [2] において「可視化」というキーワードで検索すると 28,759 件（2015 年 11 月 18 日現在）が検索される．しかし，そのほとんどは「〜の可視化」や「〜を可視化する」などのタイトルとなっており，それぞれの対象に対する可視化のケーススタディ論文である．可視化の原理や理論を根源的に問う論文は見当たらない．キーワードを「可視化　原理」とすると 182 件,「可視化　理論」では 475 件が見つかるものの，可視化の原理や理論に関するものはやはり見当たらない．

一方，英語論文においては「Visualization Theory」に関する可視化の理論に関する研究が存在する [13]．たとえば，Rogowitz[14] は色に関する人間の特性も考慮した Visualization の原理を解説している．可視化とは何かを解明するためには，様々な観点からの問いかけが必要であり，まだまだ多くの研究が必要である．

このような現状認識を受けて，本稿では,「可視化とは何か」という素朴な疑問への解答の糸口を求めて，我々が小学校以来なじんできた棒グラフや折れ線グラフなどの基本的なグラフ表現を対象に，可視化された図形が我々の直観的な情報認識とどのように関わっているのかを視覚の基本的なメカニズムに基づき考察する．

この目的へ向け本稿は以下のように構成される：

まず第 2 節で可視化とは何かを議論する．可視化のためのグラフ表現に関する起源を確認し，可視化をその狙い，すなわち，何を可視化しようとしているのかの目的によって分類を試み，また，グラフ分析に用いる我々の目の特性を整理する．

それを受けて第 3 節では棒グラフ，折れ線グラフ，円グラフなどのグラフ図形について，それらの図形が可視化の観点からどのような性格をもっているのか，その性格を踏まえて，どのような用途に適しているのかを個別に考察する．

最後に，第 4 節では本稿における分析や考察の結果をまとめ，また今後に残された課題を展望する．可視化に関する我々の追究は素朴な疑問から始めたばかりであるものの，可視化の本質を理解するためには必要で重要な第一歩である．

2　可視化とは何か

2.1　グラフのはじまり

可視化の手段としてもっとも良く用いられるのは統計的なデータを表示するためのグラフであろう．代表的なグラフとしては，棒グラフ，折れ線グラフ，円グラフなどがある．このようなグラフを最初に考案したのは 18 世紀の William Playfair と言われている [16]．Playfair のグラフ（Chart）には，現在のグラフでも用いられている補足要素が既に備わっている．図 1 は折れ線グラフの例である．確かに，グラフのタイトル，領域を示す枠線，軸目盛，キャプション，折れ線へのラベルなどが描かれている．図 2 は棒グラフの例である．横軸の目盛ラベルが上に，棒のラベルが右にあるなど，現在の典型的な棒グラフと比較して若干の違いはあるものの，基本的な形状は同じである．Playfair はそれ以外に，円グラフなど様々なグラフ形状を考案している．Playfair のおいたちを含めた初期のグラフに関しては文献 [16] が詳しい．

2.2　可視化とは

本稿で用いる「可視化」は，Visualization の訳である．Information Visualization とも呼ばれる．

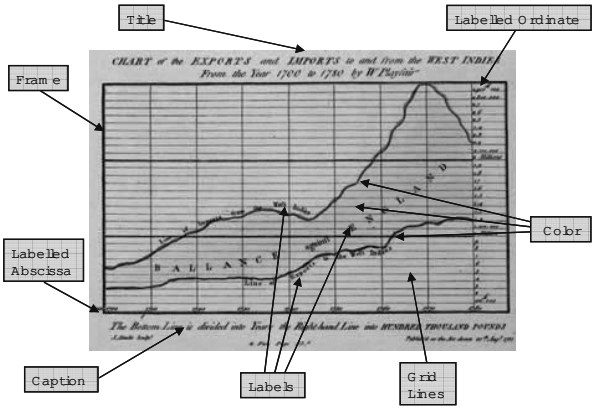

図 1: William Playfair の折れ線グラフ [16]

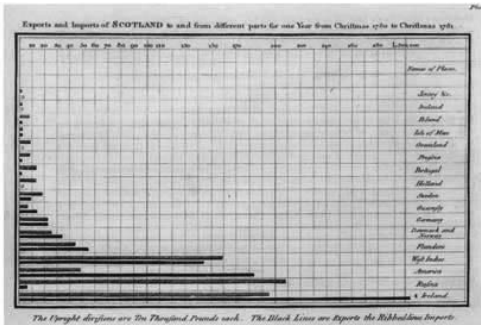

図 2: William Playfair の棒グラフ [16]

日本語訳としては可視化と共に「視覚化」という用語も用いられる．また,「みえる化」という表現も目にする．これらの用語を使い分ける場合もあるが，本稿では，特に断らない限り「可視化」という表現を用いる．

本稿の最大の目的は「可視化とはそもそもどういうものであろうか？」という疑問への解答の糸口をつかむことである．

一口に可視化と言っても，その目的や重点がおかれるポイントによって様々なタイプが考えられる．たとえば，可視化の目的を次の 2 つに分けることができる．

- **（文字通りの可視化）**（本来は）見えないモノやコトを見えるように，あるいは何らかの形のあるものとして表現する意味での可視化
- **（見えやすさ）** 元々見えるものであっても，それをより見やすく，より理解しやすくする意味での可視化

これら２つのタイプは相反するものではなく，両者を同時に実現しようとすることもある．たとえば，多次元の現象を次元を落として２次元表示するという可視化技術においては，前者の意味での可視化と同時に後者の意味での可視化の両者を同時に狙っていることも多い．

次に，別の観点から可視化をタイプ分けする．

- **（客観性重視）** データをありのままに表現することを主な目的とする
- **（特性の強調）** ある意図をもって特定の性質自体やその性質に関する差異を強調するために，我々の目の特性を考慮してその性質が際立つように表現することを目的とする

客観性重視の可視化の適用例として探索的データ解析（Exploratory Data Analysis, EDA [9]）がある．EDA ではデータ全体の傾向や特徴を把握するなどの際に可視化が用いられる．そのような目的のためには，箱ひげ図や散布図は非常に強力なツールとなる．

一方，性質を強調する可視化が有効な場面としては，プレゼンテーション [1] がある．プレゼンテーションでは，データ自体やデータ間の関連や違いを正確に伝えることだけではなく，伝えたい目的やメッセージに応じて，印象深い表現を用いることが求められる．そのような性格の違いにより伝えたいポイントをより強調して表示するなどの工夫が必要となる．

このように可視化というテーマにはさまざまな側面がある．本稿では，可視化というものを，その根本にさかのぼって考察することを目指す．すなわち，様々な可視化技術によって，どういう部分が見えるようになったのか，ヒトは可視化された図形のどの部分に注目し，どういう情報を読み取っているのかなどを低レベルの認知のレベルから分析し可視化のもつ可能性をより客観的に理論化するための第一歩を踏み出すことが本稿の最大の目的である．

その最初の試みとして，本稿ではグラフの持つ可視化機能を検証する．グラフの持つ属性には，形の他，色，大きさなど様々な要素がある．本稿では，特に形（形状）による可視化に注目する．

2.3　本稿で用いる我々の目の特性

本節では我々人間の目（視覚）の基本的特性について簡単な考察を行う．第３節ではそれを踏まえて，可視化機能の観点から，棒グラフや折れ線グラフ，円グラフなど我々が慣れ親しんできた何種類かのグラフに対する分析を行い，その特性を検討する．

まず確認しておくべきことは，神経系の大きな特性の１つは変化に強く反応することである．この特性を視覚に当てはめると，視覚野の中で，明るさや色などの属性が変化する場所（点）は他の場所よりも強く知覚されることを意味する．

形状を認識するためには，明るさの変化の知覚がポイントになる．明るさが変化する部分としてエッジ（境界）が認識される．エッジ点の繋がりとして輪郭（境界線）が認識される．境界線の繋がりとして，全体形状が認識される．

変化の逆概念として，流れ（トレンド）がある．ある傾向が継続するならば，その部分は，変化のない部分として認識される．たとえば，直線はエッジとして認識された点の集まりが一定方向に繋がったものである．方向がある割合で変化する図形として曲線が認識される．一定の変化が継続すると，形成された図形全体は円や円状の図形として認識されることになる．

直線の端点は一定の流れで繋がっていたエッジが途切れる場所として強く認識される．直線と２つの端点が線分を構成する．２つの線分がある端点を共有する場所が角になる．

３つの端点（頂点）を３つの直線（辺）がつなぐ形として三角形が認識される．１つの端点をいくつかの直線（線分）が共有する図形として放射線状（星形）の図形が認識される．アスタリスク文字（＊）は，この形を文字として抽象化したものである．

流れとしての認識は慣性を持っている．すなわち，端点で終了したはずの点の流れの，その先に再びその流れに乗っている線があると，それら全体で１つの線を構成しており，その一部が欠けていると認識される．破線や点線，一点鎖線などの線図形は，このような認識によって１つの図形として認識されたものである．

3　可視化手法の解釈

本節では前節で示した我々の視覚認識の特性に基づき基礎的可視化手法であるグラフを例に，それぞれがどのように可視化に役立つのかを分析する．

3.1　棒グラフ

グラフの可視化に関する分析の最初の事例として棒グラフを取り上げる．棒グラフは，様々なグラフの中でも特に基礎となる．基本となる図形は長方形であり，数値の大小を直観的に表現するものである．

図 3 に棒グラフの構成要素である長方形がどのように認識されるかを考察する．図左 (a) は棒グラフの基本形を示す．1 つの棒は 4 つの辺から構成されている．2 本の垂直方向の辺が 2 本並行に並び，同様に 2 本の水平方向の辺が並ぶ．

図中 (b) に (a) の図に対して我々が注目するポイントを示す．我々は，まず，明暗の境界である辺に注目する．破線で辺を示す．辺の中でも特にその端点に注意が注がれる．長方形の場合は，2 つの辺が接続されているため，端点は角となる．(b) では角の部分が実線表示されている．

長方形の持つこのような特性から，ヒトの目は幅と高さに注目することが分かる．図右 (c) に 2 つの長方形（棒）を並列させた図を示す．棒グラフでは棒の幅（横棒グラフでは高さ）が一定であるため，我々の注目点は自然と高さに向かう．高さを比較することにより，2 つの項目の相互関係を

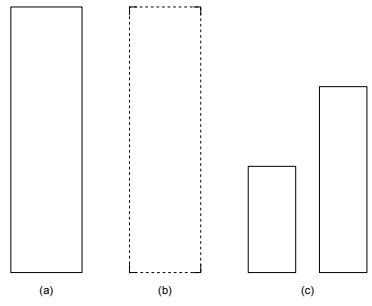

図 3: 長方形からの情報の読み取り．(a) 長方形，(b) 四隅の角と角と角の連結，(c) 2 つの長方形（棒）の比較

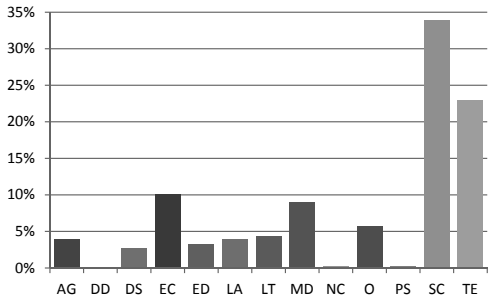

図 4: 棒グラフの例

見い出そうとする．相互関係としてまず考えられるのは高さ（長さ）の差である．それと同時に高さの比にも注意が注がれる．本例では右側の棒は左側の棒の 2 倍弱と読み取れる．

図 4 に棒グラフ（縦棒グラフ）の実例を示す．本グラフは，九州大学附属図書館の 1 年分の貸出記録データから，学部別に貸出冊数の総数に対する割合を示した棒グラフである．まず目につくことは右端の 2 つの学部 SC（理学部）と TE（工学部）の貸出冊数がずば抜けて多いことである．これらが第 1 のグループを構成する．これらを除くと他の学部はいずれも半分以下の割合に過ぎないことが分かる．

次に，これらの 2 学部を除いた学部群の中での比較を行うと，EC（経済学部）と MD（医学部）が多く，少し離れて O（その他）があることが分かる．さらに残りの学部を比較すると，それらはいずれも 5 %以下であり，その中で，DD（歯学部），NC（21 世紀プログラム），PS（薬学部）の 3 学部はほぼ 0 %であることが分かる．

前段落で示したグラフの解釈過程は著者自身の例であるが，多くの人が同様な読み方をするものと考えられる．このような解釈により，学部全体を貸出冊数の大小に関して，多数の図書を借りる上位グループの 2 学部，それに次ぐ中位グループの 3 学部，多数派を構成する 5 学部，そして，ほとんど貸出のない 3 学部とグループ分け（クラスタリング）できる．棒グラフの 1 つの役割は，直観的に大小比較を行うことを促し，全体をグループ分けして捉えることができることである．

さらに細かく比較すると，上位グループを構成する 2 学部の貸出冊数割合には 1.5 倍弱程度の差

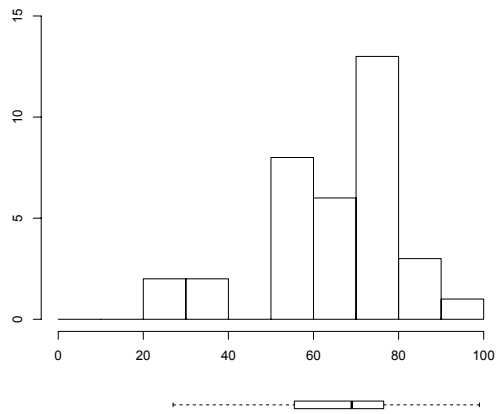

図 5: ヒストグラムの例（ヒストグラムと箱ひげ図の並列表示

があることを読み取ることができる．正確な数値は与えられていなくても，大雑把な比較が可能である．

棒グラフと同様なグラフ表示としてヒストグラムの例を図 5 に示す．本例は，ある授業の成績の頻度（人数）分布を示している．また，ヒストグラムの下には箱ひげ図も併せて表示している．値の大小を棒の長さで表現するという意味ではヒストグラムは棒グラフの一種と考えることができる．ヒストグラムは値の分布を表すための棒グラフである点が一般の棒グラフとの違いである．ヒストグラムは連続した値を範囲によって分割して表示するという性格を持つため，通常は棒と隣の棒は連結されて表示される．

ヒストグラムはこのような目的で用いられるため，通常の棒グラフのように値の比較の側面と値の変化側面との両方を同時に，かつ直観的に把握できるところに特徴がある．本例の場合，70 点台に値のピーク（モード）があり，その次が 50 点台である．ヒストグラムでは，全体として成績がどのように分布しているかも重要である．

たとえば，60 点台の値が低いことはさほど重要ではなく，全体として，50，60 点台の人数も全体としてはピークである 70 台点に向かって全体としては増加に向かっていることを示していると読み取ることができる．このような読み方をすることは，値を比較することに重点を置いた通常の棒グラフとの大きな違いである．

ヒストグラムから読み取れる値の増減に関する全体の傾向パターンは箱ひげ図からもある程度読み取ることができる．箱ひげ図は，値の順に中央値（Median，Q2），中央値以下の値群の中央値（第 1 四分位値，Q1），中央値以上の値群の中央値（Q3) の 3 つの値に最小値と最大値を加えた 5 つの値で全体の値分布の傾向を把握する．本例の場合，中央値 Q2 は 70 点近くにあり，Q2 と Q1 の距離の違い（値の差）と Q2 と Q3 の距離の違いを比べると前者の方が 2 倍近くになることが図から読みとれる．値の定義より，それぞれの範囲に同数のサンプルが位置していることから，サンプル全体の中央の半数全体が値の大きい側に偏っていることを示している．これはヒストグラムから読み取れる分布の傾向と同様の結果である．

箱ひげ図のひげ部分に関しては箱部分の反対側が最小値と最大値を表現しており，これから値全体の範囲（Range）が分かる．本例では最低値は 20 点台にあり，最大値は 100 点満点であることが図から読みとれる．ひげの長さも，全体個数の $\frac{1}{4}$ のサンプルが存在する範囲を示しており，長さに反比例した頻度で，その範囲内にデータが存在することを示している．上位の値に対応する右側のひげの長さは箱部分の長さ（幅）と同等か幾分長い．これは頻度としては半分程度の分布になっていることを示している．一方，下位の値に対応するひげは，数十％程度長くなっており，頻度の平均値が上位ひげと比べてさらに低いことを示している．このように箱ひげ図表示に用いられる 5 つの値だけでも，頻度の分布に関して相当の情報が得られる．

一方，ヒストグラムによると 20〜30 点台の頻度が少ないのみならず，40 点台の件数は 0 件である．すなわち，20〜30 点台はいわば外れ値となっているものと解釈することもできる．これらの値は，箱ひげ図で用いられる外れ値の定義には合致していないものの，ヒストグラムからは，より細かなデータの傾向が読み取れる．このように，データ全体の頻度の変化の概要は箱ひげ図だけでかなりの程度把握することが可能である．ヒストグラムでは，このような全体的な分布概要に加え，より細かい分析を行うための情報が得られる．

3.2　折れ線グラフ

折れ線グラフは時間経過による変化状況を表現することを目的に用いられることが多い．折れ線グラフのひな形を図 6 に示す．

我々はまず折れ線グラフの構成要素である線分とその端点（小さな○印）に注意を向ける．折れ線全体は，4 つの線分と 5 つの端点から構成されている．通常，左から右方向に配置されるため，視点は図左下の端点から出発する．その端点から始まる線分を順に追っていくと，まず値が増加（右上がり）し，次に減少し，大きく増加し，最後に少しだけ増加して折れ線が終了する．さらに詳しく見ると，第 1 の線分と第 3 の線分の傾きを比較すると，第 3 の線分の方が傾きが急であることが読み取れる．これは値の上昇が大きいことを意味する．また，最後（4 番目）の線分の傾き（値の上昇の程度）は，第 1，第 3 の線分よりもなだらかであることがひと目で読み取れる．

このように折れ線グラフとして値の変化を視覚化することにより，値の変化の全体的傾向が直観的にひと目で把握できる．これが折れ線グラフによる可視化の特性である．

図 7 に折れ線グラフの実例を示す [12]．本グラフは図書館の貸出記録データから 11 名の学生の貸出図書の傾向を学部全体の貸出傾向とのコサイン類似度を表示したものである．学部の配置に関しては，図書の貸出からみて文系学部の代表である LA（法学部）を左端に，理系学部の代表である SC（理学部）を右端に置いた．また，学部全体は，文系学部→文系・理系の色分けの少ない学部→理系学部という順序で配置されている．

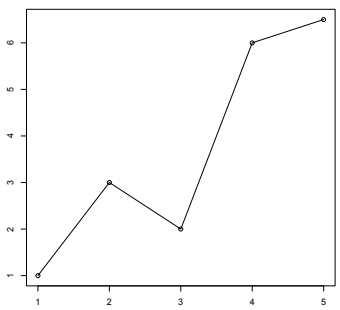

図 6: 折れ線グラフの特性

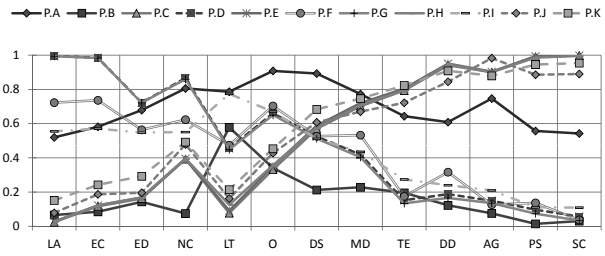

図 7: 折れ線グラフの実例 [12]

P.A と表示された学生（ひし形◇，実線で表示）の場合，左端の LA（法学部）から中央近くの O（その他）に向かって全体的に類似度が増加している．その後，途中 AG（農学部）で値が増加しているものの，全体としては右端の SC（理学部）に向かって徐々に類似度が減少している．すなわち，P.A の学部との類似度グラフは全体としては山形になっており，P.A は文系・理系という分類では，両方から離れた LT（文学部）や O に近い読書傾向を示していると結論づけられる．

P.B（□印の実線）は LT で最大の類似度になっており，全体的には P.A と同様の山形になっている．P.A との大きな違いは，両端との類似度が非常に小さく，LT とその周辺に偏っているところである．本学生は P.A 以上に LT に近い傾向を示している．

一方，P.K（□印の破線）は P.A や P.B とは異なり LT との類似性が小さい．全体としては LA から SC に向かって類似度が増加するパターンである．これは典型的な理系の傾向を示している．実際，この学生は理学部所属である．

P.K の逆パターンを示しているのが P.F（○印の実線）である．すなわち，LA や EC（経済学部）との類似性が高く，多少上下しながら，全体としては SC に向かって類似度が減少していっている．途中の O から MD（医学部）ではある程度大きな値を示したりしており，全体的には文系の傾向ではあるものの，理系・文系に囚われないである程度広い分野の読書をしていることが窺える．

本例の場合，通常は棒グラフにより学部毎の比較を行うのが通例であるが，数個以上の系列や項目について 1 つのグラフで同時に傾向を把握するためには，棒グラフは適していない．折れ線グラ

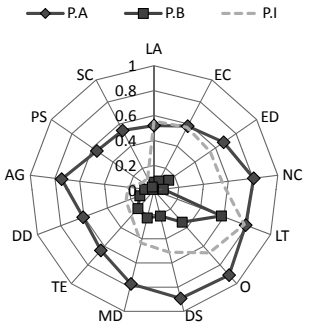

図 8: レーダーチャートの実例 [12]

フは実線や破線などの細い線分によって値の関連を表現するため，このような込み入った状況においては，棒グラフよりも適している．

図 8 にレーダーチャートの実例を示す [12]．データは折れ線グラフの例と同様に，利用者の学生と学部との貸出図書の類似度である．定義が異なるため厳密には同一ではない．通常レーダーチャートは棒グラフや折れ線グラフとは別のタイプであると見なされているが，折れ線グラフの水平軸（x 軸）方向を角度により表現したものと見なすこともできる．値の変化が円状に表されているため，我々は変化のパターンを 1 つの閉じた図形として把握できる．その結果，折れ線グラフにおける値の上下をそのまま表現された「折れ線」パターンよりも，パターンの違いがより強調される．

P.A の線を見ると，全体として円に近い形をしている．これは学部との類似度に大きな変化がないことを示している．しかし，子細に見ると，O や DS（芸術工学部）との類似度が大きく，LA や SC との類似度は小さいことが分かる．いわゆる理系の中では AG（農学部）との類似度が高い．

P.B の図形は P.A と比較してサイズがかなり小さい．これは全体として学部との類似度が小さいことを示している．その中で，LT（文学部）との類似度だけが突出して高値となっている．これは P.B が典型的な文学部学生であり，それ以外の分野の読書をほとんどしないことを示している．言い換えると，自分の専門分野以外には興味を持たないタイプの学生であることになる．

P.I は P.A と P.B の間に位置する．P.B と同様に LT（文学部）との類似度が最も大きく，全体として LT タイプの学生であると判断できる．しかし，P.B と比べると他の学部でも類似度が高い学部が存在する．たとえば，EC（経済学部）と LT との間に存在する学部との類似度は全体的に高い．それに対して SC から MD の間に位置する学部との類似度は押し並べて小さい．すなわち，P.I は基本的には文学部タイプであるものの，文系分野には広く興味を持っている一方，理系分野に関する興味が低いことがわかる．

このように，レーダーチャートは閉じた折れ線の図形によって，図 7 の折れ線グラフと同様の結果をより見やすく表現できる．

3.3　円グラフ

円グラフは全体に対する割合を表示するのに適したグラフである．円グラフでは割合の大小が中心から放射状に伸びた線分の成す角度の大小により表現される．レーダーチャートでも述べたように，我々は長さの大小よりも角度の大小に敏感に反応する．円グラフはこの能力を利用している．

図 9 に円グラフの実例を示す [11]．本例は大学図書館の貸出データより，利用者のタイプ別の貸出件数割合を表示している．グラフより，B3（学部 3 年生）や M(修士学生）の割合が高く，それに次いで B4（学部 4 年生）や D(博士学生）の割合が高いことがひと目で見て取れる．このように全体の中での割合の度合いが直観的に読み取れることが円グラフの大きな特性である．

また，グラフ上部 B1 から右回りに見て B4，そして，図では見づらいが B6 に至る部分は学部学生の割合を示しており，学部学生全体で，貸出件数の半分（円中心から真下に当たる線）を幾分越え

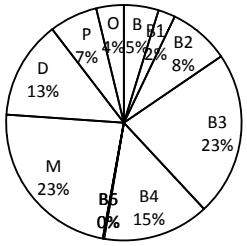

図 9: 円グラフの実例 [11]

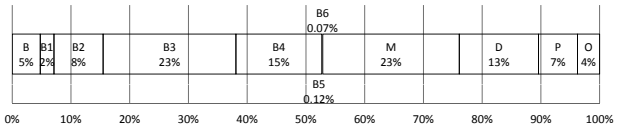

図 10: 100%積み上げ棒グラフの例 [12]

ていることが分かる．すなわち，おおよそ半分の貸出は学部学生によるものであることも容易に見て取れる．さらにその先を見ると，残り半分，すなわち円の左半分の $\frac{2}{3}$ ほどは M と D であり，残りが P（教員）や O（その他）であることも図より読み取ることができる．

図 10 は同じデータを 100%積み上げ棒グラフとして表現したものである [12]．100%積み上げグラフも円グラフと同様に，全体に対する割合の程度を表現するのに適したグラフである．

円グラフと同様に割合の大小を読み取ることができ，また，学部学生が全体の半分強の割合を占めていることも本グラフから容易に読み取ることができる．円グラフとの違いは，円グラフの方が割合の違いがより強いインパクトを持って直観的に読み取ることができるのに対し，100%積み上げグラフではより意識して注意を払わないと差の程度が見て取れないことにある．

この差異が生じるのは，円グラフの場合は，中心から放射線状に境界線が引かれており，中心に注目するだけで，その大小に関する印象情報が直ちに得られるのにたいして，100%積み上げグラフでは，視線を左右に移動させ，比較する必要があるからと理解できる．

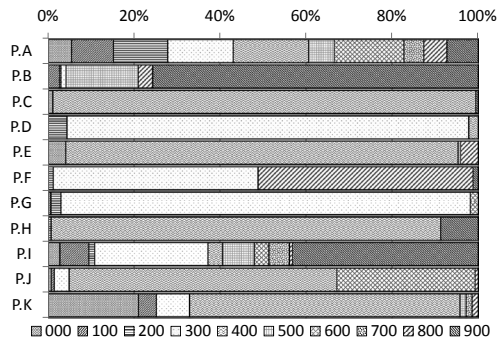

図 11: 100%積み上げ棒グラフの実例（複数グラフの並列表示）[12]

一方，100%積み上げグラフは全体の形状が棒状（長方形）であるため，複数のグラフを並べることにより比較できるグラフがコンパクトに作成できる．図 11 に例を示す [12]．本例は，11 名の学生について，貸出総数に対する日本十進分類（NDC）に関する割合（興味分野のプロファイル）を示している．このように並べて表示することにより，各学生の興味の違いが容易に見て取れる．同様の比較は円グラフでも可能ではあるものの，11 もの円グラフを表示し，比較することは，視覚的にはたやすくない．

3.4　散布図（相関）

属性と属性の間の相互関連に関する分析はデータの特性を発見する上で，重要な分析対象である．2 つの属性に関する線形関連の度合いは相関係数値によりある程度のメドをつけることができる．しかし，全体としてどのような相関関係にあるのかは，相関係数だけでは不明である．たとえば，相関係数＝ 0 という結果であったとしても，両者に関連がないということではなく，線形ではない関連がある場合が存在する．したがって，属性間の相互関連を的確に把握するためには散布図など他の手段と併用する必要がある．

図 12 に散布図の実例を示す [8]．本例は，大学図書館の貸出データのうち，NDC400 番台（自然科学）の 10 のサブカテゴリ毎に，図書館登録から時間が経つにつれての貸出冊数の減少指数を横軸に，貸出の中の古い図書の割合を縦軸に，両者の関連を散布図として表したものである．全体的に負の相関がある．これは時間に伴う貸出価値の減

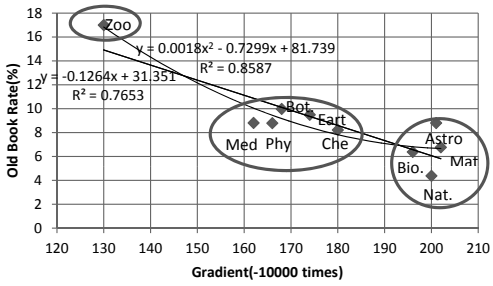

図 12: 散布図の実例 [8]

少が大きい（x 軸の右方向）分野では古い図書貸出の割合が減少する（y 軸の下方向）ことを示しており，自然な結果である．

本グラフで重要な点は自然科学の中で NDC が 400, 410, . . ., 490 の 10 分野が大きく 3 つの分野にグループ化できることである．特に，グラフ左上に位置する Zoo（動物学）は他の分野と比べ，ひとり減少率が低値，すなわち古い文献が比較的多く借りられ続けていることを示しており，興味深い．

また，素人判断では，知識の移り変わりが激しく，グラフの右端に位置することが予想される Phy（物理学）や Med（医学）が本グラフでは中央の主グループ内の，左端に位置していることも興味深い．動物学と同様，このような結果になった理由を追究するためには，さらなるデータの分析や，新たなデータの入手が必要である．このように新たな結果の発見や，それに伴う新たな行動の必要性を発見することもデータ解析やそのための可視化の重要な役割である．

図 13 に別の散布図例を示す [11]．本例は，大学図書館の貸出記録データを基に貸し出された図書の専門度なる概念を導入し，それぞれの利用者に関する借りた図書の専門度のレンジ（最大値から最小値を引いた値）を縦軸に，その利用者の専門度（借りた図書の専門度の平均値）を横軸に，相互関連を散布図として表したものであ

横軸のメモリは 1 から 6 はそれぞれ学部の 1 年生から 6 年生に対応した専門度であり，8, 9 はそれぞれ修士課程学生，博士課程学生のレベルに，そして，最大値の 10 は教員の専門度レベルとして設定してある．また，図書の専門度も同様に学部 1 年生レベルに相当する専門度 1 から教員レベルの

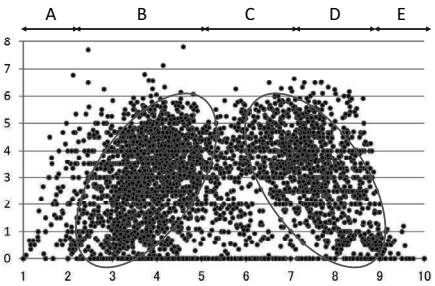

図 13: 散布図の例（全体分布）[11]

10 までの値で設定されており，その差であるレンジは最小値が 0，最大値が 9 となる．

横軸 (x 軸）の最小値である 1 と最大値である 10 の部分は，それぞれ 1 年生だけが借りる図書のみを借りた利用者，教員のみが借りる図書を借りた利用者に対応するため，レンジは 0 となっている．図より，全体を 5 つに分けることができる．左から，専門度 1 から 2 までの区間，専門度 2 から 4.5，もしくは 5 までの区間，4.5，もしくは 5 から 7.5 の区間，7.5 から 9 までの区間，9 以上の区間である．それぞれを A から E 区間と呼ぶことにする．

A 区間と E 区間は先に述べたように，自分の専門度レベルに見合った図書のみを借りた利用者に対応する．レンジレベルは他の区間と比べて自ずと小さい．

B 区間は学部の 3，4 年生を中心とする学生達のグループである．レンジ 0 の専門度幅の小さな学生からレンジが 7, 8 に至る様々な専門度レベルの図書にチャレンジする学習意欲旺盛な学生に至る幅広い学生群を含んでいる．D 区間は院生を主力とするグループに対応する．B 区間と同様に，多様な学生が存在していることが見て取れる．

C 区間は隣の B, D 区間とは異なる様相を示している．この区間は学部の 5, 6 年生から直接学生のレベルには対応していない専門度 7 を含む区間である．すなわち，この区間に属する利用者のほとんどは，学部向けの図書と同時に院生などを対象とする研究者向け文献なども借りるような学部生や逆に専門書だけではなく学部向けの図書も多く借りる院生が占めているものと考えられる．

C 区間の特徴は，隣の B, D 区間と比較してレンジが小さい利用者の割合が小さく，レンジ値 4 近辺に集中していることである．たとえば学部 4 年生が多く読む図書（専門度 4）と修士学生が多く読む図書（専門度 8）をもっぱら読むような学生はレンジが 4 となり，本区間に位置する利用者の 1 つの典型例である．

本例の場合も，このような考察により大学図書館の利用者，とくに学生の動向を掴む上で散布図は極めて有効であることが分かる．このような分析を通じて得られた知見に基づき，さらなる研究の方向性が見えてくることは，前例と同様である．

4　まとめと課題

近年，ビッグデータ解析やビジネスインテリジェンス（BI) などの概念が注目されており，マスコミなどでもたびたび取り上げられている．これらは，いずれも現在大量に取得することが可能となったデータをデータマイニングなどの技術を活用し，それをビジネスや社会に活かすことが目的である．

このようなデータ活用には，従来の手法である統計解析による仮説検証的なデータ解析（Confirmatory Data Analysis, CDA）とは異なる探索的データ解析（Exploratory Data Analysis, EDA）が有効である [9]．CDA においても可視化は有効なツールではあるものの，EDA は仮説の存在を前提としないため，まず与えられたデータがどういうものであるのかを知ることから解析を始める必要があり，そのためのツールとして可視化技術は欠かせない．

我々も大学図書館の貸出記録データや授業の振り返り評価アンケートなどのデータ解析を行って来ており，その研究過程において棒グラフ，折れ線グラフ，散布図などのグラフを多用してきた [3][4][5]．その中で，そもそも我々はグラフをどのように読み取り，そこから何がしかの特徴を発見しているのか，その原理に興味を持ったことが本研究の出発点である．

様々な場面において広く用いられている可視化技術であるが，文献や図書，Web サイトでは，どのように工夫すると，より効果的な可視化ができるか，あるいは，どういうツールがある，などの情報は満ち溢れているものの，なぜそうなのかを原理的に遡って解明しようというものは少ない．

このような現状認識を受け，本稿では，可視化図形，特にグラフ類に注目し，それらがどのように我々に認識され，可視化ツールとしての性格に結びついているのかを，基礎的レベルから分析した．本稿の分析を通して明らかになったことは，我々は流れの認識を中心に図形を認識していること，流れの認識力を利用して，図形全体のおおまかな形状の把握しており，また変化の認識力を利用して，角などを認識していることである．

さらに，長さの違いや線分の角度の違いなどを比率的にも認識しており，可視化技術により，比較による特性の抽出が容易になっているということも明らかになった．

本稿では可視化の原理に関する第一歩として可視化特性の現象面からの理解に重点を置いた．我々の最終的な目標は，仕組みを理解する科学的アプローチに留まらず，可視化の仕組みを自動化する，すなわち，計算機のアルゴリズムとして記述するという工学的なアプローチに繋げることである．可視化をアルゴリズム化できると，実際の可視化を行わず，可視化に必要なデータから自動的に特徴発見を行い，それに基づき，次の解析を行う EDA 過程を，少なくとも一部は，自動化できることになり，我々のデータ解析力を飛躍的に向上させることができる．本稿の研究は，この将来目標に向かっての小さな，しかし重要な，第一歩である．

このような認識の下，今後の課題として次のテーマがある：(1) 可視化に関する分析の際の指標となるヒトの目の特性や注視点の移動などに関するさらなる追究を行うこと，(2) 関係グラフなど他の可視化手法に関する分析を行うこと，(3) 形状だけではなく色や動きなど，我々の視覚の認知特性を活用した可視化手法を新たに開発すること，(4) 可視化という視覚以外の五感を利用した「感じる化」に関しても研究対象を広げること，(5) 可視化を用いたデータ解析，特に EDA の試行錯誤過程を記述できる枠組みを構築し，それに基づき，我々が行ってきた研究過程を自動化，あるいは半自動化することなどである．このような自動化を推進することにより 4 次元の可視化など人間の能力を超えた可視化も実現できる可能性がある．

以上見てきたように，可視化という現象は奥の深い研究対象であり，その解明には長い期間と多大なる労力が必要である．しかし，その潜在的な可能性の高さを考えると，今後ますます発展させるべき研究分野である．

謝辞

本研究の出発点となったデータ解析研究の共同研究者である九州情報大学の大浦洋子教授，及び富士通研究所（前九州大学附属図書館研究開発室准教授）の馬場謙介研究員に深く感謝いたします．

参考文献

[1] 飯田英明: ひと目で伝わる！図解表現のテクニック，PHP ビジネス新書，PHP 研究所. (2012)

[2] CiNii Articles, 日本の論文をさがす，国立情報学研究所. http://ci.nii.ac.jp/ [2016/2/11]

[3] 南俊朗：図書館データマイニングのすすめ－図書館マーケティングの可能性を広げるために，現代の図書館，Vol.51, No.3，日本図書館協会，pp.172–179. (2013)

[4] 南俊朗：図書館マーケティングのためのデータ・アナリシス－データに耳を傾けてサービスの改善を目指す－，情報の科学と技術，Vol.65, No.10，特集：データ分析によるサービス改善，pp.412-417. (2015)

[5] 南俊朗： 図書館サービスのためのデータ解析の新展開－新しいパーソナルサービスの可能性－，九州大学附属図書館研究開発室年報 2014/2015, pp.11-18. (2015)

[6] 森藤大地，あんちべ: エンジニアのためのデータ可視化 [実践] 入門，技術評論社. (2014)

[7] 湯之上隆:　日経エレクトロニクスの記事「ムーアの法則，黄昏の時」に意義あり！今後当分微細化が止まる兆候はない，Yahoo Japan!　ニュース. (2015) http://bylines.news.yahoo.co.jp/yunogamitakashi/20150727-00047852/ [2016/2/11]

[8] Kensuke Baba, Toshiro Minami, and Sachio Hirokawa: Should University Library Collect New Books or Old Books? – An Obsolescence Analysis for Book Selection –, The International Symposium on Advanced and Applied Convergence (ISAAC 2014), in J.J.Kang et al.(eds) ISAAC 2014 & ICACT 2014, AACL 03, pp.34-37. (2014)

[9] John T. Behrens: Principles and Procedures of Exploratory Data Analysis, Psychological Methods, Vol. 2, No. 2, pp.131–160. (1997)

[10] Enrico Bertini, and Denis Lalanne: Surveying the Complementary Role of Automatic Data Analysis and Visualization in Knowledge Discovery, Proc. ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration (VAKD '09), pp.12-20. (2009)

[11] Toshiro Minami and Kensuke Baba: Investigation of Interest Range and Earnestness of Library Patrons from Circulation Records, International Conference on e-Services and Knowledge Management (ESKM 2012), as a part of the 1st IIAI International Conference on Advanced Applied Informatics (IIAI-AAI 2012), IEEE CPS, DOI 10.1109/IIAI-AAI2012.15, pp.25-29. (2012)

[12] Toshiro Minami: An Analysis of Interest Area Similarities by Utilizing the Loan Records of Library, IADIS International Journal on Computer Science and Information Systems (IJCSIS), Vol. 8, No. 1, pp. 112-129. (2013)

[13] Helen C. Purchase, Natalia Andrienko, T.J. Jankun-Kelly, and Matthew Ward: Theoretical Foundations of Information Visualization, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North (Eds.), Information Visualization Human-Centered Issues and Perspectives, LNCS 4950, pp.49-64. (2008)

[14] Bernice Rogowitz: Theory of Visualization Panel, IEEE Visualization, (2010)

[15] Ben Shneiderman: Inventing Discovery Tools: Combining Information Visualization with Data Mining, Information Visualization, 1, pp.5-12. (2002)

[16] Ian Spence: William Playfair and the Psychology of Graphs, Proc. American Statistical Association, Section on Statistical Graphics, Alexandria VA: American Statistical Association, pp.2426–2439. (2006)