

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

国連欧州経済委員会

Making Data Meaningful

Part2: A guide to presenting statistics

データを有意義なものとするために

パート 2: 統計を表現するための手引



国際連合

ジュネーブ, 2009

※ 総務省統計局(統計情報システム課)仮訳 1.0 版

注 記

この文書中の称号や表記は、いかなる国、領土、都市、地域又はそれら領域における当局の法的立場、あるいは境界の画定に関して、国連事務局の見解その他を表すものではない。

ECE/CES/STAT/NONE/2009/3

目次



序文	iii
1. メッセージを理解させる.....	1
2. 統計の視覚化	7
3. 統計表	12
4. グラフ	18
5. 地図	31
6. 新たな視覚化テクニック	42
7. アクセシビリティの問題.....	47
8. 参照・参考文献.....	52

序文

この手引『データを有意義なものとするために』(Making Data Meaningful)は、ヨーロッパ統計家会議¹の事業計画の下、統計のコミュニケーション及び普及に関する国連欧州経済委員会(UNECE)作業部会²の枠組みの中で作成された。

この手引は、統計組織、特に啓発戦略を開発中の統計組織の幹部、統計家及び広報担当者を助ける実用的ツールとなることを意図している。この手引は、統計家ではない人々が統計を活かせるようにするための文章、表、グラフ、地図及びその他の工夫の使い方に関するアドバイスを提供する。この手引は提案、手引ライン及び実例を含むが、厳格なルールあるいはひな形を示すものではない。

効果的なデータの公表は、文章、表及びグラフを組み合わせて使い、様々なタイプの情報を伝達する際にその強みを最大限に発揮している。「データを有意義なものとするために Part1: 数値についてのストーリーを書くための手引(2006年発行)」では、効果的なライティング技術の用法に焦点を合わせた。「データを有意義なものとするために Part2: 統計を表現するための手引」が目指しているのは、あまたのコミュニケーション手法の中から現在利用可能なツールや技術の最適な組合せを使って専門家でない人々にメッセージを伝える最良の方法を読者が見つけ出すよう手助けすることである。

統計組織によって多くの実務的・文化的な違いがあり、そのアプローチは国によって様々であろうことをこの手引は認識している。

統計のコミュニケーション及び普及に関する専門家グループがこの手引を作成した。

専門家グループのメンバー(アルファベット順)は以下のとおりである。

- Petteri Baer(国連欧州経済委員会)
- Colleen Blessing(米国エネルギー情報局)
- Eileen Capponi(経済協力開発機構)
- Jérôme Cukier(経済協力開発機構)
- Kerrie Duff(オーストラリア統計局)
- John Flanders(カナダ統計局)
- Colleen Flannery(米国センサス局)
- Jessica Gardner(国連欧州経済委員会)
- Martine Grenier(カナダ統計局)
- Armin Grossenbacher(スイス連邦政府統計局)
- David Marder(イギリス国家統計局)
- Kenneth Meyer(米国センサス局)
- Terri Mitton(経済協力開発機構)
- Eric St. John(カナダ統計局)
- Thomas Schulz(スイス連邦政府統計局)
- Anne-Christine Wanders(国連欧州経済委員会)

¹ 統計のコミュニケーション及び普及についての国連欧州経済委員会の作業部会に関する情報は同委員会のウェブサイトから利用可能である。

<http://www.unece.org/stats/archive/04.05.e.htm>

² ヨーロッパ統計家会議に関する情報は国連欧州経済委員会のウェブサイトから利用可能である。

<http://www.unece.org/stats/archive/act.00.e.htm>

Martin Lachance (カナダ統計局)、Andrew Mair (オーストラリア統計局)、Alan Smith (イギリス国家統計局)及びSteven Vale (国連欧州経済委員会)の貢献に大いに感謝する。

1. メッセージを理解させる

1.1 書き言葉

報道発表資料(News releases)は、統計組織が、意図する対象者—ほとんどの場合は一般の人々に、統計や分析の主要な発見(key findings)を伝達する手段である。文章は、発見を説明し、傾向を概観し、背景・状況に関する情報を提供するための主な手段である。

本章では、「効果的な」報道発表資料あるいは報告書や分析記事のような他の文書を作成するための多くの提案を示す。

報道発表資料、報告書あるいは分析記事を効果的にするものは何か。おそらく、最もよい説明は最初の手引「データを有意義なものとするために Part1」にある。それは、数値についてのストーリーを書くための手引であり、メッセージをうまく相手に伝えるための最初の提言集である。効果的な報道発表資料とは、

- データについてストーリーを語っていること、
- 一般の人々のニーズに適合し、読者がなぜこれを読みたいのかとの間に答えていること、
- 見出しやイメージで素早く読者の関心をとらえること、
- 容易に理解できて、興味深く、しばしば楽しめること、
- 伝えようとしていることにインパクトを加えるために統計を適切に使うよう、メディアを含めた他者を後押しすること。

文章作成にとりかかる手助けになるいくつかのコツがある。

1.2 対象とする読者: 最初に決定すべきこと

最初に行うべき重要な決定は、読者を正確に特定することである。誰に向けて書いているのか？簡単に言えば、読者は運転席に座っている。概して、読者の望むものがく報道発表資料を通じて与えられるべきある。読者に耳を傾け、その関心をとらえる正しい叙述、言葉、視覚に訴える表現方法を見つけ出し、選択しなければならない。

今日、インターネットのおかげで、選択すべき読者は多様化している。ほとんどの統計組織は、専門家ではなく一般の人々かなりの教育を受けてはいるが専門家ではない人々に情報を伝える責任を負っている。報道発表資料が印刷物であった時代には対象とする読者はもっぱらメディアであり、統計組織は主要な発見を大衆に伝えるのにメディアを頼っていた。

しかし、今日、統計組織は、ウェブサイト、電子メール、その他の形態のインターネットベース配信を通して、多数の直接的な読者層を開拓してきている。このことは、統計組織が同時に多数の読者とコミュニケーションしていることを意味する。一般の人々、データユーザー、銀行家、金融アナリスト、大学教授、学生など、それぞれにそれぞれのデータへの要求がある。

コミュニケーションの世界は、休むことなく進化している。成功した商業メディアはこのことを知っており、どのストーリーが最も注目を集めたかを絶え間なく—しばしばリアルタイムで—注視している。そのようなメディアは、それから、各読者との対話をより深めようと、ビデオ、追加の写真あるいはより多くの分析といった手段を使って一層豊富なコンテンツを作成することにリソースを振り向ける。

いずれにせよ、ここで述べたいことは、貴重なリソースを何らかのコミュニケーション手法—新規のものであれ、既に確立されているものであれ—に投ずる前に、まず、読者や利害関係者は誰か、統計組織に何をどのように求めているのかを見極めることである。

もし、統計組織が複数の読者に話しかけたいのであれば、それぞれに応じた最も適切なコミュニケーション手段・技術を用いて、メッセージを伝えなければならない。

しかし、時間と資源が足りないということは、すべての読者に常に情報を届けることが不可能であることを意味することが多い。統計組織には選択の余地がある。＜情報を伝えようとする読者に＞優先順位をつけるか、できるだけ広い読者に伝えたいのならくそのような読者の間に＞最も明確な共通の基盤を見出すこともできる。

これは、多くの統計組織が行っていることである。こうした統計組織は一般の人々を対象としているが、報道機関を「パイプ」のように使って、その読者に情報が伝わるように協調している。対象とする読者は一般の人々であるが、報道機関はそれらの人々とコミュニケーションするための手段である。このアプローチは専門家のためにもなる。広範な読者に情報を届けるための簡潔で分かりやすい技術は、最も専門的な読者にさえも温かく歓迎されることが多い。

1.3 コミュニケーションを行う背景を理解する

統計のコミュニケーションは単独で起きるものではない。それゆえ、統計組織がコミュニケーションを行う背景を理解することが重要である。読者がメディアを利用する方法は絶えず変化している。また、技術的な能力及び統計の理解力に世代間の明らかな違いもある。

統計のコミュニケーションを立案するときは、オンラインメディアの利用に関する四つの顕著な傾向に留意すべきである。それらの傾向は、好機であると同時にリスクでもある。

1. ウェブは、ますます娯楽のためのメディアとなっている。興味深い方法で表現されないどんなメッセージも、若い読者とかみ合わないリスクがある。
2. 社会は、情報の利用に関して「スナック文化」を作り上げた。読者は、すばやく利用できる、よりコンパクトな情報の断片をますます求めるようになっている。
3. インターネットを利用する読者は、必要最低限のものでよしとする傾向がある。そのような読者は、何となく関連性がある情報の断片を見つけると、それと関連の深い情報をさらに探すことなく、そこで止めてしまう。
4. 様々な読者や表現スタイルを扱うことにより、統計の情報発信をより楽しくあるいは利用しやすくする重要な読者を排除しないようにしよう。

さて、インターネットを最大限に利用するために、統計組織には何ができるだろうか。統計組織のメッセージを理解してもらうため、最も適切な言語・構成・表現を用いなければならない。以下の節でその方法を解説する。

1.4 叙述: ストーリーを語る

何よりもまず、ストーリーを見つける

一般の人々にとってデータが有意義であるためには、数字の中に意味を見出すことが重要である。「ストーリー」という言葉は、統計や科学の世界ではしばしば人々に警戒感を与える。それはデータの間違った解釈につながりかねない虚構ないし装飾という含みを持っているからである。もしも分析者が注意深く、丁寧にデータにアプローチしないとすれば、この見解は正当化されるかもしれない。

しかし、その代わりの手段、例えばストーリーを避けることは、はるかに悪いことであろう。人々はしばしば統計に不信感を抱き、統計が誤解を招いていると感じているが、これは人々がデータを理解できないためである。データを作成する者がデータを適切なものとし、人々が理解できるようにデータを説明することができていないためにこのようなことが起きる。ストーリーの筋がなければ、報道発表資料は単なる数字の記述になってしまう。

統計のストーリーは、研究中のデータと現象に関する十分な知識に基づいていなければならない。さもなければ、そのようなストーリーはおもしろいが、実際はすべて間違いということにもなり得る。統計のストーリーを準備する際は「公的統計の基本原則」³も思い起こさなければならない。

1. 公平性
2. 専門性
3. メタデータ
4. 誤った解釈に対する意見
5. 多様なデータ源
6. 秘密保護
7. 透明性
8. 国内調整
9. 国際基準
10. 国際協力

統計組織が公正を保ち、回答者及び小さな部分集団の秘密を保証することが不可欠である。

文章は、短期・長期の傾向の中で最も重要で有意義な発見を扱うべきである。それは、証拠によって裏付け得る範囲で、関係、原因及び効果を探るものでなければならない。そして最新情報の意義を読者に示すべきである。

ジャーナリストのスタイルで書く

ジャーナリストが使っている「逆ピラミッド型」の文章スタイルを使おう。最も重要な事実を最初に示し、重要度が落ちる順に補助的なポイントを続ける。読者は急速に興味を失うので、最も重要な情報は文章の最初になければならない。

方法論で文章をスタートさせ、結論で終わるということは避けよう。方法論に関する要点は注記—解説は簡潔なほどよい—に記述すればよい。結論をリード部分<柱書き>あるいは冒頭の段落に置くべきである。

リード部分は、文章の最も重要な要素であり、データに関するストーリーを語るべきである。リード部分は、ストーリーの筋を簡潔、明瞭、単純に要約し、そのストーリーを文脈の中に置く。リード部分は、一つのメッセージ又はテーマに集中し、最小限のデータのみを含むべきである。

リードパラグラフの良い例

オランダの非金融会社の純益は、2008年の第2四半期に190億ユーロとなった。これは3年間で最も低い水準である。利益は2007年第2四半期より11パーセント低かった。純益の低下は二つの主な要因によってもたらされた：高い利息コスト—会社はより多くの純利息を支払った—と、海外子会社の低調な利益である。

出所：オランダ統計局

³ これらの原則は1994年に国連統計委員会によって採択された。UNECEのウェブサイト <http://www.unec.org/stats/archive/docs.fp.e.htm> で詳細に説明されている。

文章の中に数字を書き込み過ぎて読者に負担をかけないように、かつ、数字は主要なものにとどめ、丸めた形にして使おう。重要度の低い数字は付属表に任せよう。文章は分析、傾向及び前後関係を述べるために使い、表の中の値を繰り返さないようにしよう。

構造に注意を払う

文章を構造化し、各部分がそれぞれ自身において筋が通っており、語ろうとしている全体のストーリーに合う構成にしよう。小見出しは、報道発表資料の構成を強化するための効果的なツールである。小見出しは、報道発表資料を扱いやすく意味のあるかたまりに小分けにする。

簡潔な小見出しは、各サブセクションの主要な発見を要約する。小見出しが動詞を含んでいると、より人を引きつけ、理解しやすいものとなる。

小見出しの良い例

「在庫水準わずかに緩和」
「エネルギー製品の成長が輸入増加を牽引」

インターネットによる情報伝達の場合、各サブセクションはそれぞれ意味を持つべきである。つまり、用語は文字を省略しないように正しく書き、資料出所を明記しなければならない。検索エンジンは、訪問者を望ましい目的地へと導くために統計組織が設けたホームページやその他の入口よりも、ウェブサイト内にある深いリンクへとユーザーを導いてしまう傾向がある。

読者の様々な情報ニーズに対応できるように、メッセージが積み重ねられるべきである。各サブセクションは、そこでの主要な発見を明確に述べるピック文で始めよう。後に続く段落でより詳しくこの発見を述べることができる。

この構造は、詳細に興味を抱く読者に対して、より詳しいデータ、分析及び技術情報への確かな道筋を示す。電子出版では、読み手をさらに複雑な分析に導くため、いつでもリンクを使用することができる。

1.5 言葉:明確、簡潔、単純に

単純な言葉は、成功を収める情報伝達における核心である。「短く、単純に」(keep it short and simple:KISS)という古い格言は、これまでと同じように今も真実である。これは文章だけに適用されるものではなく、必要以上に手を加えたいくなる表・視覚表現・図案を気の利いたものにするためにも当てはまる。

平易な言葉を求めて懸命に努力しよう。非効率でしばしば不親切あるいは不明瞭な「お役所的」文体で書かれた文書を容認してしまうことがあまりに多い。

平易な言葉を用いることは横柄でもなければ、内容を軽視するものでも、単純化しすぎることでもない。ましてや、文法を踏み外すことを意味するものでもない。平易な言葉は、明瞭で簡潔なメッセージを伝達する。それは、読者を念頭に置いて、正しいトーンで用いられる。

平易な言葉はより速く読むことができ、メッセージをより頻繁・簡単に、親切な方法で理解させることができる。

覚えておこう: 明瞭で単純なメッセージは「度を超えてやさしく書き直す」とことと同じではない。

明瞭に書くためのコツ:

- 短い文を用いる。
- 一つの文に一つのアイデアを目指す。
- 長い文を分割する。
- 各段落は最も重要なメッセージで始める。
- 段落は短く保つ。
- 歯切れのよい書き方を保つ。

動詞の場合、受動態を避け、能動態を用いよう。受動態はわかりにくく、冗漫で、かつ、直接的ではない。

悪い例:

「失業率の上昇は、経済危機によってもたらされた。」

良い例:

「経済危機が失業率の上昇をもたらした。」

可能な限りどこでも、専門用語は避けよう。専門用語は、特定のグループの人々に特別な意味を持つ言葉であり、幅広い読者にとって、ほとんど常にふさわしくない。メッセージに適し、読者が理解する最も単純な言葉を使おう。しかし、読者すべてがある特定分野の専門家で構成されているときは、専門用語が最も適切な使用言語であるかもしれない。

1.6 影響力を評価する:メディアの分析

読者は、データを誤りなく、知的に解釈したか。読者は、その分析を利用して、経済、環境及び社会について理に適った判断や政策立案を行うことができたか。

統計組織によるデータの公表によって読者が何をしたのかは、分析しなければ分からない。メディアと他の読者の両方がどのように情報を利用しているのか、統計組織がモニターしておくことはよいことである。その方法には次のようなものがある。

- 「顧客の声プログラム」又は顧客のフィードバックを収集する他のプロセスを設ける。
- ユーザーグループ及び的を絞った主要な利害関係者のグループを通して、一般の人々の反応を測定する。
- ユーザー調査を実施する、あるいは影響力のある顧客に一对一で意見を聞く。
- 微妙な論点又は接触が難しい読者について深い理解が必要なら、マーケットリサーチ・サービスを用いる。
- どのデータ及びストーリーが読者から最も需要があるのか判断する手助けになるよう、インターネットのウェブサイトへのアクセス量をモニターする。
- 読者がどんな情報を探しているのかを理解するための、単純であるが強力な技術として、検索キーワードのパターン及び用語を分析する。

メッセージが正確かつ公正に報道されることを確保するため、統計組織はメディアに直接働きかけることができる。メディアは、統計のメッセージを幅広い読者に伝達する最も簡単、安価かつ効率的な方法である。

残念ながら、いくつかのメディアは独自の基本方針(agenda)を持っていて、それがデータを誤用・誤解する結果となることがある。統計組織のメッセージがメディアを通して届けられたことと、あわせて、報道の論調と質をモニターすることも不可欠である。明らかな誤用の場合には、迅速に、一貫した態度で、断固として対応すべきである。

さて、文章の主要な要素をカバーしたので、次は統計組織がデータを説明する際の手助けとなる視覚的な要素に注意を向けよう。

2. 統計の視覚化

2.1 「百聞は一見に如かず」の理由

誰もが「百聞は一見に如かず」という古い格言を聞いたことがある。データを理解するための最良のテクニックの一つは、多くの数字を一枚の絵として視覚化することである。そうすることでパターンをずっと見やすくしたり、そうしなければ隠れてしまったであろうパターンを表示したりすることができる。

簡単な棒グラフからより複雑な散布図、主題図<(thematic maps) 訳注:あるテーマに関するデータを示すための地図。5.4で「thematic (statistical) maps」としているとおりの、この手引では「統計地図」とほぼ同義>、アニメーションで形の変わっていく人口ピラミッドまで、多くの異なった方法でデータを視覚化できる。データの視覚化に関する書物があり、この話題を専門にする多くのウェブサイトがあるなど、技術的サポートにも不自由はない。多岐にわたるソフトウェアとダウンロード可能なプログラムが入手できるようになっており、あらゆる目的に対応できる。

本章では、データ視覚化の歴史を簡単に眺め、優れたデータ視覚化のためのガイドラインを提供する。

2.2 視覚化は統計業務プロセスに不可欠な部分

統計組織は、データを効果的に表現することを統計作成プロセスに不可欠な部分にすべきである。データの視覚化は、統計的な傾向と関連性を伝える大変重要な部分であるから、後知恵ではなく、継続した取組でなければならない。UNECE、ユーロスタット(欧州統計局)及びOECDによる統計メタデータに関する運営グループでは、包括的統計業務プロセスモデル⁴を開発しており、視覚化はその中の「普及」局面に含まれている。

読者は視覚的なプレゼンテーションを容易に理解できる。新聞、テレビ、インターネット、書籍を通して、グラフと地図はほとんどすべての人々に影響を与える。数字の並んだ長いリストより、グラフ又は地図として示された統計を理解することははるかに容易である — もちろん、視覚化された表現が正しく作られているとすれば。

表現は、時間をかけずかつ簡単に傾向と関連性を解説すべきである。そのような表現は、データベースから情報を引き出して読者の頭の中に入れる効率的な方法である。

しかし、注意しよう。統計情報のつたない視覚化は誤解を招くものになり得る。わざとであろうと、多くの場合そうであるように意図したものではないにしても、誤解を招く情報を提供する多くの方法がある。デザインと機能の間でバランスが保たれなければならない。込み入った視覚化はしばしば情報を伝え損ねる。グラフを解釈することは骨が折れるから、読者にメッセージの「発掘」を強いることのないようにすべきである。

様々な文化的な慣習から誤った理解や解釈が引き起こされることもある。例えば、色は、世界の様々な地域で異なる象徴的な意味を持っているであろう。

2.3 あるスコットランド人によるデータ視覚化への歴史的影響

統計グラフの歴史は、今日私たちが理解しているように、300年に達していない。古代ローマ人とアラブ人は数字を使うのは得意だったが、棒グラフや曲線を使って数字を視覚的に表現することはなかった。

⁴ <http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

フランスの百科事典編集者・哲学者、René Descartes (1596-1650)は、観察結果を視覚化するために、初めて直交座標系を用いた。しかし、今日よく知られている方法で統計データを初めてグラフで表現したのは、スコットランド出身の技術者であり経済学者のWilliam Playfair (1759-1823)であった。

Playfairは、19世紀のかわり目(1786年及び1801年)に世界貿易の発展に関する2冊の本を出版し、その中で自身が「グラフ法」(graphic method)と呼ぶ方法を用いた。彼は熱心にこの新しいアプローチの普及に務め、視覚的な表現の利点を以下のように唱えた。

1. グラフによる表現は、観察することが難しい込み入った関係を単純化するのに役立つ。

「これらのグラフの手助けにより、全体を構成する項目を詳しく検討することによる疲労や苦勞なしに、人々が情報を手に入れるようになることが期待される。」

2. 事業主、政治家及び政策決定者は、詳細を検討する時間がないため、統計情報の核心につながる視覚的な近道が必要とする。

「高い地位の人々や多忙な事業家は、一般的な概略にしか注意を払うことができない...」

3. 統計表、特に分量の多い統計表は、退屈で理解し難いものになることがある。グラフによる表現はユーザーを助ける。

「...これまで長く続けられてきた方法、数値の並んだ表を使ったら、記憶に刻み込むために終日を要するであろうほど多くの情報が<グラフを用いた表現なら>5分で得られるであろう。」

4. グラフは目を引きつける。Playfairは、<データの>関係と量を理解するための彼の方法の有用性を強調する際、この点を力説した。

5. グラフ法は、知覚を助けるだけでなく、<データの>構造と関係の理解を助けることで、知性を引き付け、意欲をかき立てる。

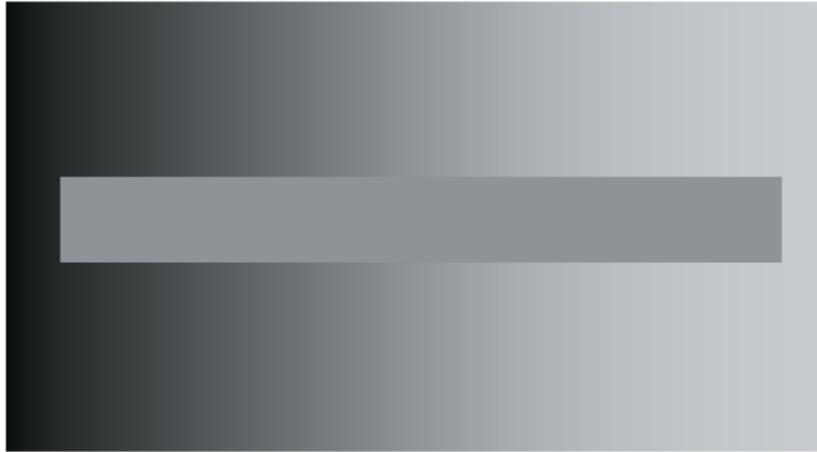
Playfairの時代から視覚化の方法に革命があったが、まだ精緻化と新たな発展の余地がある(第6章参照)。しかし、あまりに派手な、観察された内容をゆがめる表現を避けることが重要である。**表現技術は手段に過ぎないこと**<訳注：原文直訳「技術は単なる召使いであること」>**を忘れてはならない**。そうすることができるからといって、役に立たない注釈やあいまいな要素を加えるべきではない。読者のためにメッセージは単純にしておくべきである。

今日、グラフ表現と人間の知覚に関するより良い理解を促進している新しい名人の中に、Jacques Bertin, William Cleveland, Jan-Erik Kristiansen, Vesa Kuusela, Hans Rosling, Edward Tufte及びHoward Wainerがいる。こうした最近の専門家の業績はあるものの、統計グラフの基礎はWilliam Playfairによる革新的創造の上に作られ続けている。

2.4 人の認知に関する基本的事実

視覚による観察を素早くかつ簡単に行う能力は、規則性と不規則性を認識する脳の能力に基づいている。このような脳の能力の多くは無意識に働いている。考え始める前に、比較<するという脳内の認識のプロセス>はほとんど終わっている。

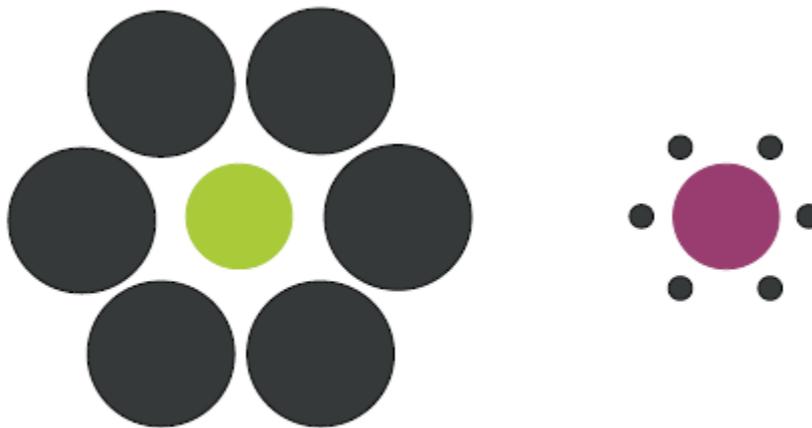
下の図を見てみよう。横棒のどちらの端がより暗いだろうか？



Source: Helmholtz, H. (1821-1894), "Simultaneous Contrast Illusion", In *Wikipedia, The Free Encyclopedia*, retrieved in July 2009 from http://en.wikipedia.org/wiki/Optical_illusion.

横棒の両端は同じ濃さである。背景の違いが誤った認識を生む。背景を紙で覆ってみれば、容易にこれに気付く。

次に、この二つの図を考えてみよう。左図の中心の円と、右図の中心の円ではどちらの円がより大きいだろうか？



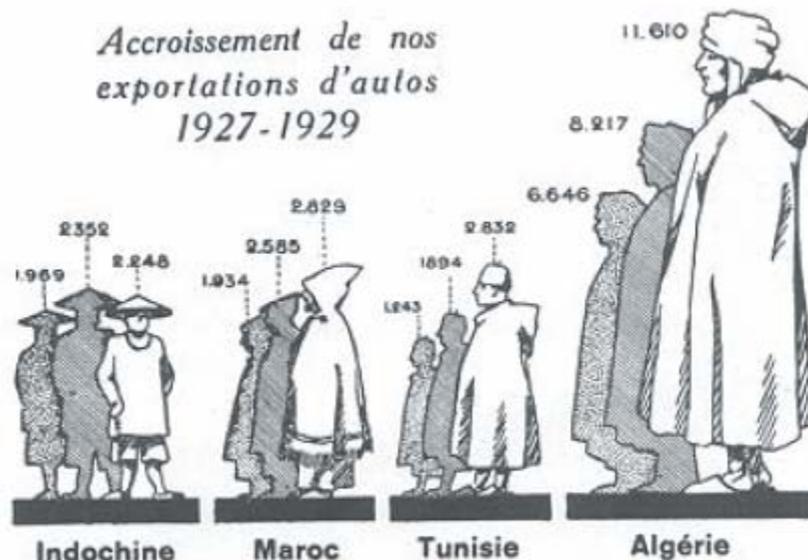
Source: Ebbinghaus, H. (1850-1909), "The Ebbinghaus Illusion", In *Wikipedia, The Free Encyclopedia*, retrieved in July 2009 from http://en.wikipedia.org/wiki/Hermann_Ebbinghaus.

それぞれの図の中心にある円は同じ大きさである。信じないなら、二つの円の大きさを計ってみよう。

これら二つの例は、人間の心が、意識することもなく周囲の環境の中で対象を見ていることを明らかにしている。周囲のものとの対比こそが、横棒の濃さと二つの円の大きさについての<錯覚に基づく>結論に導いているのである。周辺のものを取り除けば、正しい結論により達しやすくなる。

統計家へのメッセージはこれである。統計的な観察について視覚的な表現を作り出すときには、慎重でなければならない。発見したことが示されている背景・状況は、ユーザーの受け止め方をゆがめるかもしれない。

3番目の例を見てみよう。以下のグラフは1927年から1929年までのインドシナ、モロッコ、チュニジア、アルジェリアにおける自動車の輸出の増加を伝えようとするものである。男性の絵の大きさは、実際のデータの変化をどれほどよく反映しているだろうか？



Source: Satet, R. (1932), *Les Graphiques*, Paris. Quoted in Tufte (2001).

自動車輸出の増大(1927年～1929年)

※ 比較されている国名は左から「インドシナ」「モロッコ」「チュニジア」「アルジェリア」

絵を見るとき、私たちの心はそれぞれの対象物の相対的な大きさを比較する。このグラフでは、1927年と1929年間の車の輸出に関するデータは4セットの絵の高さと体積(ボリューム)によって示されている。<これら4か国の間の>相対的な値は二つの意味で不正確に伝えられている。

1. 目盛が正確でない。アルジェリアを示す絵の高さはずっと大きいはずである。4倍大きいものは4倍大きく又は4倍高く示されるべきである。
2. 以前の年の絵の上に最新の年の絵を重ねることで、対象年における値の差を計算するのがより難しくなっている。前面の絵はその後ろにある絵よりもずっと大きく見える。対象年における変化を示しているのは、それぞれの絵の相対的な高さだけである。

2.5 知覚も経験に基づいている

グラフがどのように受け止められるかについて、経験も一定の役割を果たす。読者とその能力、経験、見込まれる違いを知ろう。統計のことであれ、主なテーマのことであれ、統計組織が知っていることを読者も知っていると思い込んではいならない。

統計家は、数字の中にパターンを見分ける自身の経験に意識を払わなければならない。そのような経験は、その点に関しておそらく経験不足の読者とは対照的なものである。統計数値を分析する専門家として、統計家は根底にあるメッセージを理解する点で一般の人々より優れていると思われる。多くの読者が統計情報を認識するのに多くの手助けを必要とするため、統計専門家は<統計情報からの>発見を説明するために視覚化されたものを使う。データの視覚的表現は、主要な発見を観察しやすく、また、理解しやすくするためのものである。

2.6 ツールを誤用しない

数値情報を視覚化する際には多数の技術的ツールが利用可能である。グラフ表現の方法論、デザイン又は技術に焦点を置くのではなく、グラフの本質とメッセージに必ず焦点を当てよう。

基本的な表計算ソフトでは、70種類以上の標準的な種類のグラフを選べ、カスタマイズされたグラフを作ることもできる。ワクワクするほど幅広いこれらの選択肢は、しばしば情報提供者にデータを視覚化するあらゆる種類の方法を使う気にさせる—それらが使えるというだけの理由で。しかし、その結果として、統計情報が複雑過ぎたりゆがんだりすることもしばしばである。

一貫して成長を続ける派手で多彩な新しいツールが利用できるようになって、この危険は増している。これらのツールを使うときには、技術は目的のための手段に過ぎないことを心に留めておくべきである。統計組織がどれほど多くの派手なツールの使い方を知っているかを示すことより、統計情報に含まれるメッセージの方がはるかに重要である。

2.7 データの良い視覚化を行うためのチェックリスト

視覚的な表現を作るときには、以下のことを考えるべきである。

- **対象にするグループ:**異なる読者(例えば、ビジネス界、学界、専門家又は一般の人々)には、異なる形の表現が必要になる。
- **表現全体の中でのグラフ・図案の役割:**全体像を分析するか、要所に焦点を合わせるかによって、視覚に訴える異なるタイプの表現が必要になる。
- **どこでどのようにメッセージが示されるのか:**長くて詳細にわたる分析なのか、それとも時間をかけないスライドショーなのか。
- **理解をゆがめるかもしれない背景・状況:**<対象とするのは>専門家か、それとも初心者のデータユーザーか。
- **文章による分析又は統計表の方が<グラフ等よりも>より良い答えである可能性はないか。**
- **アクセシビリティへの配慮:**
 - グラフや画像のような非文章の要素には、その代替となる文字情報も提供しよう。
 - 色だけに頼らないようにしよう。その色を使わないとしても、その表現は分かりやすいか?色の組み合わせには、十分なコントラストがあるか?その色は色覚障害(赤・緑)のある人にも見分けられるものか?<訳注:後述のとおり、赤と緑を使うと、色弱者の人々には違いが見えづらくなるので、赤は紫に置き換えるべきである。>
 - 時間に依存するコンテンツ<訳注:アニメーションや動画のように時間の経過とともに変化するコンテンツ>をユーザーがコントロールできるようにしよう(例:動画で表示されるグラフを停止できるようにする)。
- **データを視覚化する際の一貫性:**視覚化された要素が一貫性を保ってデザインされ、できるだけ一般的な慣例に従うようにしよう(例:地図上では青で水面を表す)。
- **<文章の>大きさ・所要時間・複雑さ:**表現は分かりやすいか?与えられた時間で読者が内容を把握するのに多過ぎないか?
- **誤解される可能性:**同僚、友人又はターゲットグループの何人かの人々に成果物をテストしてみて、意図しているメッセージを彼らが受け取るかどうかみてみよう。

3. 統計表

3.1 統計表が重要な理由

良い統計表というものは、報道発表資料であろうと、分析記事であろうと、研究論文であろうと、その欠くべからざる一部である。統計表を有効に使うことは、文章中で扱われるデータの数を最小限にするのに役立つ。ストーリーの筋に関わりの薄い、それほど重要ではない変数について論ずる必要もなくなる。

数字について書いた本の中で、Miller(2004年)は良い統計表の作り方について以下のガイドラインを示している。

- 統計表の中で、読者が数字を見つけやすく、理解しやすくしよう。
- 読者の注意が統計表の構造よりもデータが伝える本質的な点に向くように、直接的で控えめな方法で統計表のレイアウトとその表題を考えよう。

この章では、効果的な統計表の作り方を扱う。

3.2 二つの種類の統計表

統計表には二つの種類があることを理解しなければならない。まず一つは、プレゼンテーション(あるいはデモンストレーション)表と呼ばれる小さな統計表である。これは報道発表資料、ウェブページあるいは分析を述べる出版物において、重要な数字を強調するために使うことができる。

二つめは、参照表と呼ばれる大きな統計表から成り立っているものである。それらは、ユーザーがオンラインで自ら統計表を作成することができる双方向のデータベースにますます取って代わられつつある。参照表はく視角化のツールというより>むしろ解析的なツールであるので、ここで議論することはしない。

プレゼンテーション用の統計表において、付随する分析結果をサポートするように、簡潔で十分に体系だった方法でデータが示されなければならない。小さくて配列がよく考えられた統計表は大量の情報を提供でき、読者はすぐにそれを理解することができる。

報告書、記事、出版物あるいはウェブページの中で公表されているかどうかにかかわらず、統計表はそれ単独で成り立つべきである。各統計表は、他の文書にそのまま貼り付けることができるように、かつくそのような形で利用されても>なお意味を持つように、記述的な表題や資料出所の表示など十分なメタデータを含むべきである。もし、統計表がそれ単独で成り立つようにできれば、統計表はもともとの文脈の内側であれ外側であれ、正確に理解される可能性が高くなる。

3.3 良い統計表を設計するためのチェックリスト

統計表に表されたデータを説明するには、それをサポートする五つの要素が必要である。

- **表題(table title)**はデータについて明確かつ正確な記述を表すべきである。表題は、「何が」「どこで」「いつ」という三つの質問に答えるべきである。短くて簡潔なものとし、動詞を使うことは避けよう。

良い例
職業別退職予定年齢、カナダ、2007年
この文章は、データを正しく理解し、使うために必要なすべての情報を提供している。

- **表頭(column headers)**は、統計表の上部にあり、表のそれぞれの列に示されているデータを特定し、関連するメタデータ(例:測定単位、期間、地理的な領域)を提供すべきである。
- **表側(row stubs)**は、統計表の最初の列にあり、表の各行に示されているデータを特定すべきである。
- **脚注(footnotes)**は、統計表の下の部分にあり、正しくデータを理解し利用するために必要な追加情報(例:定義)を提供する。
- **資料出所(source line)**は、統計表の下の部分にあり、データの出所、すなわちデータを作成した機関及びデータの収集方法(例:人口センサス、労働力調査)を提供すべきである。

下図に、これら五つの表の要素がどのように表示されるべきかを示す。

Table title 表題	
Row stubs 表側	Column headers 表頭
	Data データ
Footnotes 脚注	
Source 資料出所	

プレゼンテーション用の統計表においては、以下に示す良い例のように、メッセージを最もよく伝えるよう、数を絞った数字の組み合わせだけが表示されるべきである。

プレゼンテーション用の統計表の良い例

カナダにおける州・準州別製造業売上高、2008年6-7月(季節調整済)

GOOD EXAMPLE of a presentation table

Manufacturing sales in Canada, provinces and territories, June-July 2008
Seasonally adjusted

	June 2008 ^r	July 2008 ^p	June-July 2008
	\$ millions		% change ¹
Canada	52 685	54 105	2.7
Newfoundland and Labrador	692	674	-2.5
Prince Edward Island	123	115	-6.1
New Brunswick	1 914	1 872	-2.2
Quebec	13 019	13 280	2.0
Ontario	23 902	25 015	4.7
Manitoba	1 360	1 445	6.2
Saskatchewan	1 079	1 108	2.8
Alberta	6 298	6 316	0.3
British Columbia	3 347	3 306	-1.2
Yukon	3	4	45.5
Northwest Territories and Nunavut	4	3	-27.4

^r Revised

^p Preliminary

¹ The percentage change is calculated from data in thousands of Canadian dollars

Source: Statistics Canada

※ 脚注部分

^r 改定値

^p 速報値

¹ パーセントポイントの変化は、千カナダドル単位のデータを基に計算したもの。

データの値は、重要な情報が容易に引き出せるように並べられなければならない。メッセージ次第で、ユーザーがより簡単に統計表の行や列に目を走らせることができることに気づくだろう。統計表を縦長にするか横長にするか決める際には、こうした点を考えなければならない。また、罫線や網掛けは、ユーザーが統計表を水平あるいは垂直に読むよう促すために利用できる。余白を空けたり影をつけたりすることで、表の読まれ方を変えることができる。

<統計表で縦に並ぶ>列は均等の間隔に区切られるべきであり、あまり離れていてはならない。統計表は、データに必要なだけの幅とすべきである。

統計表を理解しやすいものにするために、以下のようなガイドラインを考慮に入れるべきである。

- 不要な文章は避ける。
- 時系列に沿って、あるいは何らかの標準的な分類に従ってデータを示す。長期の時系列では、例えば月次の失業率のような場合には降順の時系列(最新の期間から始めて過去に遡る)を使う方が適切なこともある。
- 小数点以下の桁数を最小限にする。
- 千の位を示す分離符号を使用する。符号の代わりに空白を使うと、言語間で翻訳をしなければならない問題を避けることができる。<訳注:分離符号には日本ではカンマ(,)を使うが、ピリオド(.)を使うことが一般的な国もある。このため、具体的な符号を使うと、翻訳の際に符号を変換する手間が生じることになる場合がある。>
- 数字の相対的比較が明らかになるように、小数点の位置を合わせる(小数点がないときは右詰めで)。数字がすべて同じ桁数でない限り、数字を行の中央に揃えることはしない。

- データの入るセルを空のままにしない。欠測値は「利用不可(not available)」又は「該当なし(not applicable)」と特定されなければならない。「NA」はどちらの略語としても通用するので<いずれであるのか>定義する必要がある。<訳注:「not available」は当該データが秘匿すべきものである場合や回答が得られなかった場合、「not applicable」は当該セルに含まれるべき対象が存在し得ない場合などに用いられる。>

これらガイドラインのいくつかは、次の節で論述・説明される。

出版物やウェブサイトのために一連の統計表を作成する際には、すべての統計表で同じレイアウトを使うべきである。統計表の表題にはどのくらいの情報が必要なのか(何が明確で、何が明確でないのか)よく考えよう。略語の使用にも一貫性を持たせよう。

3.4 数字の丸めと小数の利用

統計に詳しくないユーザーの多くは、3桁以上の異なる数字の間に違いを見つけ出すことは難しいと感じている。統計表に示された数値を丸めることで、そのようなユーザーを助けることができる。データが高精度でないときにも、数字の丸めを利用することができる。場合によっては、丸められたデータのみが使ってもらいやすいため、丸められた数字のみが統計表に表示されるべきである。しかし、データを丸めるときには、情報を失い過ぎないように注意しなければならない。

良い例	悪い例
1 320 000	1324567
1 670 000	1673985
1 830 000	1829456

上の例では、左側の丸められた数字は、右側の正確な数字よりも分かりやすく記憶に残る。1,000の区切りとして空白を使っている例も示されている。

小数点以下の異なった桁を表示する必要があるなら、数字を右に詰めるのではなく、小数点の位置をそろえるべきである。以下の例では、左側の値は右側の値より読みやすい。この例は、すべての値に小数点以下の同じ位を持たせることが<小数点以下の桁がそろわない数字を示すより>はるかに良いことも示している。

良い例	悪い例
93.2	93.2
1045.0	1045
385.6	385.63

数値は右詰めにされるべきである。同じ例を使って数字が下のように左詰めにされると、数値を読みづらいことがわかるだろう。

良い例	悪い例
93.2	93.2
1045.0	1045.0
385.6	385.6

3.5 統計表の改善例

3.3 節で示したガイドラインが効果的であることを示すため、悪い統計表の例とそれがどのように容易に改善できるかを示す。

悪い例

分野別エネルギー最終消費—パーセント

BAD EXAMPLE

Final energy consumption by sector - Percentages

	1980	1985	1990	1995	2000	2002	2003
Transport	27.81	27.92	28.24	31.12	36.82	39.48	39.13
Residential	31.11	33.91	30.41	27.61	24.33	23.71	23.97
Industry	31.47	27.21	23.86	22.11	21.41	19.53	18.78
Agriculture	n/a	n/a	3.51	3.7	3.11	2.91	2.82
Services	9.61	10.96	13.98	15.46	14.33	14.37	15.3
Total	100	100	100	100	100	100	100

上の統計表で何が悪いのか？

- データがどの地域について述べているのかわからない。
- 資料出所が確認できない。
- 値が右詰めではなく中央に並んでいる。
- 値は小数点第 2 位で表されるべきではない(過剰な情報である)。
- 合計の値は、他の値と同じ小数点の桁にそろえるべきである。
- 略語「n/a」が説明されていない。
- 灰色の陰影と、各行と各列の間にある同じ太さの線は、表で示されている各データの相違を理解する助けになっていない。
- 表が不必要にページの幅全体に広がっている。

良い例

エネルギー消費の割合・分野別(%)

アイルランド、1980年-2003年

GOOD EXAMPLE

Share of total energy consumption, by sector (in percent)

Ireland, 1980-2003

	1980	1985	1990	2000	1995	2002	2003
Transport	27.8	27.9	28.2	31.1	36.8	39.5	39.1
Residential	31.1	33.9	30.4	27.6	24.3	23.7	24.0
Industry	31.5	27.2	23.9	22.1	21.4	19.5	18.8
Agriculture	n/a ¹	n/a ¹	3.5	3.7	3.1	2.9	2.8
Services	9.6	11.0	14.0	15.5	14.4	14.4	15.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

¹ Data on energy consumption for the agricultural sector was not collected until 1990.

Source: Department of Public Enterprise, Ireland

※ 脚注部分

¹ 農業分野におけるエネルギー消費については 1990 年以前はデータがない。

この表はどのように改良されただろうか？

- データを理解するために必要なすべての情報が表題及びサブタイトルで示されている。
- 資料出所が特定されている。
- すべての値が右詰めされており、小数第 1 位まで表されている。
- 略語「n/a」が脚注で説明されている。
- 統計表の様々な構成要素(ヘッダー、データ、脚注、資料出所)を分けるための線だけが表示され、意味のない陰影は取り除かれている。
- 表は、表頭やデータをすべて載せるのに必要なだけの幅に収まっている。

4. グラフ

4.1 なぜグラフを利用するのか？

統計は表よりグラフで見せられたときの方がしばしば良く理解できる。グラフは統計データを視覚的に表現したもので、データは棒や線といった記号で表される。グラフは、データを手早くかつ簡単に示し、比較を促し、データ内の傾向や関係を明らかにすることができるので、非常に効果的な視覚的ツールである。

一般的にグラフは、棒グラフや折れ線グラフのような一次元又は二次元の形をとる。三次元のグラフも使うことができるが、そのようなグラフは容易に理解するには複雑過ぎると通常は考えられている。

グラフは、大量のデータの中の傾向を示したり、主要な発見やメッセージを伝えたりするために用いることができる。以下のことを示したいなら、グラフを使うことを考えるべきである。

- **比較**:どれくらいか？どちらが大きいか、小さいか？
- **時系列的変化**:変数はどのように変わっているか？
- **度数分布**:項目はどのように分散しているか？違いは何か？
- **相関関係**:二つの変数には関係があるか？
- **全体に占める相対的な割合**:一つの項目は合計とどのように比較されているか？

この章では、最も一般的な種類のグラフについて吟味し、良いグラフを作るためのガイドラインを示す。

4.2 良いグラフを作るためのチェックリスト

データを示す方法としてグラフが最も適当な方法であると決めたなら、どんなタイプのグラフを利用するのであれ、以下の三つのガイドラインを心に留めておく必要がある。

1. **ターゲットとする読者を明らかにしよう**:対象者はその題材について何か知っているか？
2. **伝えたいメッセージを確定しよう**:データは何を示しているか？複数のメッセージがあるか？
3. **メッセージの性格を確定しよう**:項目を比較したいのか、時間を追った傾向を示したいのか、あるいはデータ内の関係を分析したいのか？

良いグラフは:

- 読者の注目を集める
- 簡潔、明確、正確に情報を提供する
- 判断を誤らせない
- 集約的な方法でデータを表示する(例:多くの円グラフの代わりに一つの折れ線グラフ)
- データの比較を容易にし、傾向や違いを強調する
- 付随する文章でメッセージ、テーマ又はストーリーの筋を説明する

4.3 グラフの使用が適切とは言えないとき

グラフは、統計情報を提供するためのツールとして常に最も適しているというわけではない。時として、文章と統計表の両方又はいずれかの方が<グラフよりも>読者にわかりやすい説明を提供し、かつ、統計組織における相当の時間や手間を省くであろう。

以下のようなデータのときは、グラフを使うことを考え直すべきである。

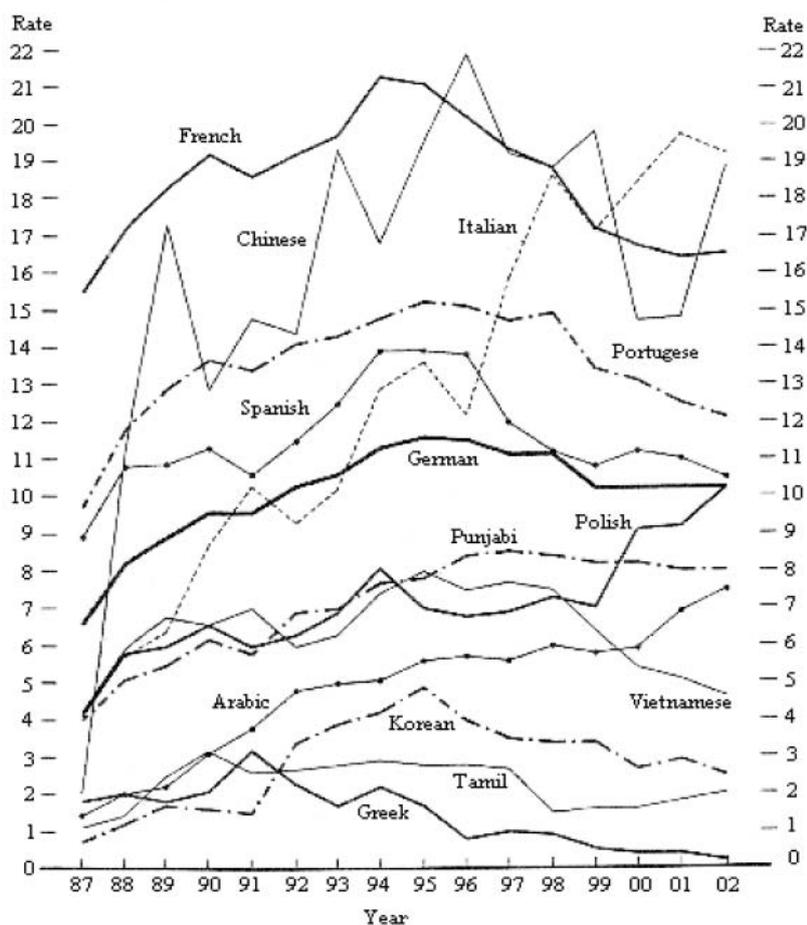
- 分散が大きい場合
- 値が少ししか値がない場合
- 値が多過ぎる場合
- 全く変化がない又はわずかな変化しかない場合

折れ線グラフの悪い例

West High Schoolにおいて英語を第二言語として
選択している生徒の数・母語別・1987年から2002年

BAD EXAMPLE of a line chart

Number of students taking English as a second language
at West High School, by first language spoken, 1987 to 2002



Source: Statistics Canada, *Learning Resources: Using graphs*⁵.

どのようなものであれ、上の折れ線グラフに似たようなものは避けなければならない。データがあまりに多過ぎ、分析者がどのようなストーリーの筋を説明しようと望んでも、折れ線のジャングルの中でそれは見えなくなっている。

⁵ <http://www.statcan.gc.ca/edu/power-pouvoir/ch9/using-utilisation/5214829-eng.htm>

4.4 適切なグラフ種類の選択

どのようなタイプの情報にどのようなタイプのグラフを使うべきかを知ることが重要である。データの性格によって、最も適当なグラフの種類は異なる。この節では、最も一般的なグラフ、すなわち、棒グラフ、人口ピラミッド、折れ線グラフ、円グラフ、散布図について、ガイドラインを示す。

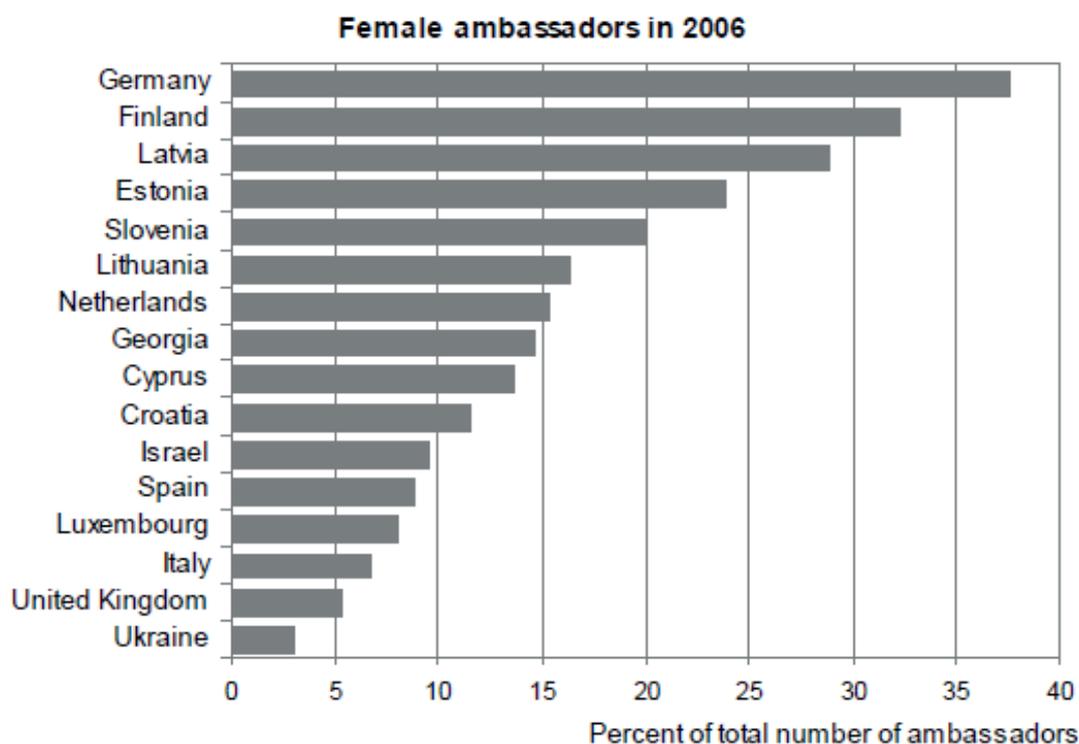
棒グラフ

棒グラフは、描いたり読んだりするには最も簡単なグラフである。棒グラフは異なるカテゴリーやグループの度数や値を比較することに利用される。

棒グラフの良い例

2006 年における女性大使の数

GOOD EXAMPLE of a bar chart



※ 横軸表記

全大使に対する割合(パーセント)

棒は垂直方向又は水平方向どちら向きにもできる。水平方向の場合、上の例のように、文章は簡単に読むことができる。棒は任意の順で並べるよりも、一番小さなものから大きなものへ順番に並べられれば、異なる値を比べやすくなる。

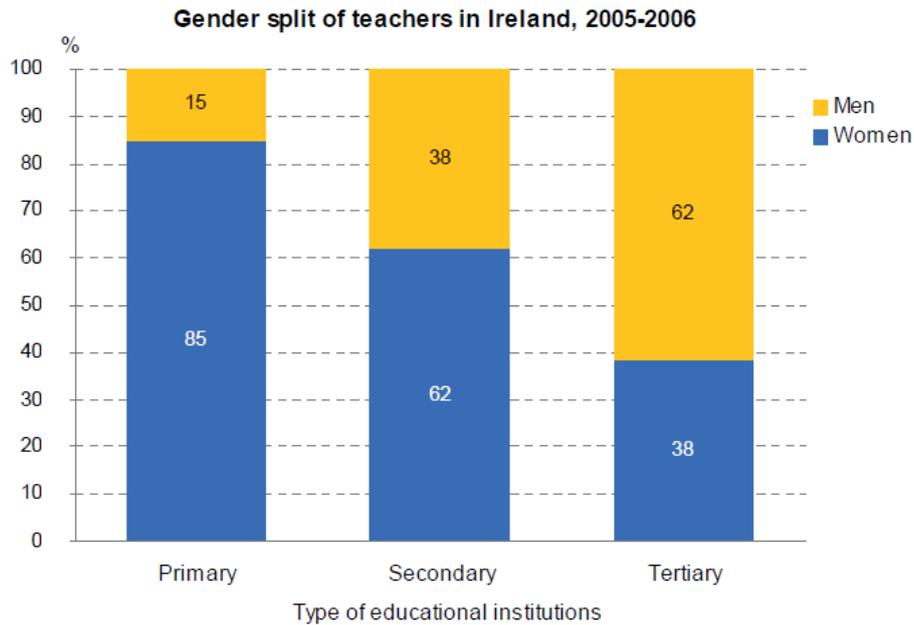
棒は、棒と棒の間隙間よりもずっと太くすべきである。隙間は棒の幅の 40%を超えるべきでない。

帯グラフ<訳注:原文 stacked bar chart は直訳で「積み上げ(積み重ね)棒グラフ」であるが、例示されているグラフに対応させるため「帯グラフ」と訳した。>は、全体の内訳を示したり比較したりすること利用することができる。このタイプのグラフを利用するときは注意しなければならない。それぞれの帯の中にある項目が多過ぎたり、ほとんど同じ大きさの項目が数多くあると、分析や比較が難しくなる。

積み上げ棒グラフの良い例

アイルランドにおける教師の性別割合、2005～2006年

GOOD EXAMPLE of a stacked bar chart



※ 横軸表記

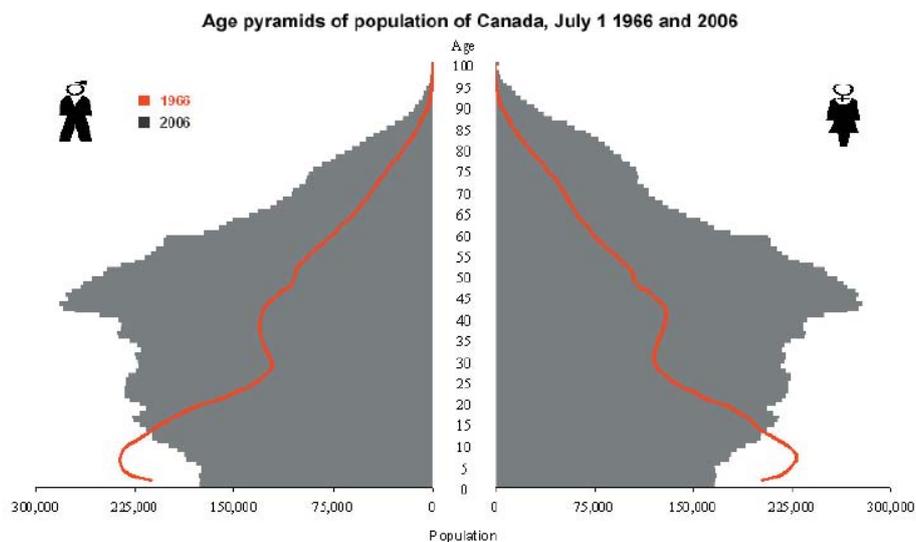
初等 中等 高等
教育機関の種別

人口ピラミッドは、二つの水平な棒グラフの組み合わせであり、国や地域の男性と女性の年齢別構成を表している。慣習として、左側は男性を表し、右側は女性を表す。異なる人口ピラミッドを比較したい場合、通常は、人口そのものの数ではなく、全人口に占める割合で表す方がよい。

人口ピラミッドの良い例

年齢各歳別カナダ人口ピラミッド、1966年及び2006年各7月1日

GOOD EXAMPLE of a population pyramid



⁶ <http://www.statcan.gc.ca/daily-quotidien/061026/figure.htm>

ほとんどのヨーロッパの国々では<訳注：日本においても同様>、人口ピラミッドはもはやピラミッドの形をしていないが、今なお、人口における年齢や性別に関する膨大な情報を効果的に示す表現である。人口ピラミッドが時間とともに形を変えていくようなアニメーションで示される場合は特にそうである。

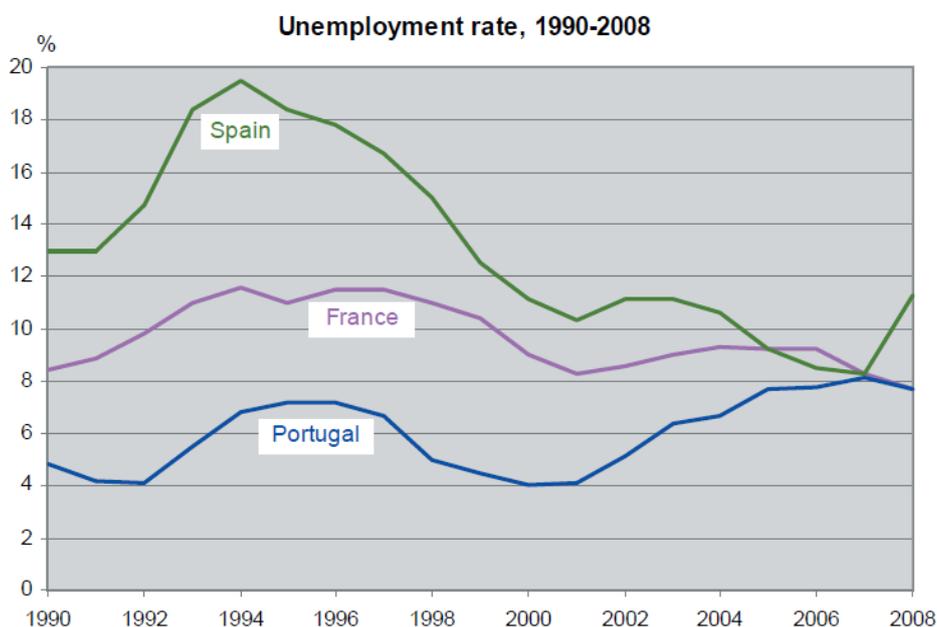
折れ線グラフ

折れ線グラフは、時系列の傾向を見えるようにするための効果的なツールであり、時系列を表す最も適切なグラフである。メッセージをよりよく伝えるためにグラフのパラメータを調節することができるが、データをゆがんだものにならないよう注意しなければならない。この問題は 4.6 節で取り上げられる。

折れ線グラフの良い例

失業率 1990 年～2008 年

GOOD EXAMPLE of a line chart



Source: UNECE Statistical Database

円グラフ

円グラフは一つの変数に係る構成比を示すことに使われるが、限られた数の項目だけ—通常は 6 以内—を表示できる。一つの円内で異なる内訳を比較することは難しく、ましてや異なる円グラフ間でデータを比較するのはより難しいため、この種のグラフを使うことを多くの統計家が勧めていない。この問題点を克服するため、内訳に実際の値を表示することができる。グラフの凡例を不要とするため、カテゴリーの名前をラベルとして書き込める場合もある。内訳は大小ばらばらに並べずに、普通は最も小さいのから最も大きいものへという順序で示すと最も見栄えがよいく<訳注：我が国では、構成比の大きな項目から始めて時計回りに小さな項目になるように並べることが円グラフ作成の基本と考えられている。>。

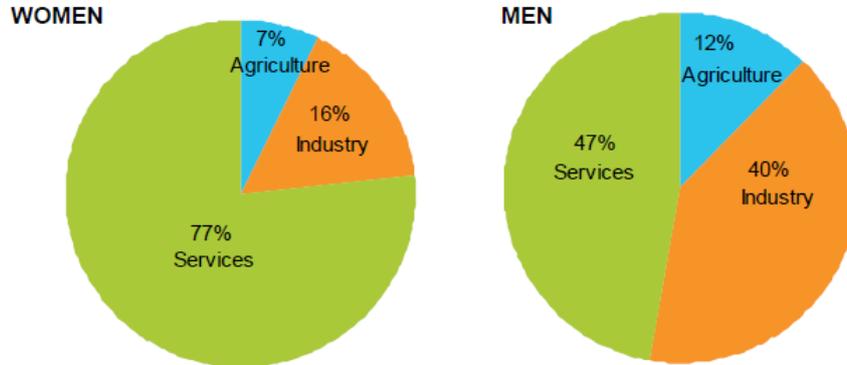
多くの場合、他のタイプのグラフ(例：棒グラフ)の方がより適切であるが、あるカテゴリーが全体の中で占める重要性を見せるためには円グラフは効果的なツールであるから、円グラフを完全に除外してしまうべきではない。円グラフは、以下の例のような状況を概観するにはよく適している。

円グラフの良い例

ラトビアにおける主要分野別雇用 2007 年

GOOD EXAMPLE of a pie chart

Employment by major sectors in Latvia, 2007



Source: UNECE Statistical Database

散布図

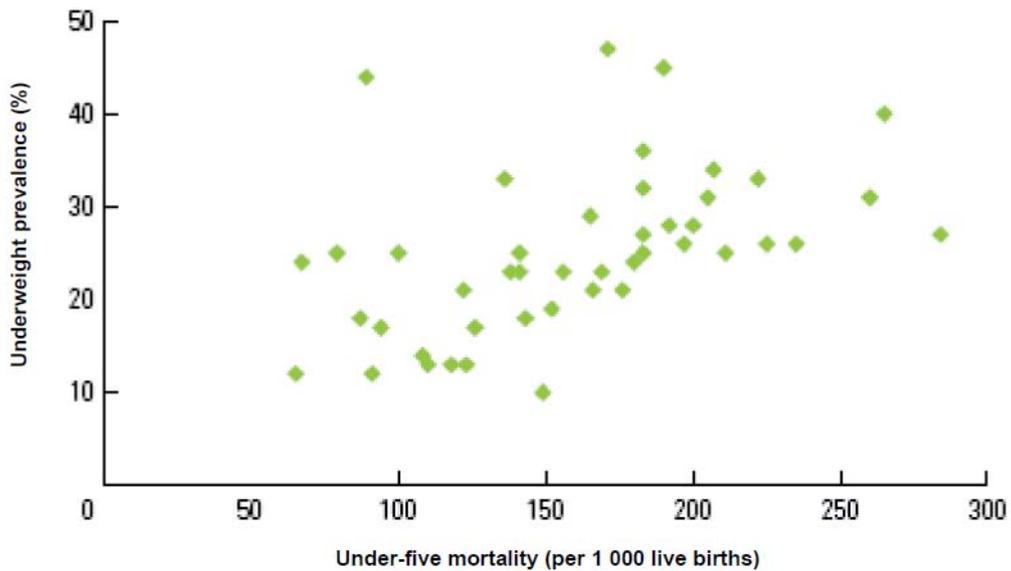
散布図は二つの変数の関係を表すのに利用される。以下の例に示されるように、相関を表すには最も正確な方法である。しかし、散布図は解釈が難しいことがあるので、分析者によっては棒グラフの方を好んで使う。

散布図の良い例

サハラ以南のアフリカ諸国における5歳児未満死亡率及び低体重者率 2003 年

GOOD EXAMPLE of a scatter plot

Under-five mortality and underweight prevalence in Sub-Saharan African countries, 2003



Source: Jamison et al. (2006) *Disease and Mortality in Sub-Saharan Africa, 2nd edition*, Washington D.C., The World Bank⁷.

※ 縦軸表記 低体重者率(%)

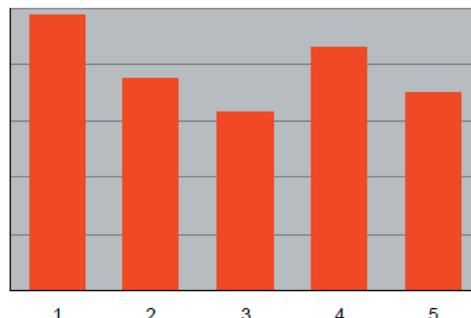
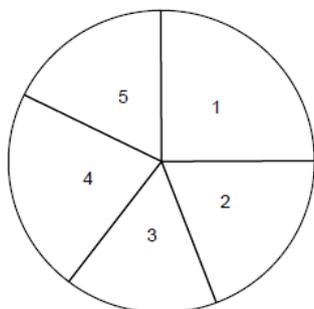
横軸表記 5歳児未満死亡率(出生 1000 件当たり)

⁷ [http://www.dcp2.org/file/66/Disease and Mortality in SSA.pdf](http://www.dcp2.org/file/66/Disease%20and%20Mortality%20in%20SSA.pdf)

様々な種類のグラフを試そう

どのタイプのグラフを使うべきだろうか？それを決めなければならない。一つの良い方法は、様々なグラフを試してみ、メッセージを伝えるのに最も適したツールを選択するというものである。

以下には、同じデータを二つの異なる方法でグラフにしたものがある。どちらがより明確だろうか？



円グラフでどの内訳が一番大きいが見分けることができるだろうか？相当数の読者は、棒や折れ線よりも円グラフで大きさを表す角度を比較することは難しいと思う傾向がある。円グラフでは内訳の1と4はほとんど同じように見える。一方、棒グラフではそれらの相対的な大きさの違いが直ちに明瞭になっている。

4.5 効果的なグラフを作るには

グラフの要素(Chart components)

グラフを構成する様々な要素は、読者の注意を引く上で互いに競合する。グラフに含まれる特徴が多ければ多いほど、重要な点が分かりにくくなる。

グラフの要素は三つのカテゴリーに分かれる。

1. データを表す**データ要素(Data components)**:棒、線、領域又は点
2. データの理解を助ける**補助要素(Support components)**:表題、凡例、データラベル、目盛り線、脚注、資料出所
3. データと関係がない**装飾的要素(Decorative features)**

データ要素だけでは決して十分でない。グラフを正しく理解してもらうために、以下の補助要素を必要とする。

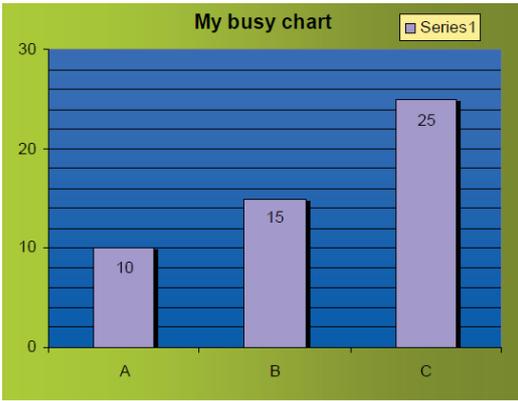
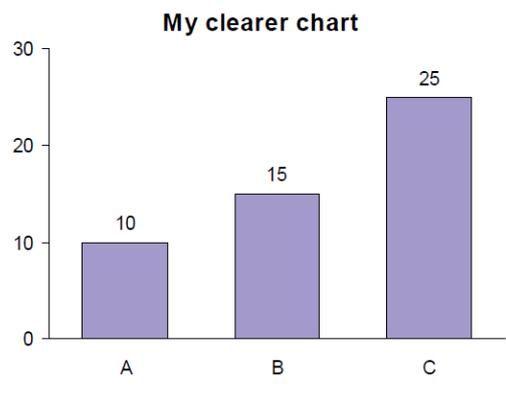
- **グラフの表題**は、グラフが何についてもものなのかはっきりした考えを示すものとすべきである。表題は短く簡明でなければならない。二つのタイプの表題があり得る:
 - **情報伝達型の表題(informative title)**は、そのデータを理解するのに必要なすべての情報を提供する。「何が」、「どこで」、「いつ」という三つの質問に答えるべきである。
 - **記述型の表題(descriptive title)**は、グラフに表われている主なパターンや傾向を強調する短い説明である。それは、グラフが表しているストーリーをわずかに数語で語る。
- **軸のラベル**は、グラフに示された数値を特定しなければならない。ラベルは<縦軸と横軸の>両方の軸に水平に表示される。
- **軸のタイトル**は、データの単位(例:「千」、「%」、「年齢(歳)」、「\$」<ドル>)を特定すべきである。単位が明白な場合(例:時系列の場合における「年」)には、軸のタイトルは必ずしも必要ない。

- ユーザーがデータの値を読んだり比較したりすることを助けるために、棒グラフや折れ線グラフに目盛り線を追加することができる。＜訳注:「目盛」が軸に刻む短い線であるのに対して、「目盛り線」はその目盛をグラフのXY平面(プロットエリア)全体に延ばして書き入れる長い線。＞
- 凡例やデータラベルは、グラフの中でデータを表現するために利用されている記号、形、色を特定すべきである。グラフに表される数値の系列が一つだけの場合には、凡例は表示すべきでない。可能なときは常に、凡例ではなくデータラベルを用いるべきである。データラベルは、確認や理解することを容易にするためにデータ要素(棒、領域、線)の上又は隣に表示される。
- 脚注は、定義や方法に関する情報を提供するために用いられる。
- 資料出所は、グラフの下に明示されるべきである。

すべてはデータに関して

グラフの効果を最大限にするために、データが最も重要な場所を占めるべきである。補助要素は以下のようにあるべきである。

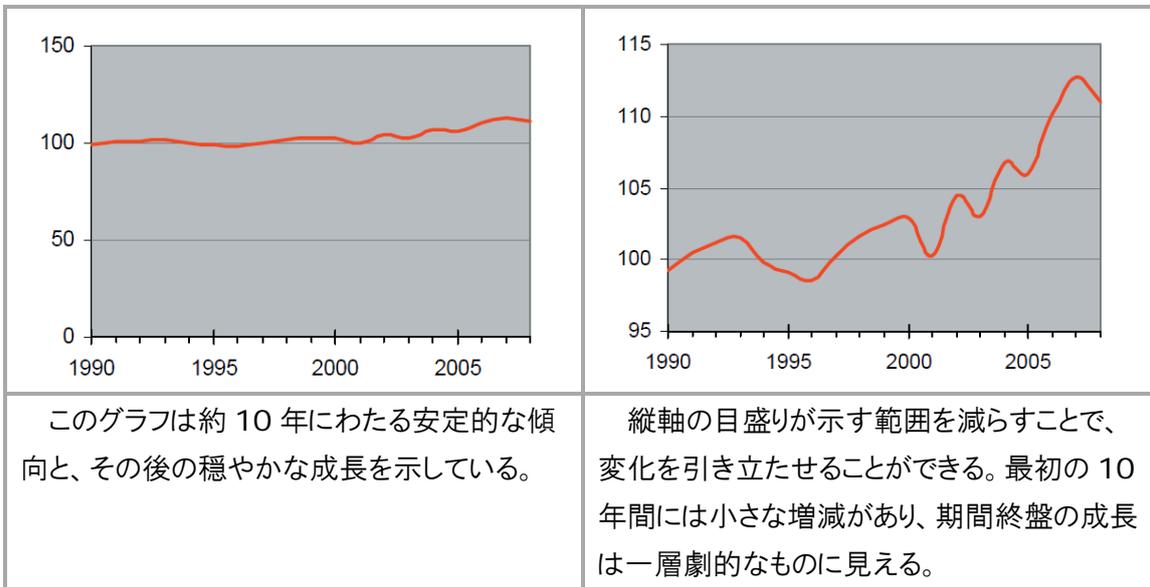
- <補助要素は> 適当な場合にだけ使おう。データの性格によっては、軸のタイトル、凡例、データラベルはグラフを正確に理解するために不可欠かもしれないし、全く不要かもしれない。
- 控え目にしよう。軸や目盛り線はデータ要素の線よりも薄い線を使おう。装飾的要素で、読者の注意をそらはならない。

悪い例	良い例
	
<p>全ての要素が最大限の印象を与えている。その結果、三つの数値しか表していないにもかかわらず、見た目がわずらわしく、読み取りにくいグラフになっている。</p>	<p>このグラフの方がより読み取りやすい。補助的な要素をできるだけ使わなかったことで、データが主役になることが保証されている。</p>

データ要素がお互いにぶつかり合うこともある。表示したい変数や数値が多ければ多いほど、データをはっきりと示すことが難しくなる。効果的なグラフには、明確で視覚的なメッセージがある。あまりに多くのことをグラフにさせようとする、あまりに骨の折れる難解な代物になる。最悪の場合、そのグラフは単に誤解を招くだけになる。

4.6 グラフのパラメータの調整

グラフを描くとき、メッセージを最もよく伝えるために目盛りの範囲を調節することができる。以下に同じデータを示す二つの折れ線グラフがあるが、二つのグラフは見た目が大変異なっている。



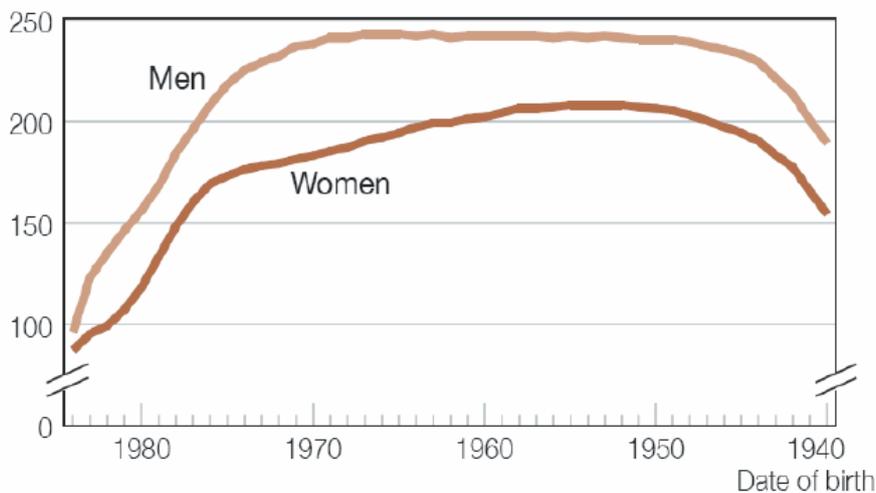
上の右図の Y 軸のように目盛がゼロから始まらないとき、記号を使ってそれを示すのが良いやり方である。最も良い選択肢は、以下の例に描かれたように、ゼロから始まりく軸の途中で>ジグザグの線を引くか間を空けることである。

縦軸が0から始まっていないグラフの良い例

スウェーデンにおける 20 歳から 64 歳の年金基礎収入 2004 年
平均収入・千スウェーデンクローナ

GOOD EXAMPLE of a chart with a y-axis not starting at zero

Pensionable income for persons aged 20-64 in Sweden, 2004
Average income in thousands Swedish Krona



Source: Statistics Sweden (2006), *Women and Men in Sweden: Facts and figures 2006*⁸.

※ 横軸に出生年を表示

4.7 グラフに対する認知負荷のコントロール

グラフを使うことにより強調したいというメッセージはいくつかあるであろう。出版物におけるすべての要素と同様に、グラフには「認知負荷」(cognitive load)が伴い得る。認知負荷は、基本的には、統計組織が伝えようとしていることを読者が理解するのにどれくらい懸命に頭を使わなければならないかということを意味する。認知負荷が高い図は理解

⁸ http://www.scb.se/statistik/publikationer/LE0202_2006A01_BR_X10ST0602.pdf

することや覚えることが難しいものになり、メッセージは伝えられ難いものになる。認知負荷が低い図は一目で理解でき、そのメッセージは明確になる。効果的なグラフ設計についてのガイドラインは、そのほとんどが認知負荷を低く保つことを意図したものである。

グラフを設計するとき認知負荷を調整する。適切な決まり事や様式を使うことによって、認知負荷を低くし、明確なメッセージを送ることができる。読者にデータの仔細な局面に注目してほしいなら、グラフの認知負荷を意図的に引き上げるよう試みることもできる。認知負荷を引き上げることで、読者は違った観点からグラフを考えるよう仕向けられる。下図は認知負荷の高いグラフの例である。

高い認知負荷を伴うグラフの良い例

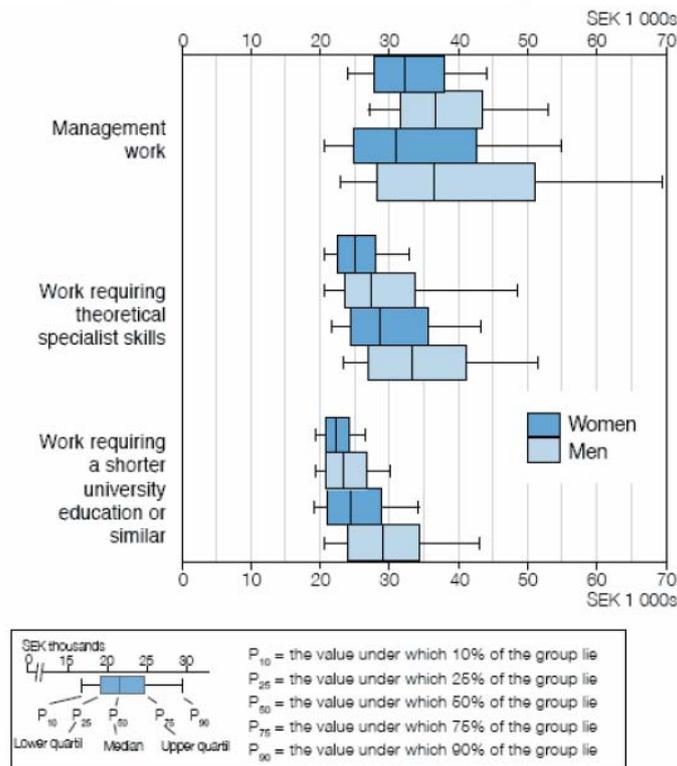
高等教育を必要とする職業集団における賃金の散らばり スウェーデン 2004 年
 月給 千スウェーデンクローナ(SEK)
 各職業集団において上段 2 本の横棒は公的分野における女性及び男性、
 下段 2 本の横棒は民間分野を示す。

GOOD EXAMPLE of a chart with a high cognitive load

Wage dispersion in occupational groups that require higher education in Sweden, 2004

Monthly salary in Swedish Krona (SEK)

The two upper bars in each occupational group show women and men in the public sector, the two lower bars show the private sector.



Source: Statistics Sweden (2008), *Women and Men in Sweden: Facts and figures 2008*⁹.

- ※ 縦軸: 上から「管理的業務」「理論的な専門技術を要する業務」「短期の大学教育又は同等の教育を必要とする業務」
- ※ 凡例を示す枠囲み内には、箱ヒゲ図の各点に関する解説(10 パーセンタイル点、第一四分位点、中央値、第三四分位点、90 パーセンタイル点)が記載されている。

⁹ http://www.scb.se/statistik/publikationer/LE0202_2008A01_BR_X10BR0801ENG.pdf

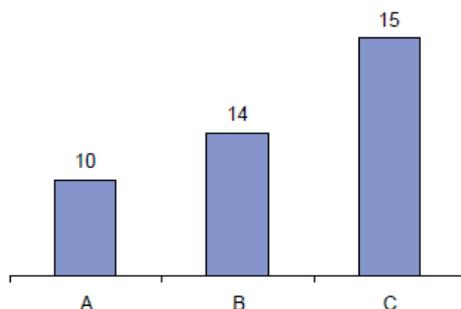
4.8 グラフを改良するための提案

正確にしよう

図表は、比率を正確に示す大きさでなければならない。以下の例のように、不釣り合いな大きさを示しているグラフは誤解を招く。

グラフ要素の相対的なサイズが悪い例

BAD EXAMPLE of relative size between chart objects



データを並べかえよう

棒グラフ又は円グラフを使うときは、比較が容易になるように、最も小さい値から最も大きい値にデータを並べ替えるべきである。〈訳注：前述のとおり、我が国では通常、大きい値のデータから順に並べることが多い。〉

悪い例	良い例																												
<p>思春期妊娠率 2006年</p> <table border="1"><thead><tr><th>Country</th><th>Rate</th></tr></thead><tbody><tr><td>Albania</td><td>13</td></tr><tr><td>Georgia</td><td>36</td></tr><tr><td>Greece</td><td>11</td></tr><tr><td>Hungary</td><td>20</td></tr><tr><td>Romania</td><td>35</td></tr><tr><td>Serbia</td><td>23</td></tr></tbody></table>	Country	Rate	Albania	13	Georgia	36	Greece	11	Hungary	20	Romania	35	Serbia	23	<p>思春期妊娠率 2006年</p> <table border="1"><thead><tr><th>Country</th><th>Rate</th></tr></thead><tbody><tr><td>Greece</td><td>11</td></tr><tr><td>Albania</td><td>13</td></tr><tr><td>Hungary</td><td>20</td></tr><tr><td>Serbia</td><td>23</td></tr><tr><td>Romania</td><td>35</td></tr><tr><td>Georgia</td><td>36</td></tr></tbody></table>	Country	Rate	Greece	11	Albania	13	Hungary	20	Serbia	23	Romania	35	Georgia	36
Country	Rate																												
Albania	13																												
Georgia	36																												
Greece	11																												
Hungary	20																												
Romania	35																												
Serbia	23																												
Country	Rate																												
Greece	11																												
Albania	13																												
Hungary	20																												
Serbia	23																												
Romania	35																												
Georgia	36																												
<p>データは各国アルファベット順に表示されている。数値は比べにくい。特別な意味を持たない最初と最後の数値に関心が向く。</p>	<p>データは小さいものから大きなものに順番に並べられている。数値を比較しやすい。データセットの中の最小と最大に関心が向く。</p>																												

誤解を招く相関関係を避けよう

同じグラフの中で異なる目盛に変数をプロットすることは、誤解を招く結果を生みがちである。二つの折れ線が一緒に動いているという事実は、〈二つの変数の間に〉相関関係を確立するには十分でない。

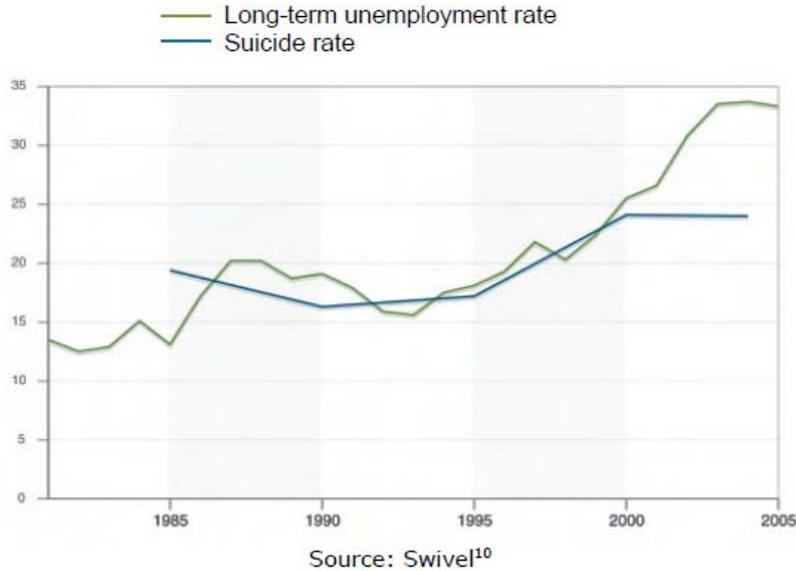
下図は日本の自殺率と長期失業率との関係づけを試みたものである。二つの折れ線は一緒に動いているように見えるが、二つの変数は違っている。一方の折れ線は10万人あたりの自殺者の数を表しているが、他方は失業者数全体に占める百分率として示された12か月以上の失業者数である。二つの変数はたまたま全体の期間を通して10から35の間に値があるが、定義又は目盛をわずかに変えるだけで全く異なったグラフになったことであろう。このグラフは両方の変数が関係していることを印象づけることに成功しているが、〈このグラフによって〉それが立証されているわけでも、立証できているわけでもない。

悪い例: 二つの変数間の誤った相関関係

日本における自殺率と長期失業率

BAD EXAMPLE: misleading correlation between two variables

Suicide and long-term unemployment in Japan



2本の縦軸は慎重に使おう

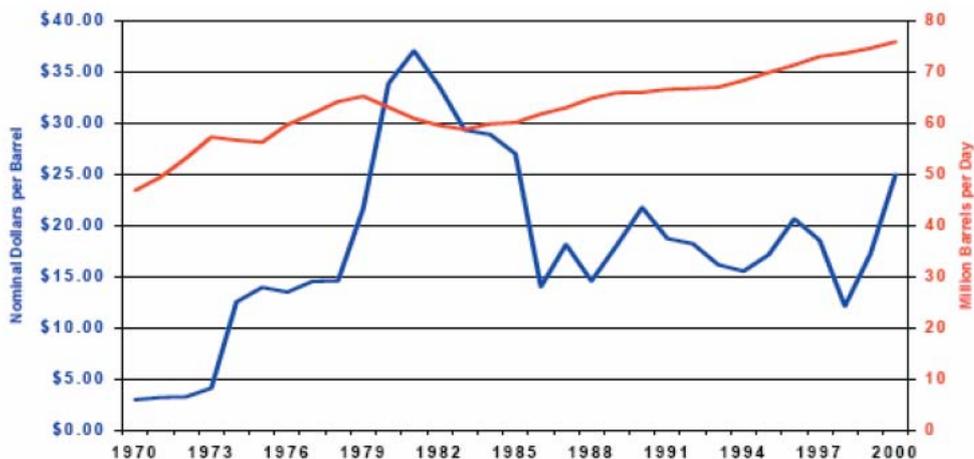
2本の縦軸は混乱を引き起こす可能性がある。二つの異なる変数(例:価格と量)があるときは、このタイプのグラフがうまく利用できるかもしれない。しかし、<縦軸の>表示に十分な注意を払い、以下の例で示されているように、それぞれの折れ線はユーザーが参照すべき軸と同じ色で示される必要がある。

二本の縦軸を持つグラフの良い例

石油に対する世界の需要と価格

GOOD EXAMPLE of a dual y-axes chart

World demand and world oil prices



Source: Blessing et al. (2003), *Cognitive Testing of Statistical Graphs: Methodology and Results*¹¹.

同じグラフで二つの異なる目盛に似たような単位がある場合、専門家でさえ混乱してしまうことがある。同じ目盛で描かれれば互いに近づくことのなかった線でも、2本の縦軸を使ったグラフでは交差するかもしれない。

¹⁰ <http://www.swivel.com/graphs/show/28847825>

¹¹ <http://www.fcs.m.gov/O3papers/BlessingBradsher.pdf>

必要のない図柄を避けよう

提示しようとしているデータセットについて何も表していない図柄はグラフを読みにくいものにする。特に、単純なデータセットを表すときに三次元のグラフを使う場合に言えることである。簡単に三次元のグラフを作成することができるソフトウェアがあり、三次元のグラフを魅力的だと考える人がいるかもしれないが、三次元のグラフはしばしばデータをゆがめてしまう。下に示したように、単純な二次元の形式で示される方がはるかに容易にメッセージを理解することができる。

悪い例

アイスランドにおける法的婚姻関係別 18 歳以上
人口 2004 年

BAD EXAMPLE



良い例

アイスランドにおける法的婚姻関係別 18 歳以上
人口 2004 年

GOOD EXAMPLE



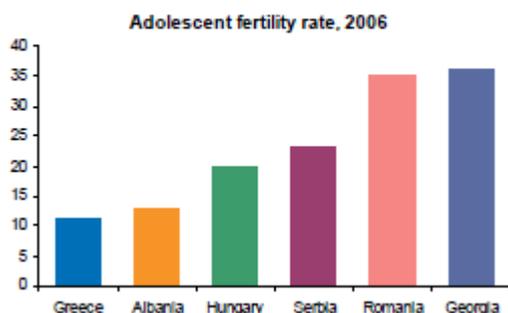
三次元のグラフはほとんどの場合に価値を高めることはなく、しばしば読者を混乱させる。図が奥行きを持つと、ある部分は近くに感じるし、それ以外の部分は遠くに感じる。脳は遠くにあるように見えるものは大きいと思い込むことでこれを補正している。しかし、二つより多い属性を持つデータを表しているときには、奥行きを利用することで、読者に重要な点をより明確に理解させることができる。

一般に、いかなるデータも伝えない要素を加えることは避けるべきである。例えば、以下の例に示したように、棒グラフの中でいくつもの値に異なる色を使うことは、グラフを読み取ることを難しくする。一つの色にそろえるべきである。

悪い例

思春期妊娠率 2006 年

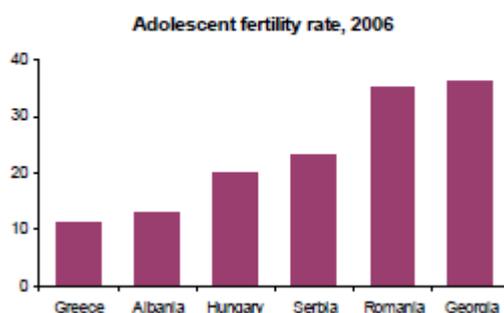
BAD EXAMPLE



良い例

思春期妊娠率 2006 年

GOOD EXAMPLE



5. 地図

5.1 地図が千の数字に匹敵する理由

地理情報はすべての統計データに不可欠な部分となっている。地理的領域は、場所を特定し、統計情報を関連づけることを可能にする境界、地名、その他の情報を有している。この空間的關係は特にセンサス(＜訳注：日本なら国勢調査に代表されるような全数調査)データにとって重要である。

地図は、空間的なパターンを視覚化する最も効率的なツールである。慎重に設計・提示されると、地図は統計の表現において単なる飾り以上のものになる。それらは統計表やグラフからは明らかにできない分布やパターンを特定し、強調するのに役立つ。

「一枚の写真は千の言葉に匹敵する」＜訳注：英語の原文は「百聞は一見に如かず」の意味で用いられる＞なら、「一枚の地図は千の数字に匹敵する」。視覚の時代にあつて、地図は強力な情報メディアである。それらは専門家、政治家や一般の人々のための貴重な意思決定ツールとして役に立ち、社会のあらゆる部分で増大している情報に対する要求を満たすものである。

地図の力

理解することが容易な地図は上手に設計されている。そのような地図は、膨大な情報をひと目で理解するのに役立つ。それらは、ボリュームのあるデータ表や、長く込み入った文章を要約することができる。国のすべての地域の統計情報を提供しようとするとき、グラフの完全なセットを作ることもできる(＜訳注：グラフで表そうとすれば膨大な数のグラフが必要になるとの趣旨＞、すべての情報をただ一つの地図の中に示すこともできる)。

幅広い分析機能を提供し地図の構成要素をまとめる地理情報システム(GIS)から、統計組織における専門的な地図及び地図帳の作成者向けの高性能な地図製作情報システム(CIS)まで、今や多くの地図製作技術が利用できる。紙による出版物の時代には、地図は、正確な数値を示さないがゆえに統計学者にはあまり使われなかった。“地図の背後にある”実際のデータをユーザーが引き出せる双方向のマッピングツールが登場したことで、この欠点は解消された。

統計における地図の利用

地図は統計調査の準備、結果の分析と報告といういずれにおいても非常に有用である。以下を望むなら、地図を使うことを考えるべきある：

- 地理的な場所とデータの空間的な分布を示そうという場合；
- 異なる地域間を比較しようという場合；
- 大きなボリュームのあるデータを要約して複雑さを減らそうという場合；
- 明確なメッセージを伝えようという場合；
- 見い出されたことを検証しようという場合；
- 人々の注意を引き付けようという場合；
- 空間情報を地理情報システムに格納しようという場合。

この章では、最も一般的な地図を検討し、良い地図を作成するためのガイドラインを示す。

5.2 良い地図を作るためのチェックリスト

地図製作は、芸術・科学・技術を混ぜ合わせたものである。レイアウトの整え方には無限の可能性があり、地図製作は複雑な作業である。

Web2.0 技術(6.2 及び 6.4 参照)の成長に伴い、多くの双方向のマッピングツールがオンライン上で利用可能であり、ユーザーはデータをアップロードし、手間をかけずに主題図(thematic maps)を引き出すことができる。主題図の製作ははるかに安く、速くなったが、それによってメッセージを正確に伝える、上手に設計された地図が自動的にもたらされるわけではない。

良い地図を設計するために、次の四つのガイドラインを考慮する必要がある:

1. **対象とする読者を明確にしよう:** 地図はどのように、どのような文脈の中で使われるのか? アクセスしやすさに制約があるか?
2. **伝えたいメッセージを決めよう:** データは何を示しているか? 一つ以上のメッセージがあるか?
3. **データの性格を決めよう:** いくつの変数が地図上に落とし込まれるべきか? 時間軸があるか?
4. **適切な地図作成技術、色及び記号を決めよう:**

データはどのような性格か(量的データか質的データか、絶対値か相対値か)? 技術的な制約はあるか(例:フォーマット、モノクロ再生)? 色や分類についての決まりごとは何か?

良い地図は:

- 単純で容易に理解される;
- 明確で客観的なメッセージを持つ;
- データを正確に表示し、誤解を招かない;
- 最も重要な情報に読者の注意を引き付ける;
- ふさわしい形で示され、興味を引く;
- 成果物のフォーマットとしてふさわしく、読者にも適している;
- さらに説明を加えなくても、それだけで成り立つ;
- 色覚障がい者にとって利用しやすい。

5.3 地図を作る意味がないとき

地図を作り始める前に、地図がデータを提供する最も適切なツールであるかどうか考えるべきである。グラフ又は統計表の方がメッセージを伝えるのにより良い方法になり得るなら、時間と労力を無駄にしてはならない。

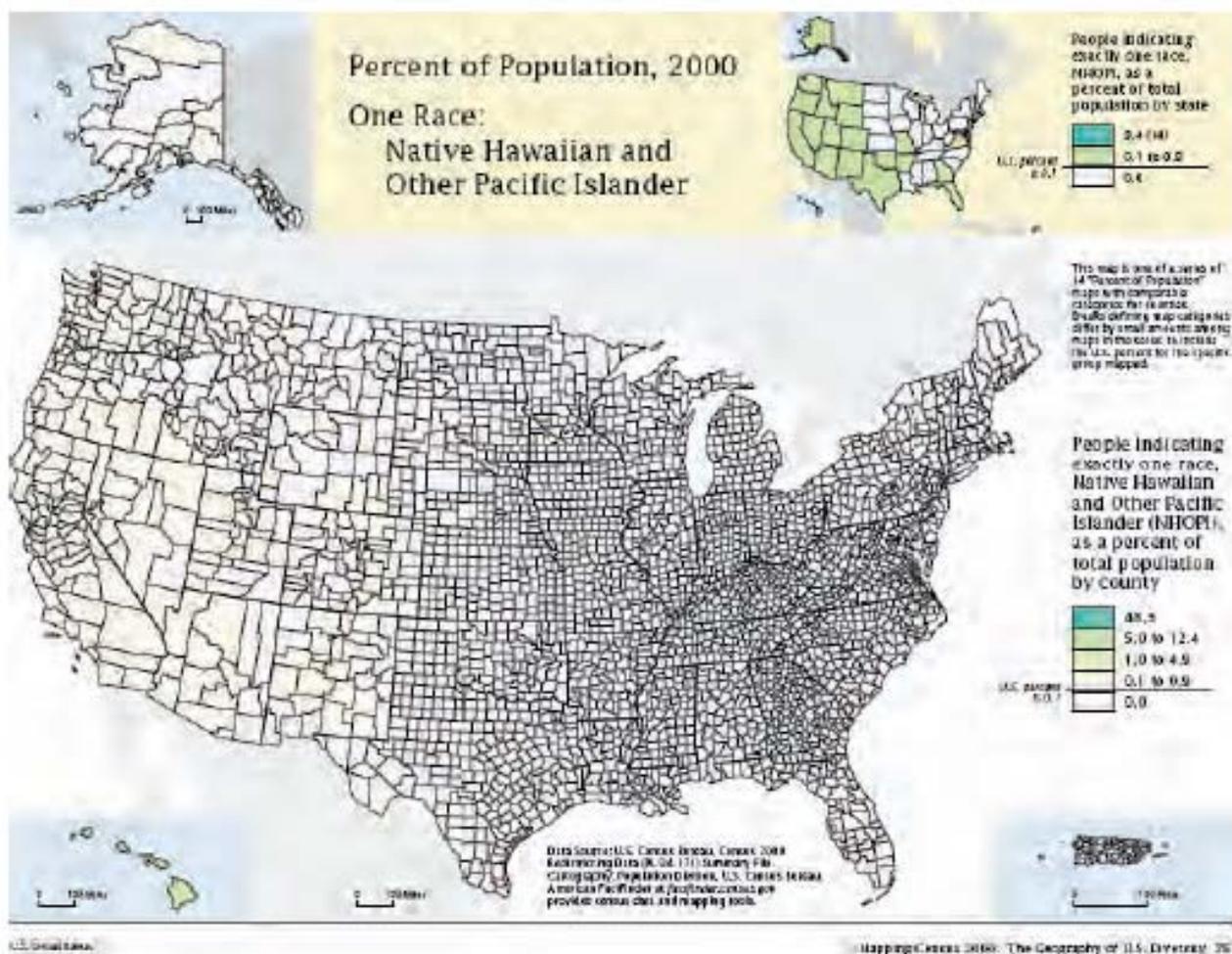
以下のような場合には、データを地図に表わす意味がない:

- データが地理的に詳細化できない場合;
- データに意味のある変化がない場合;
- 対象となる読者が地図を理解することが苦手であろう場合;
- 適切に読まれ、理解されるような地図を示すための十分なスペースがない場合。

地図の悪い例

人口の割合 2000年 ハワイ原住民と他の太平洋島しょ民族

BAD EXAMPLE of a map



Source: Brewer, C.A. and Suchan, T.A., U.S. Census Bureau (2001), *Mapping Census 2000: The Geography of U.S. Diversity*, U.S. Government Printing Office, Washington DC¹².

上の例は、地図をに十分な情報がないと、地図がいかに無駄になるかを示している。上の地図は米国における少数民族、すなわちハワイ原住民と他の太平洋島しょ民族(NHOPI)の空間的分布を表すことを狙っている。このグループはほとんどの郡で総人口の1%未満しかいないため、地図はほぼ空白である。<総人口の>1%より多いNHOPIがいる郡はほとんどなく、容易に見い出すことができない。

5.4 様々なタイプの地図

地図は縮尺、機能、デザイン、製作技術あるいは出版物の中での使われ方によって分類することができる。

一般に、二つのタイプの地図がある：

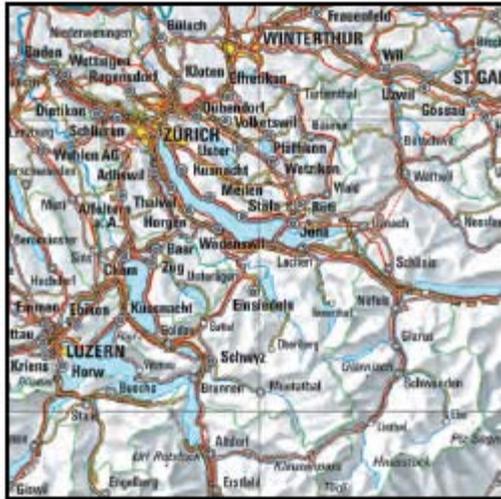
- **一般的参照(地理的な)地図**は、空間を方向付けることを助け、湖、川、山、海岸線、道路などの様々な地形の場所を示す。そのような地図は、ユーザーが地理的エリアの状況を識別する手助けになる。

¹² <http://www.census.gov/population/www/cen2000/atlas/index.html>

- **主題図(統計地図)**は、一つ又はそれ以上の統計的属性の空間的な分布を示すために使われる。主題図は、常に、目的にかなない、政治的、社会的、文化的、経済的、農業あるいは自然に係る現象についての具体的な質問に答えることに役立つように設計される。

一般的参照地図の例

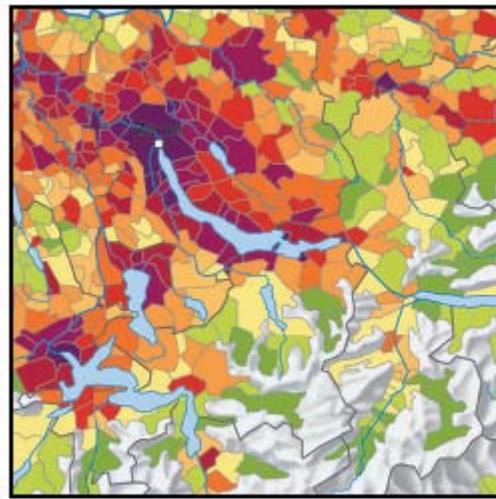
Example of a general reference map



Source: swisstopo, the Federal Geo-Information center¹³

主題図の例

Example of a thematic map



Source: Swiss Federal Statistical Office¹⁴

静的又は双方向型の地図

地図は静的にも、双方向的なものにもなり得る。静的な地図はユーザーが編集することはできない。双方向的な地図には柔軟性があり、ユーザーはデザインを変更したり、データを選択したり引き出したり、地図を動くようにしたり、テーマを変えたり、あるいは主要な関心となっている側面に焦点を当てることができる。

地図はイメージかデータをベースとすることができる。地図がイメージだけをベースにしているなら、それは前もって作り出されて、画面上では静止している。データをベースにした地図を作成するときには、地図を作成するために必要とされるすべての情報(データ及びメタデータ)がデータベースに格納される。〈データをベースにした〉地図そのものはユーザーがインターネット上でそれを求めるときに作成されるだけである。この技術によって、ユーザーは地図を再作成することなく、容易にデータと地図のパラメータを更新・変更することができる。

主題図帳

地図帳は集約された地図が装丁された最も単純な形式である。主題図帳は、包括的な方法で統計を提供し、文章、グラフ、統計表の形で貴重な情報を豊富に掲載している。ほとんどの統計組織はそのような地図帳によりデータを伝える可能性を認識し、人気の高いセンサス地図帳あるいは主題(例:人口、健康、経済)を表す地図帳を作成している。

地図帳に係る技術は最近10年の間に著しく改良された。最新のオンライン地図帳情報システム(AIS)では、ユーザーが地図の背後にあるデータを探し、地域をクリックし、自分にピッタリの地図を作り、自分に必要なデータをまとめ、地図を作成した著者又は事務所に連絡することもできる。地図製作、グラフィックデザイン、データ分析、著述・翻訳など様々なタイプの専門知識の統合を容易にする新たな生産方法が姿を現しつつある。

¹³ <http://www.swisstopo.admin.ch/internet/swisstopo/en/home.html>

¹⁴ http://www.bfs.admin.ch/bfs/portal/en/index/regionen/thematische_karten/maps.html

オンライン上の主題図帳の良い例 GOOD EXAMPLE of an online thematic atlas



Source: International Monetary Fund, IMF Data Mapper¹⁵.

5.5 適切なタイプの地図を選択すること

グラフのためのアドバイスと同じことが地図にも当てはまる。どんな種類の情報でどんな種類の地図を作成すべきかを知ることが重要である。どのような地図作成技術を選択することが適当なのかはデータの性格による。この章では、最も一般に利用される三つのタイプの主題図、すなわち階級区分図(choropleth maps)、ドットマップ(dot maps)、比例記号地図(propotional symbol maps)についてのガイドラインを示す。

階級区分図 (Choropleth maps)

最も一般的なタイプの地図は階級区分図で、表示されている変数の値に比例して各区域が段階的に色分けされている。下の例のように、この種の地図は、全体を見渡した形でパターンを視覚化する簡単な方法を提供する。

色分け向けのデータが豊富になり、GISを使って容易にデザインできるようになったことで、残念なことに、階級区分図のテクニックがしばしば誤って使われることになった。比率(すなわち割合、率又は密度)だけがこのテクニックで地図化できる。人口規模のような絶対値を表すために階級区分図のテクニックを使うべきではない。

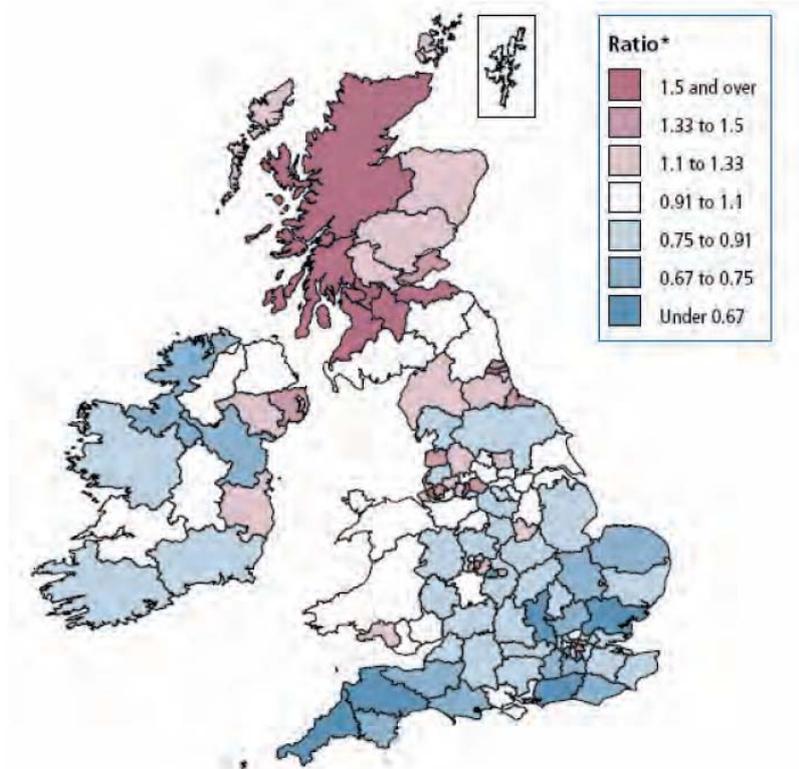
¹⁵ <http://www.imf.org/external/datamapper/index.php>

階級区分図の良い例

咽喉ガン：保健機関<の管轄区域>別・男性・英国及びアイルランド・1991年～1999年

GOOD EXAMPLE of a choropleth map

Cancer of the larynx: incidence by health authority
Males, UK and Ireland, 1991-99



* Ratio of directly age-standardised rate in health authority to UK and Ireland average

Source: Office for National Statistics (2005), *Cancer Atlas of the United Kingdom and Ireland 1991-2000*, London¹⁶.

階級区分図は、上記の例における英国とアイルランド保健機関のように、前もって定められた区域に関して集計されたデータに基づいている。このタイプの地図は、これらの保健機関の管轄地域に関する議論に適切である。しかし、現実世界のパターンはしばしば前もって定められた区域どおりに形作られるわけではなく、大きな誤解につながり得る。したがって、階級区分図は、望ましくはそれぞれの区切られた地域内では均等に分布する現象を示すために使われるべきである。

データ分類は色分けテクニックの大きな問題である。地図に表示されるパターンはデータの値をどのようにグルーピングするかによって決定される。同じデータからであっても、異なった分類法を適用し、データを異なった階級の数に分けるか、あるいは階級の範囲を変えることで、同じデータから異なる地図を作成することができる。多くのテクニックがあるが、決まりはないので、データを分類する最も適切な方法を決定しなければならない。

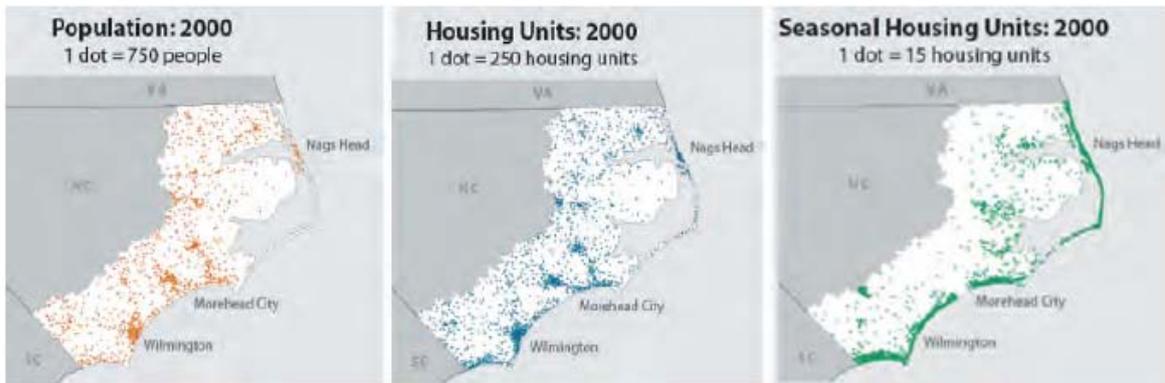
ドットマップ (Dot maps)

ドットマップ(あるいはポイントマップ)は、記号を使って人口あるいは現象の所在及び密度を示す。ユーザーは、データのおよその大きさ、集中・分散を素早く把握することができる。一つひとつの点はその値を示し、大きい数字を表す場合、以下の例のように表すことが多い。

¹⁶ <http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14059>

ドットマップの良い例

GOOD EXAMPLE of dot maps



Source: U.S. Census Bureau, *Census Data and Emergency Preparedness*¹⁷.

左から 人口:2000年 住宅戸数:2000年 季節的住戸数:2000年
 点1個=750人 点1個=250戸 点1個=15戸

比例記号地図 (Proportional symbol maps)

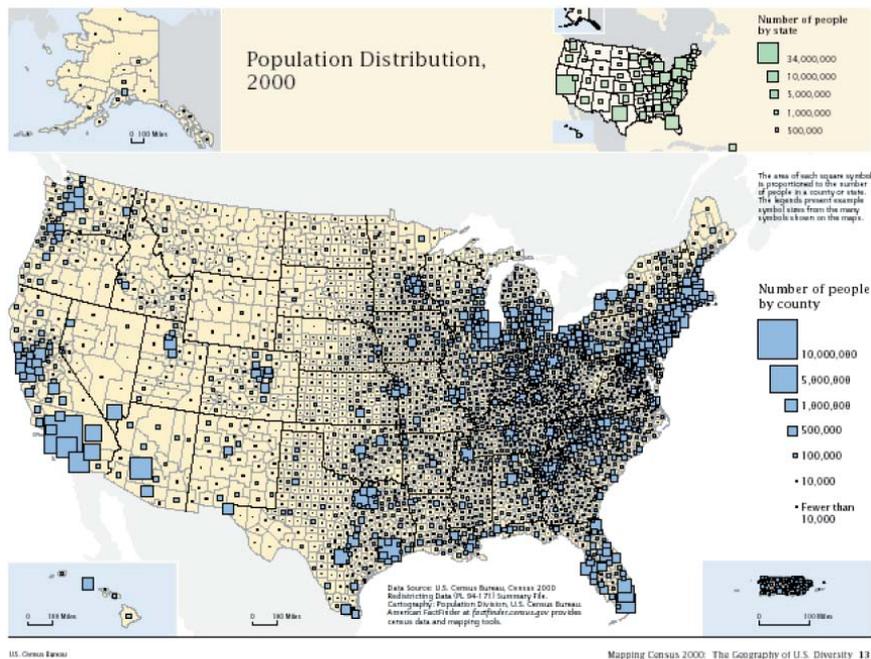
比例的(又は段階的)記号地図は絶対値を表すために用いられる。記号の大きさは人口や表される現象の大きさに比例している。それぞれの記号は、通常はそれぞれの地域単位の中央又は中心都市のある場所に置かれる。

円はコンパクトで縮小・拡大が容易なため、最も一般に使われる。しかし、下の例のように、正方形又は三角形のような他の幾何学図形を使うこともできる。

比例記号地図の良い例

人口分布 2000年

GOOD EXAMPLE of a proportional symbol map



Source: Brewer, C.A. and Suchan, T.A., U.S. Census Bureau (2001), *Mapping Census 2000: The Geography of U.S. Diversity*, U.S. Government Printing Office, Washington DC¹⁸.

17 <http://www.census.gov/Press-Release/www/emergencies/>

18 <http://www.census.gov/population/www/cen2000/atlas/index.html>

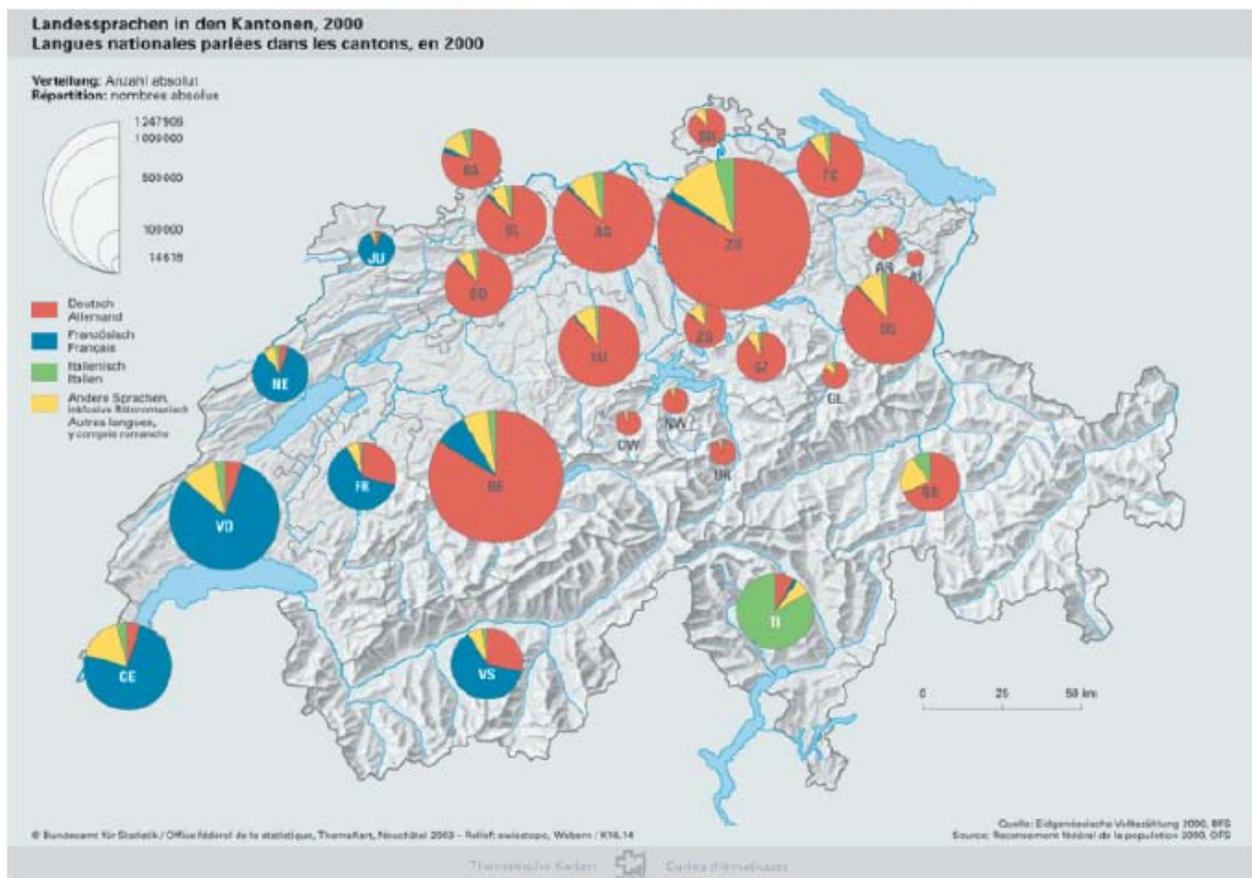
上の例は、よくデザインされた比例記号地図を作成するには、専門的な地図作成技術が必要とされることを示している。記号のサイズがそれに対応する単位となる区域のサイズより大きいと、記号が表している区域を特定することは難しいであろう。また、この例に示されているように、多くの記号が重なり合うと、それを解釈することは一層困難になる。

同じ地図上に人口の様々なカテゴリーの分布を表すために、グラフと地図作成技術を組み合わせることができる。複合的な記号地図においては、円グラフ又は棒グラフが記号として用いられる。下のスイスの地図では、円グラフは主な言語(ドイツ語、フランス語、イタリア語)又は他の言語を話している住民のシェアを26州について表している。記号の大きさはそれぞれの州の総人口を表している。

複合記号地図の良い例

〈スイス〉各州における会話用言語 2000年

GOOD EXAMPLE of a complex symbol map



Source: Swiss Federal Statistical Office, *Map Gallery Switzerland – Languages and religions*¹⁹.

地図と凡例は過度に盛り込まれやすいので、この複合技術は大きな注意を払いながら用いるべきである。そのような地図は国レベル(例:ヨーロッパの国々)あるいは地域レベル(例:上の地図に示されるスイス各州)においてのみ作成することができる。さらに、少ない数(5以下)のカテゴリーだけが円グラフ又は棒グラフで表すカテゴリーは少ないほうがよい(5以下)。

¹⁹ http://www.bfs.admin.ch/bfs/portal/en/index/regionen/thematische_karten/maps.html

5.6 デザインのコツ:単純にしておく!

あらゆるデータ視覚化に関して、メッセージを確実に理解してもらうための最も重要な提案は「地図を単純にしておく!」ということである。あまりに多くの情報を出したり、見た目にごちゃごちゃさせたりすることで、読者の注意をそらしてしまったり、混乱させたりしないように気をつけよう。

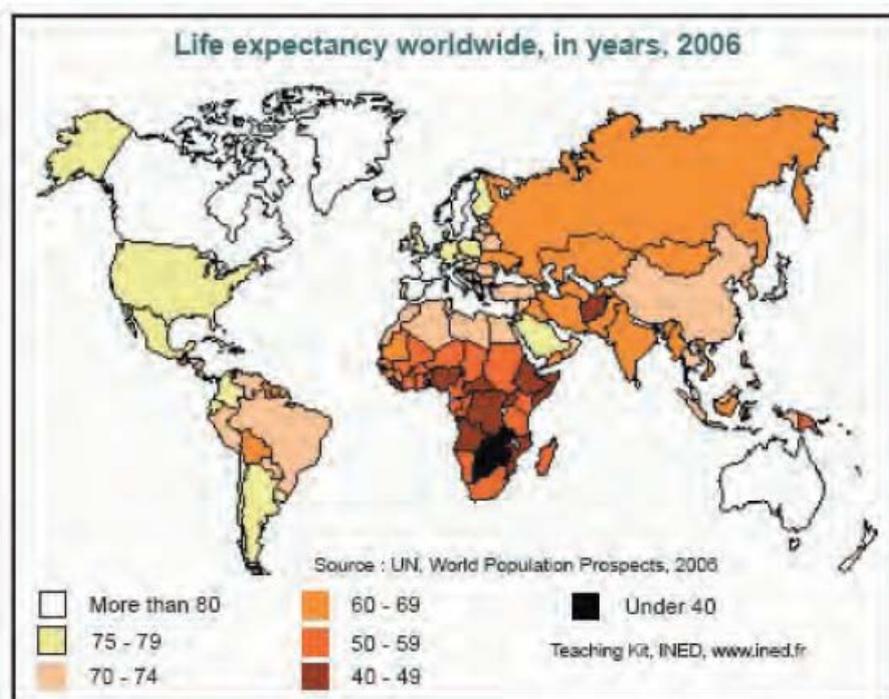
読者を知ることも極めて重要である。読者のバックグラウンドは何か? 読者は日常生活の中で地図に親しんでいるか? <統計組織が示そうとしている> 地図は何らかの意味で読者に不快感を与えるのでは? 対象とする読者の感じ方を意識すべきである。地図は強い視覚的インパクトを持ち、<地図の中で使われている> 色や記号はある人にとって否定的な意味を持っているかもしれない。

地図は、文章又は統計表から独立するように設計すべきである。グラフと全く同じように、地図は、周囲にある文章や脚注を参照しなくとも、それ自身だけで理解されるべきである。一たび地図が出版されれば、それはスキャンされ、あるいはダウンロードされて、<もともとそれが使われていた> 文脈を離れて使われることがある。したがって、以下に記述される様々な地図の構成要素を含んでいることが大切である。

更に説明・参照する必要のない地図の良い例

世界各国の平均寿命(歳)2006年

GOOD EXAMPLE of a map that does not require any further reference



Source: Institut National d'Etudes Démographiques (INED)

地図の構成要素

地図の構成要素は互いに読者の注目を競い争う。メッセージの効果を最大にするため、特に他の情報(水面、高度等)もく地図上に表されるときには、レイアウトしたものの焦点がデータに当たっていることを確かめよう。地図はプレゼンテーションに使うスペースの80-85%を占めるべきである。

以下の構成要素は、地図を理解する上でユーザーの手助けとなる:

- **地図の表題**は、その地図が何についての地図であるか明確な意図を与えるべきである。表題は短く、簡明である必要がある。一層詳細な情報(例:測定単位)を提供するために、副題が加えられることもあろう。
- **凡例**は、その地図上のデータを表すために使われるすべての記号、パターン、色を示すべきである。
- 地図中でデータが表される**地理的単位**は、表題(又は副題)あるいは凡例の中のいずれかで特定されるべきである。
- **テキストラベル**は、重要又は関連のある地名あるいは他の情報を示すため、地図の上に加えられる。
- **地図の縮尺**は、ユーザーが距離を測ったり、異なる地図を比較したりすることを助けるために提供される。
- **脚注**は、定義又は他の方法論についての情報を提供するために使われる。
- <統計データに関する> **資料出所**は地図の一番下に明記されるべきである。
- <統計データを含めてデザインした地図全体に関する> **著作権情報**は、地図の一番下でその内容に関して責任を有する著者を特定すべきである。

以下の他の構成要素が含まれる地図があるかもしれないが、それらは必ずしも不可欠ではない。:

- 地図<の上>が北になっていないときのみ、**北を示すための矢印**が必要になる。
- **緯度と経度**は世界地図又は大陸地図にのみ必要とされる。
- **位置マップ(location map)**は、その地図に表された区域がより広い範囲で見たときにどこであるかを示す基本地図の小さな複製である。読者がその区域の地理に精通していないなら有用であろう。
- **グラフ**が地図の理解を高めるなら、それも加えられる。

表題と凡例にはキーワードだけを使おう

地図は視覚に訴えるメッセージを表現するが、地図に関連づけられた文章も同じように重要である。地図の表題と凡例の言葉遣いは、ユーザーの理解と解釈を決めることになるため、慎重に考えなければならない。地図における文章の用法について、いくつかガイドラインがある:

- 正確に、しかし単純にしておこう。
- キーワードのみを使い、表題、凡例あるいは脚注で同じ言葉を繰り返すのを避けよう。
- 中立的な言葉を使おう。
- 省略語と頭字語<訳注:各語の頭字をつづりあわせて作った語で、一つの単語として発音されるもの>を避けよう。
- 凡例に使うフォントは表題より小さくし、脚注にはさらに小さいフォントを使おう。しかし、すべての文字列が読みやすいものになるように。

慎重に凡例をデザインすること

凡例のデザインは、地図の正しい理解と解釈を確かなものとするものでなくてはならない。それぞれのタイプの地図が異なるタイプの凡例を必要とする。しかし、階級区分図や比例記号地図についてはいくつかの基本的なルールがある:

- すべての階級の範囲は明確であるべきである:100-200、200-300、300-400のような範囲を設定することは避けよう。<訳注:この場合、境界値「200」、「300」がどの階級に属するのか明確でない>
- 階級間にギャップ<訳注:上下どちらの階級にも含まれない範囲>を作らない:1.0-1.5、2.0-2.5、3.0-3.5のような範囲を設定することを避けよう。
- データが利用可能でない(存在しない)区域を明示すべきである。
- 複数の変数を表している地図では、凡例は重要な変数の重要性に基づく降順で表示されるべきである。

色：もう一つの重要な選択

色は最も強力な特徴の一つである。色は<読者に>影響を与え、誤った方向に導くこともできるので、よく注意した上で地図の中で使う色を選択しよう。

テーマ地図における色の選択はデータと地図のタイプに左右される。しかし、他の三つの側面も考慮されるべきである。まず、ある色に結びついている決まり事<訳注：地図上の青は一般に水面を意味することが前述されている>や、<ある色に結びつき得る>肯定的又は否定的な意味合いに気を付けるべきである。

次に、地図に使われている色で誰もがメッセージを理解できるよう確認すべきである。例えば、色覚障がい者の人々はいくつかの色を識別することが苦手である。最も一般的なものは赤と緑の色覚障害である。成長と下落のような二つのタイプの区域間の違いを示すために赤と緑を使うと、色覚障がい者の人々には違いが見えないであろう。簡単な解決方法がある：赤を紫に置き換えれば、これらの人々にも違いがわかる。

最後に、連続的な尺度(例：人口密度)で階級の数が比較的少ないときには、様々な色を使うより同じ色で濃淡を変えることを考えるべきである。離散した階級又は正負の値を持つデータあるなら、多くの色を使うことは<同じ色の濃淡を使うよりも>より適切である。

6. 新たな視覚化テクニック

6.1 視覚化がただの画像以上である理由

新たに登場してきたツールとテクニックは、データを視覚化し、ユーザーにとってデータをより興味深いものにする新しい機会を提供している。動きのある統計表、グラフ、地図を作成できるプログラムによって、ユーザーはデータを操り、自らデータを視覚化したものを作り出すことができる。アニメーションとビデオは、あたかもテレビのように魅力的なフォーマットである。時系列変化を表すのに役立つ、数値の背後にある意味を説明するための話し言葉や文章による描写も含んでいる。スパークラインやタグクラウド(6.5参照) < 訳注: 原文では「6.4」となっているが、誤りと思われる。 > のような新しいタイプの視覚化も登場してきており、従来の方法にとって代わり得る情報表示の手法となっている。

最新のウェブ技術とそれがユーザーコミュニティの中に作り上げる期待は、統計組織が統計を普及する方法を変えつつある。インターネットは、今やユーザーが自らデータを視覚化したものを共有し、< 視覚化されたデータから > 発見したものを議論できる双方向の通信フォーラムである。Many Eyes²⁰、Swivel²¹、Data Place²²のようなウェブサイトは、データとグラフィックスを論じ、共有するオンラインコミュニティの例である。

新しい視覚化テクニックとより双方向的なウェブサイトの発展は、ユーザー側の対応の幅を広げる一方、統計組織にとって問題を引き起こすこともある。偶然か意図的にかかわらず、ユーザーが統計をゆがめたり、誤って伝えたりして、他に広めてしまうことがますます簡単になっている。それゆえ、統計組織は新しい視覚化テクニックをどのように使い、提供するかについて明確な方針を持つことが重要である。

この章では、これらの新しい視覚化ツールとテクニックを概観する。

6.2 ダイナミックな視覚化

インターネットの到来とウェブ2.0技術²³の発展に伴い、ユーザーがデータと対話しながら自分の思うように視覚化できるようになった。今や多くの統計組織がウェブサイトを通じてデータベースにアクセスできるようにしており、ユーザーは自身で統計情報を検索・ダウンロードできるようになっている。この機能性は、統計表、グラフ、地図をオンライン上で作成できるようにする一連の視覚化ツールによってますます補完されるようになっており、ユーザーが別のアプリケーションでデータをダウンロードしたり作業したりする必要はない。

ユーザーにこのレベルの自由度を与えることで、それによって引き起こされる事態が懸念される。ユーザーが意味のない図や不適当な相関を引き出してしまいう可能性がある。しかし、ユーザーがデータにアクセスして作業しているということは確かに望ましい。明確・明白な方法でキーになるメタデータを提供したり、経験が十分でないユーザーにサポートを提供したり、誤用をモニターし、修正することによって、起こり得る問題を最小化することができる。

²⁰ Many Eyes は、ユーザーがデータをアップロードして、グラフや他の視覚化したものを作成でき、そこで発見したことを議論できるウェブサイトである (<http://manyeyes.alphaworks.ibm.com/manyeyes/> を参照)。

²¹ Swivel は、Many Eyesに類似の機能を提供する (www.swivel.com を参照)。

²² Data Place は、米国の都市、町、州に関する統計をユーザーに提供するウェブサイトである (www.dataplace.org/ を参照)。

²³ 「ウェブ2.0」は、ユーザーに単にオンライン上で情報にアクセスする以上のことを可能にするインターネット技術の新たな潮流を表す用語である。ユーザーは、ウェブ上のコンテンツに追加し、変更し、内容に影響を与えることができる。例えば、ウィキペディアのようなウィキ < 訳注: 情報の集積に便利なウェブプログラムの形態を表す。一般のウェブサイトと同じように閲覧できるだけでなく、誰もが手軽に内容を追加・編集でき、編集過程の全バージョンが保存される。 >、ブログ、FacebookやLinkedInのようなソーシャルネットワーキングサイトなどがこれらの技術に含まれる。

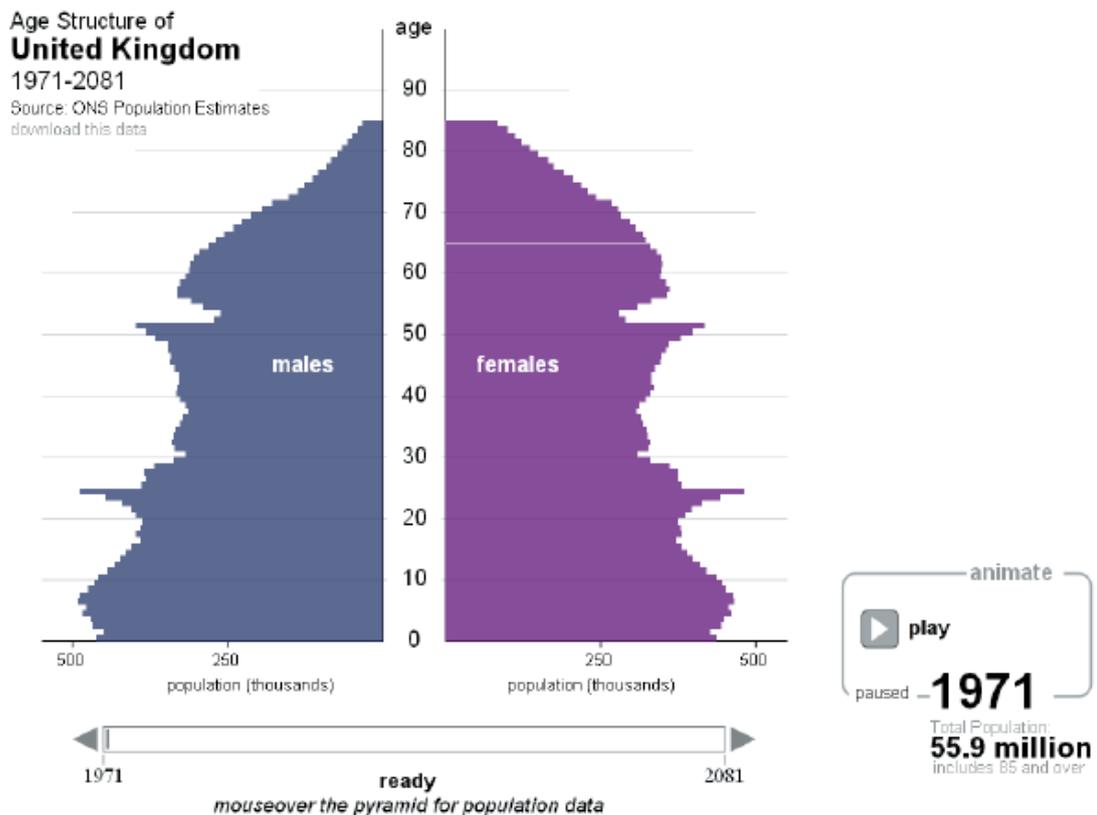
6.3 動画とビデオ

アニメーションとビデオは、姿を現しつつある二つの重要なデータ視覚化テクニックである。テレビや映画の人気を考えると、動画を経由してメッセージを受け取るというアイデアをユーザーが好むことは驚きではない。この方式は、数字の背後にある意味を伝えるために音声又は文章による描写と図形とを組み合わせることで物語を語ることを容易にする。

統計を伝えるのにアニメーションを使っている良い例

英国における人口年齢構成 1971年～2081年

GOOD EXAMPLE of using animation to communicate statistics



Source: Office for National Statistics, United Kingdom²⁴

< 訳注: ウェブ上ではグラフ右側の「play」ボタンで人口ピラミッドが形を変えていく様子を見ることができる。 >

英国国家統計局やカナダ統計局などいくつかの統計組織によって開発された動きのある人口ピラミッドは、単純なインタフェースでアニメーションと双方向性とを組み合わせた良い例である。ユーザーが「プレイ」をクリックすると、人口ピラミッドの形が長い期間にわたってどのように変化するか見ることができる。ユーザーは、年齢層を選択して、総人口に占める詳細な数と割合を知ること、そのグラフとやりとりすることができる。

Gapminder²⁵の共同創設者Hans Roslingは、統計を伝える方法としてデータを表すアニメーションを使用することで大きな成功を収めた。Roslingは、インターネット上でますます人気の高まっている機能であるオンラインビデオを

²⁴ http://www.statistics.gov.uk/populationestimates/svg_pyramid/uk/index.html

²⁵ <http://www.gapminder.org/>

通じて多くの視聴者を獲得した。2006年のTED〈技術・エンターテインメント・デザイン〉会議²⁶における彼のプレゼンテーションは、何千回もオンラインで視聴・ダウンロードされた。彼は、妊婦死亡率、グローバルゼーション、エネルギー、人間開発の傾向のような課題について短いビデオ講義「gapcasts」を開発することで、統計を伝えるこの媒体の人気を作り上げた。

統計を伝えるのにアニメーションと新しいウェブ技術を組み合わせた良い例

世界を変えた200年

GOOD EXAMPLE of combining animation and new web technologies to communicate statistics



Source: Gapminder²⁷

〈訳注: Gapminderのサイト上に置かれている動画の画面(円の大きさは各国の人口を表し、縦軸に各国の平均寿命、横軸に国民一人当たりGDPを示している。動画をスタートさせると、時代を経るにつれ、各国を示す円が大きくなりながら全体に右上に移動していく様子が観察できる。)〉

6.4 ウェブ 2.0 とデータを取り巻くコミュニティの形成

Many Eyes, Swivel, Data Place のようなウェブサイトは、データの視覚化と共有の周辺にオンラインコミュニティを築き上げることで視覚的表現に新次元をもたらしている。これらのウェブサイトは、ユーザーがデータセットをアップロードし、他のユーザーと議論・共有するためのグラフィックスを作ることを可能にしている。しばしば「mashups」〈訳注:「すりつぶす」の意味〉と呼ばれる他のアプリケーションは、二つ以上の素材からデータや機能を組み合わせて新し

²⁶ TED は、技術(technology)、エンターテインメント(entertainment)、デザイン(design)を表す。毎年一度開かれるこの会議には、これらの分野で先端を行く人々が集い、最新の話題について話し合う。プレゼンテーションは会議のウェブサイトで公表される。www.ted.com

²⁷ <http://www.gapminder.org/videos/gap-cast/>

いサービスを作り出す。mashup の一つの例は、統計データと地図データを組み合わせ、データの地理的な見え方を作り出すことである²⁸。

より広範な読者と意思疎通するため、これらのウェブサービスの可能性を試している公的統計の作成者がいる。例えば、UNECE(国連欧州経済委員会)は、様々な結果と組み合わせ、四つのデータセットをSwivelにアップロードした。肯定的な面では、これらのデータセットは当初18カ月間にそれぞれ5000回以上閲覧され、データが多くのユーザーに届いたことを示した。しかし、それらのデータセットへのユーザーのコメントは二つだけで、UNECEの統計データベースウェブサイトへのリンクはこの期間に10回たどられたただけであったので、これら新しいユーザーに関する情報は事実上得られなかった。

より広範な読者に手を伸ばすための新しいウェブサイトを使っている良い例

GOOD EXAMPLE of using the new Web services to reach wider audiences

Data	Description	Created	Views
GDP per capita at current prices and PPPs, US\$	Gross Domestic Product per Capita estimates are compiled according to the 1993 version of the System of National Accounts (SNA 1993) for all countries, except Turkey for which the 1988 SNA is still used. Common currency (US\$) estimates are computed...	about 1 year ago by UNECE	6735
Male Life Expectancy at Birth by Country and Year	Life expectancy is the average number of years of life remaining to be lived by a female or male at birth, based on current age-specific mortality rates. Footnotes: Azerbaijan - 1980: data refer to 1981 Belarus - 1980: data refer to 1981 Bosnia and...	about 1 year ago by UNECE	4358
Female Life Expectancy at Birth by Country and Year	Life expectancy is the average number of years of life remaining to be lived by a female or male at birth, based on current age-specific mortality rates. Footnotes: Azerbaijan - 1980: data refer to 1981 Belarus - 1980: data refer to 1981 Bosnia and...	about 1 year ago by UNECE	4271
Industrial Production Index (2007=100) by Year** Updated**	An industrial production index is an index covering production in mining, manufacturing and public utilities (electricity, gas and water), but excluding construction. The exact coverage, the weighting system and the methods of calculation vary from...	about 1 year ago by UNECE	3709

Showing: 1 - 4 of 4 data sets

Source: Swivel²⁹

この種のオンラインコミュニティは、これまでのところ成功の程度は様々であるが、新たなユーザーに手を伸ばせる比較的簡単な方法を提供していることは明らかで、その動向を注目すべきデータの視覚化に関する新たな分野となっている。³⁰

6.5 他の新しい視覚化テクニック

スパークライン

スパークラインとは、時間的な傾向を示す小さな、単語サイズの折れ線グラフである。スパークラインはひと目で非常に多くの情報を示せる利点があり、その意味を説明する言葉と並べて置くことができる。

²⁸ 例えば、統計提供ソフトウェアである PC-Axis の最新のバージョン(<http://www.pc-axis.scb.se/>) はデータをグーグル地図とグーグル・アース両方と組み合わせるためのオプションを提供している。

²⁹ <http://www.swivel.com/users/show/1005968>

³⁰ より詳しくは国際官庁統計協会(IAOS)統計ジャーナルの特集 Vol.25, No.3-4,2008「ウェブ2.0と公的統計」を参照のこと。

以下のアドレスで閲覧できる。

<http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>

Tufte(2006年)が最初にスパークラインを提案した。次の例は、他の通貨に対するユーロ為替相場の変動を示すスパークラインの例である。このような、“力強く、簡単で、単語サイズのグラフ”は、多くのスペースを占めることなく視覚的表現で統計表のデータを強調する。

スパークラインの良い例

GOOD EXAMPLE of a sparkline



Source: Tufte, E.R. (2006), *Beautiful Evidence*, Cheshire CT, Graphics Press.

タグクラウド

タグクラウド(しばしば「ワードクラウド」とも呼ばれる)は、ある文章又はデータセットにおける単語又はタグの頻度についての視覚的表現である。<文字フォントの>様々なサイズと色でタグの人気又は重要性を示している。タグクラウドはウェブサイト上でよく見られ、それぞれの言葉から関係する情報をたどれるリンクの張られたカテゴリーのリストになっている。

タグクラウドは、文章中でよく使われる用語を特定し、キーワードの分類を作り上げる便利な方法である。下の例は、この章の文章から作られたもので、キーワードを鮮やかに示している。

タグクラウドの良い例

GOOD EXAMPLE of a tag cloud



Tag Cloud

2006 allow animation clouds **communicate** create **data** development discuss download
dynamic emerging examples gapminder good graph http illustrating information interactive map
meaning new numbers online organizations own place popularity providing sets share sites
source sparklines **statistical** such swivel **tag** techniques time uk **users** using
video **visualization** ways web words www

Created using Tag Cloud Builder³¹

³¹ <http://tagcloud.oclc.org/tagcloud/TagCloudDemo>

7. アクセシビリティの問題

基本的な統計情報が広く簡単に使われるためには、誰にとっても利用しやすいものでなくてはならない。これは、ユーザーが使用するテクノロジーやユーザーが持っているかもしれない障害にかかわらず、誰もが情報を調べ、理解できなければならないことを意味する。したがって、コミュニケーション戦略における重要な目標はアクセス可能な情報を作り出すことである。

大勢の読者を引き付けるために、既に確立されて広く認められている基準に適合する、体系的でよく整理された方法で情報が示されるべきである。この章では、部分的に重なり合っているアクセシビリティの三つの側面を検討する：

- ユーザーがデータを理解するのを助ける適切なメタデータを提供すること；
- 携帯電話あるいは他の携帯端末のような新しいメディア経由のものを含め、多様なフォーマットでデータを提供すること；
- しばしば特定の法的又は政策的な要請に従うように、障がい者にとっての情報へのアクセシビリティを最大限保証すること。

統計情報を誰にも利用しやすくするには、特別の注意深さ、特別な作業及び特別なリソースを必要とする。しかし、基本的な統計情報に平等にアクセスできるという原則は重要である。綿密なアクセシビリティ戦略は、すべての人に役立つであろう。

この章では、統計情報を伝えるとき考慮されるべき、優れた情報に関する主な要素を検討する。

7.1 文章

ユーザーが探している情報を容易に見つけられるように、文章は簡潔で、筋が通っていて、うまく構成されていなければならない。見出しと小見出しを使って文章を明瞭かつ論理的な部分に分けて示すことで、探しやすく、また、他のフォーマットに変換しやすくなる。〈訳注：後述の通り、抑制出版物とHTML文書の間の変換などが想定。〉

すべてのターゲットグループに効果的に役立つものにするため、文章は、点字、音声又は特大文字などいくつかのフォーマットで利用可能とすべきである。オリジナルの文章を作るときには、フォーマットを変換することによる影響を最小にするため、考えられるフォーマット変換を見込んでおくべきである。例えば、それぞれの画像情報にそれを説明する文字情報を付加することは、点字、音声はもとよりウェブサイト用のHTML³²のフォーマットに変換する作業さえ減らす。

W3C(ワールド・ワイド・ウェブ・コンソーシアム)³³によって作成されたアクセシビリティの規則は、適応技術を使った電子的フォーマットの情報へのアクセスを保証するように設計されたものである。この規則には、例えば、文字読上げソフト、文字拡大システムとPDA装置(携帯型データアクセス装置)が含まれているが、それだけではない。

7.2 表

データを示す際に統計表を使用するときも、アクセシビリティに注意を払うべきである。データは、それを理解し解釈するために必要なすべての情報がないまま示されると、使うことができなかつたり、誤解を与えるものになる。

³² ハイパーテキストマークアップ言語。

³³ <http://www.w3.org/>

例えば、複数ページにまたがっている一つの表を考えてみよう。表頭又は表側が見えなくなってしまうと、ユーザーは表を読み取ることが難しくなる。そのため、それぞれページで表頭と表側を繰り返すことが重要である。

アクセシビリティの要請に適合した統計表の良い例

カナダにおける家畜頭数 2007年及び2008年

GOOD EXAMPLE of a table that meets accessibility requirements

Livestock inventory in Canada, 2007 and 2008

Type of livestock	2007		2008	
	Thousands of head	%	Thousands of head	%
Cattle	15 885	50.2	15 195	52.0
Hogs	14 690	46.4	12 985	44.4
Sheep	1 096	3.5	1 062	3.6
TOTAL	31 671	100	29 242	100

Source: Canadian Cattle Statistics August 2008³⁴

それぞれの値は単に家畜の種類や年に関連づけられるだけでなく、頭数(千単位)又は割合にも言及している。文字読上げソフトは、表中のそれぞれ数値についてすべての情報を与えることができなければいけない。ユーザーには、したがって、「2007年、牛の割合は50.2%である」と聞こえるはずである。

HTML のような電子形式では、必要なら、内容を表す短い記述とともに表頭と表側のためのタグを含めることができる。タグを使うことは誰にとっても良いことである。表がダウンロードされる時、混乱を招くことなく、表頭と表側は直接転送される。さらに、表の構造が複雑なときは、ユーザーが最小限の努力で最大の情報を入手することができるよう、データの整理方法を正確に記述することが重要である。

7.3 グラフ

グラフを作成するときは、すべてのユーザーが画像を使えるわけではないと考えてみるのがよい。グラフを説明する文字情報を付け加えることで、グラフに示されるのと同じ情報を提供することができる。以下は、全米点字協会のテープ録音マニュアルの例である。

ウェブサイトにおいては、「代替テキスト」タグ(短い等価テキスト)や、「longdesc」属性(異なるページに蓄積され、eメールアドレスによってアクセスできる長い記述)を使うことがW3Cによって認められ、記述されているテクニックである。これらの記述は、音声、点字又は他のフォーマットのドキュメントの作成に当たっても使用される。

手作業で等価テキストを作成するのが困難なグラフは少なくないであろう。グラフから自動的に情報を抽出できるツールを設計するプロジェクトが進行中である。そうしたツールによって、グラフ作成と同時に記述ページを作り出すこともできるようになるであろう。

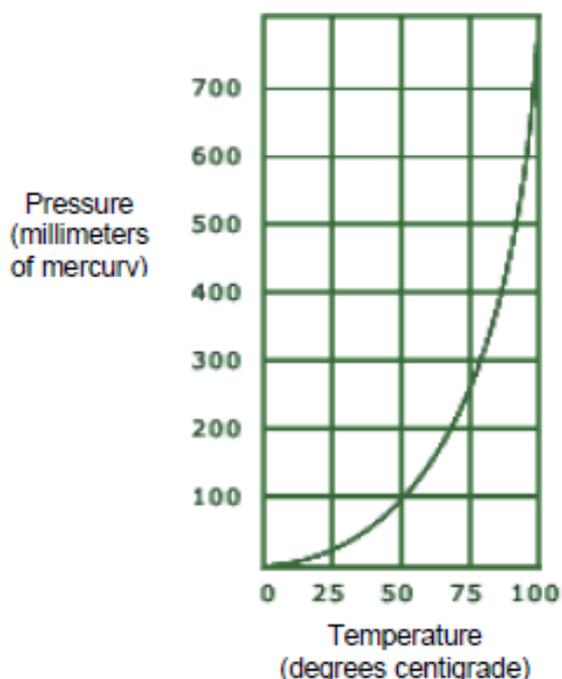
³⁴ <http://www.thedairysite.com/articles/1613/canadian-cattle-statistics-august-2008>

等価テキストを備えたグラフの良い例

水蒸気圧力と温度との関係

GOOD EXAMPLE of a chart with a text equivalent

Relationship between the vapor pressure of water and its temperature



"The relationship between the vapor pressure of water and its temperature." This is a line graph whose x-axis is temperature in degrees centigrade, running from zero to one hundred degrees. The y-axis is pressure in millimeters of mercury and runs from zero to 800 millimeters. The curve starts at the origin and rises so that when x is 25 degrees, y is approximately 40 millimeters. When x is 50, y is 100. When x is 75, y is just under 300. When x is 100, y is about 760."

Source: Accessible Digital Media: Design Guidelines for Electronic Publications, Multimedia and the Web³⁵.

※ 枠内の等価テキスト

「水蒸気圧力と温度の関係」これは線グラフで、x軸は摂氏による気温を表しゼロから百度までである。y軸は水銀計のミリメートルで示された圧力を表しゼロから800ミリまでである。＜グラフの＞曲線は原点からスタートし、xが25度のときyがおよそ40ミリであるように上がっている。xが50のときyは100。xが75のときyは300をわずかに下回る。xが100のときyは約760である。」

7.4 地図

地図は、ある特定の障害を持っている人々にアクセシビリティの点で大きな技術的障壁になる。＜地図上の＞情報は、イメージと色の組み合わせという、アクセシビリティ基準では事実上置き換えられない二つの方法によって伝達される。したがって、地図に表示されているのと同じ情報を提供するため、代替テキストを提供するか、データ表を使えるようにするか、両方又はいずれかを検討すべきである。

³⁵ http://ncam.wgbh.org/publications/adm/guideline_f.html

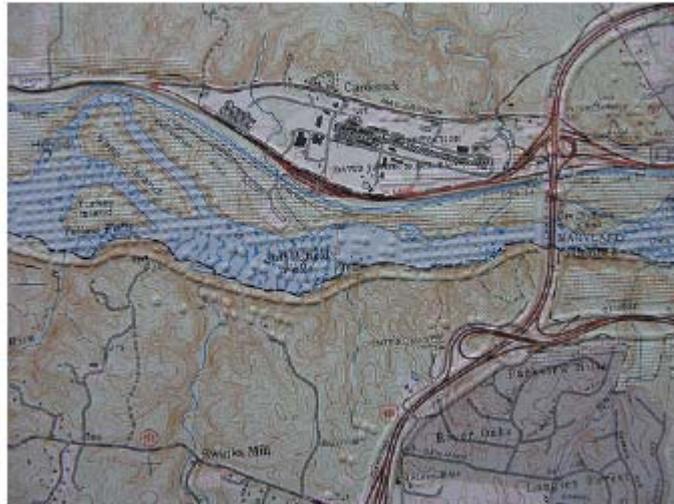
もう一つの選択肢は、ユーザーに地域、町又は街路のようなパラメータを選択させる検索ツールを提供することである。データベースの検索は、このようなものであり得る：「特定のエリアに住み、農産業部門で働いている15歳から49歳の男性の数を検索」。検索結果は、視覚的に見るのと同じ情報を提供する統計表の形式で生成されるべきである。

視覚障がい者にもアクセス可能な地図の良い例

ワシントンDCの一部を表した点字地図

GOOD EXAMPLE of a map accessible to blind people

Braille map of a portion of Washington DC



Source: US Geological Survey Library, courtesy of Flickr³⁶

7.5 メタデータ

統計組織は、そのデータの強みと限界を含め、データを理解するために必要なメタデータが確実にユーザーに提供されるようにすべきである。これらのメタデータは、定義、分類、方法論に関する直近の変更を取り込んで最新化されなければならない。

誰もがメタデータにアクセスし、使うことができるようにするため、以下の提案を考慮すべきである：³⁷

- 印刷出版物、CD-ROMなどのような様々な媒体を経由してメタデータにアクセスできるようにしよう。インターネットはユーザーが最新のメタデータを見つける上で最も利用しやすい媒体であるから、すべてのメタデータはインターネット上で利用できるようにすべきである。
- ユーザーは求めるものも統計に関する専門知識も様々であり、そのような幅広い人々のニーズに応えるような方法でメタデータを示そう。
- たとえ、統計データやそれに関するメタデータの印刷物を有償で提供していたとしても、メタデータはインターネット上では無償で提供しよう。
- メタデータと統計表やグラフへのリンク、その反対方向のリンクをしっかりと張っておこう。
- メタデータはその国の言語だけでなく、できれば英語のような共通言語でも提供しよう。

³⁶ <http://www.flickr.com/photos/98169608@N00/3296197787/>

³⁷ メタデータの報告と普及のためのガイドラインは、2007年のOECD出版「データとメタデータの報告とプレゼンテーションハンドブック」(以下のURLから利用可能)に含まれているものから採用された。<http://www.oecd.org/bookshop?9789264030329>

- 無料のテキスト検索を使ってく統計組織が運営するウェブサイトに絞ったローカルな検索エンジンを提供しよう。
- URL(Uniform Resource Locators)を頻繁に変えるようなことはせず、＜変える場合は＞移転先にユーザーを誘導する新旧URL間のリンクを提供する仕組みを設けよう。ウェブサイト間のリンクの重要性を考えると、これは重要である。
- コンセプト、定義、統計上の方法論について詳しい情報が得られるであろう問い合わせ先担当者の氏名又は電子メールアドレスを提供しよう。組織によっては、「問い合わせ先」が一般的な組織の窓口又はあらゆる顧客を対象とする問い合わせサービスということもあろう。

統計組織の中には、ピラミッドの層のようにメタデータを提示しているところがある。このアプローチでは、ユーザーがピラミッドの上から下りてくるにつれ、メタデータはより詳しいものになる：

- **ピラミッドの頂上：** データの状態(速報か確報かの別)についての情報を含め、その統計を基本的に理解するのに欠かせないメタデータ。一般的なユーザーであれば、この種のメタデータ以上のものが必要になることはほとんどない。
- **＜頂上の＞下部：** データの使用に影響を与え得る統計の短い記述(定義、重要なポイント、制約等)を提供する説明的な解説。＜一般ユーザーと専門家の＞中間的なユーザーがこのタイプのメタデータを参照するであろう。
- **ピラミッドの底部：** 方法論に関する最も詳細な情報。典型的には方法論に関するマニュアル又は類似のドキュメントなど。このタイプのメタデータには、最も経験豊かなユーザーや詳細な研究のためにデータを使っているユーザーだけが関心を寄せる可能性が高い。

8. 参照・参考文献

Bertin, J. (1981), *Graphics and Graphic Information-processing*, New York, de Gruyter.

Bertin, J. (1983), *Semiology of Graphics: Diagrams Networks Maps*, Madison Wis., University of Wisconsin Press.

Blessing, C., Bradsher-Fredrick, H., Miller, H., Miller, R. and Rutchik, R. (2003), *Cognitive Testing of Statistical Graphs: Methodology and Results*, Washington D.C., U.S. Energy Information Administration.

Bosch ten, O. and Jonge de, E. (2008), "Visualising official statistics", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 103-116, available at:

<http://iospress.metapress.com/content/v0376364l348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Cleveland, W.S. and McGill, R. (1984), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods", in *Journal of the American Statistical Association*, 79, pp. 531-554, available at:

<https://secure.cs.uvic.ca/twiki/pub/Research/Chisel/ComputationalAestheticsProject/cleveland.pdf>.

Cleveland, W.S. and McGill, R. (1987) "Graphical Perception: The Visual Decoding of Quantitative Information on Statistical Graphs (with Discussion)", in *Journal of the Royal Statistical Society Series A*, 150, pp. 192-229, available at:

http://www.wjh.harvard.edu/~kwn/Kosslyn_pdfs/1987Cave_JRoyStatSocA_CommentaryClevelandArticle.pdf.

Few, S. (2004), *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Oakland CA, Analytics Press.

Gardner, J. (2008), "Blogs, wikis and official statistics: New perspectives on the use of Web 2.0 by statistical offices", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 81-92, available at:

<http://iospress.metapress.com/content/v0376364l348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Harris, R.L. (2000), *Information Graphics*, New York and Oxford, Oxford University Press.

Kennedy, D. (2007), *Research Paper: Data Visualization*, Canberra, Australian Bureau of Statistics, available at:

[http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/7F8E375FC22D26A5CA25731C0022DF1B/\\$File/1211055001_jul%202007.pdf](http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/7F8E375FC22D26A5CA25731C0022DF1B/$File/1211055001_jul%202007.pdf).

Miller, J.E. (2004), *The Chicago Guide to Writing About Numbers*, Chicago, University of Chicago Press.

Playfair, W. (1786), *The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Exports, Imports, and General Trade of England, at a Single View*, London.

Playfair, W. (1801), *Statistical Breviary: Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe*, London, Wallis.

Robbins, N.B. (2005), *Creating More Effective Graphs*, Hoboken NJ, John Wiley & Sons.

Schulz, T. (2009), *Guidelines on the Presentation of Statistical Maps*, UNECE Work Session on the Communication and Dissemination of Statistics, Warsaw, available at:
<http://www.unece.org/stats/documents/ece/ces/ge.45/2009/crp.1.e.pdf>.

Smith, A. and Rogers, S. (2008), "Web 2.0 and official statistics: The case for a multi-disciplinary approach", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 117-123, available at:
<http://iospress.metapress.com/content/v0376364l348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Tufte, E.R. (2001), *The Visual Display of Quantitative Information (2nd ed.)*, Cheshire CT, Graphics Press.

Tufte, E.R. (2006), *Beautiful Evidence*, Cheshire CT, Graphics Press.

Wainer, H. (1984), "How to Display Data Badly?", in *The American Statistician*, vol. 38, no. 2, pp. 137-147, available at:
http://www.soc.washington.edu/users/bp Pettit/soc504/wainer_display.pdf.

Ware, C. (2004), *Information Visualization: Perception for Design*, San Francisco CA, Morgan Kaufmann Publishers.

Further useful information and discussion on the presentation of statistics can be found at:
<http://blogstats.wordpress.com/>.