

データの視覚表現と理解について

関 友 作*

(2021 年 10 月 22 日受理)

Data Visualization for Better Comprehension

Yusaku SEKI

キーワード: データ, 視覚化, グラフ, わかりやすさ

さまざまなデータを示すときにグラフによる視覚的な表現がよく使われる。これは、データを理解しやすくするために有効である。ただし、その表現の仕方によっては、見る側に誤解をまねいたり、わかりにくくしてしまったりすることもある。

本稿では具体例をもとに、いくつかのグラフ表現について留意すべきことを取り上げ、指針を示した。データの重要性がますます高くなっている現在、その視覚表現についてのリテラシーも必要になってきているであろう。

はじめに

数字が並んだだけの表ではよくわからないことも、グラフや図を見れば一目でわかることは多い。その一方で、見る人にわかりにくかったり、誤解をまねいたりするようなグラフもときにある。

いまでは表計算ソフトなどを使えば、だれでも容易にグラフを作ることができる。しかし、よりわかりやすいグラフの作り方という点については、まだ十分に理解されておらず、あまり教育もされていなのが現状である。

そこで本稿では、グラフを用いてデータを視覚的に表現するときに、受け手が理解しやすく、誤解を生まないために留意すべきことを、実例とともにいくつか取り上げてみたい。

棒グラフ

まず、身近によくあるようなグラフの例から見ていきたい。図 1 は、5 人の生徒の点数を表した棒グラフである。これを見ると、生徒 C の点数がもっとも高く、B の点数がもっとも低いことが見

*茨城大学 全学教育機構

てとれる。また、Bの点数はAの点数の半分であるようにも見えてしまう。

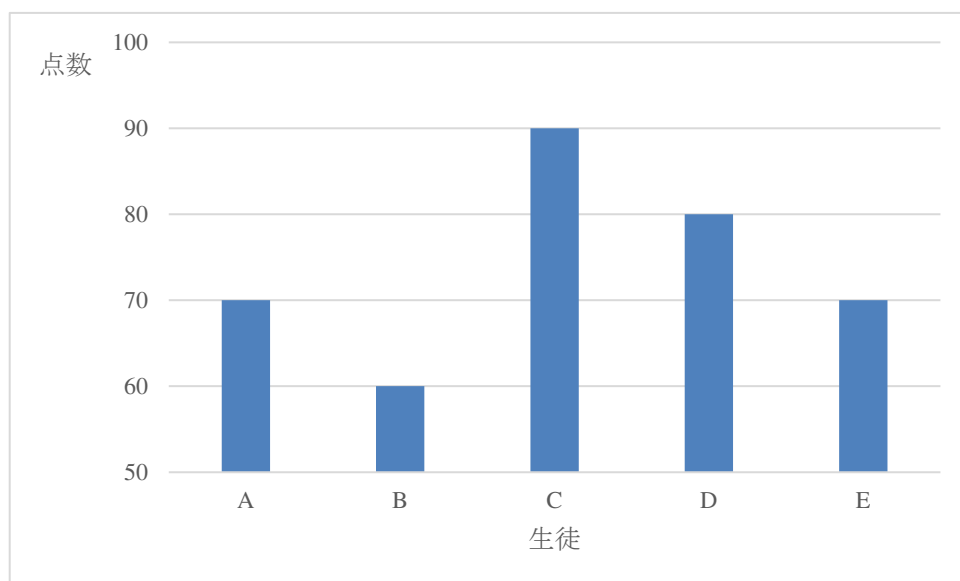


図1 棒グラフ例 A

しかし、縦軸の目盛りを見ると、Aの点数は70点、Bの点数は60点であることがわかる。Bの点数はAの半分ではない。

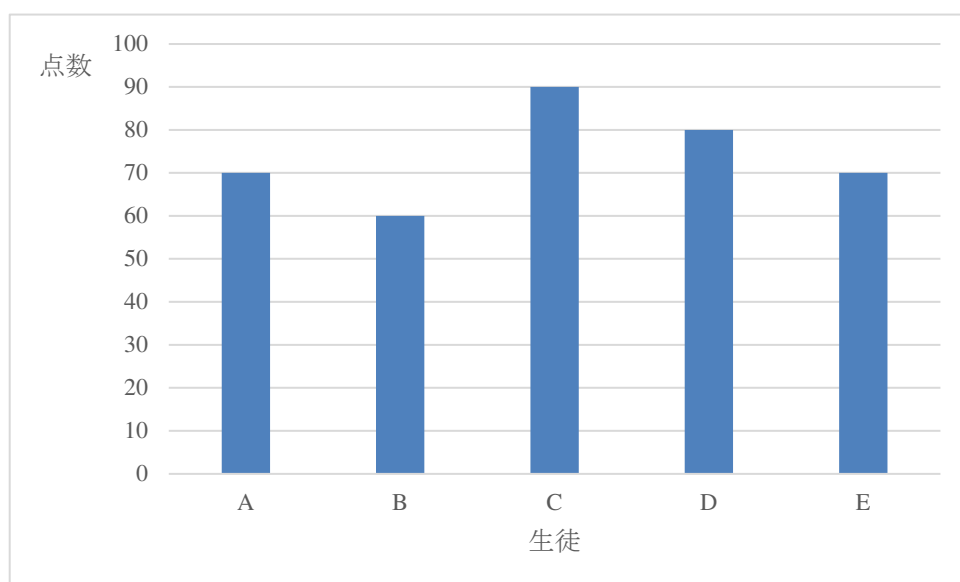


図2 棒グラフ例 B

そこで、図1の縦軸の目盛りを修正して、基点を0にしたものが、図2である。これを見ると、生徒Cの点数がもっとも高く、Bの点数がもっとも低いことは、図1と同様にわかる。ただし、Bの点数はAの点数の半分ではなく、少しの差であることが棒の長さから見てとれる。

このように棒グラフでは、数値軸の目盛りを変えることで棒の長さも変わるため、見る側に誤解を生むおそれがある。したがって、よく指摘されるように、棒グラフの数値軸は0から始めるべきである（阿部、2021；Wong、2010など）。また、棒グラフを見るときにも、この点には注意する必要がある。

折れ線グラフ

図3は、日本の出生率（合計特殊出生率）の変化を示した折れ線グラフである。これは、1950年から2019年までの出生率の変化を示している。折れ線グラフは、こうした時間的な変化を見るのに使われる。

これを見ると、1950年以降1956年ころまで出生率は急に低下し、その後は1966年の落ち込みはあるものの、1974年ころまで横ばいに近い様子が見てとれる（1966年は、丙午による出生率低下である）。その後、出生率は2を下回り、おおむね減少から横ばいの傾向にあることがわかる。

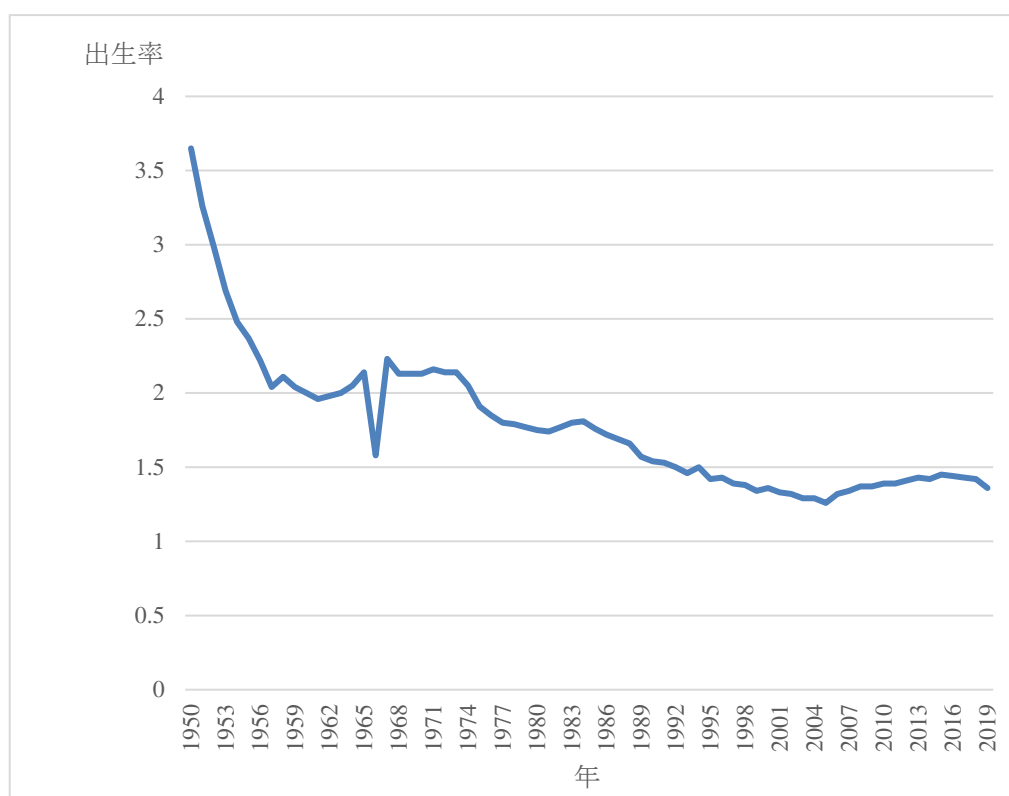


図3 日本の出生率の変化
(厚生労働省の人口動態統計より)

ここから、2000年以降のデータだけを取り出して折れ線グラフにしてみると、図4のようになる。

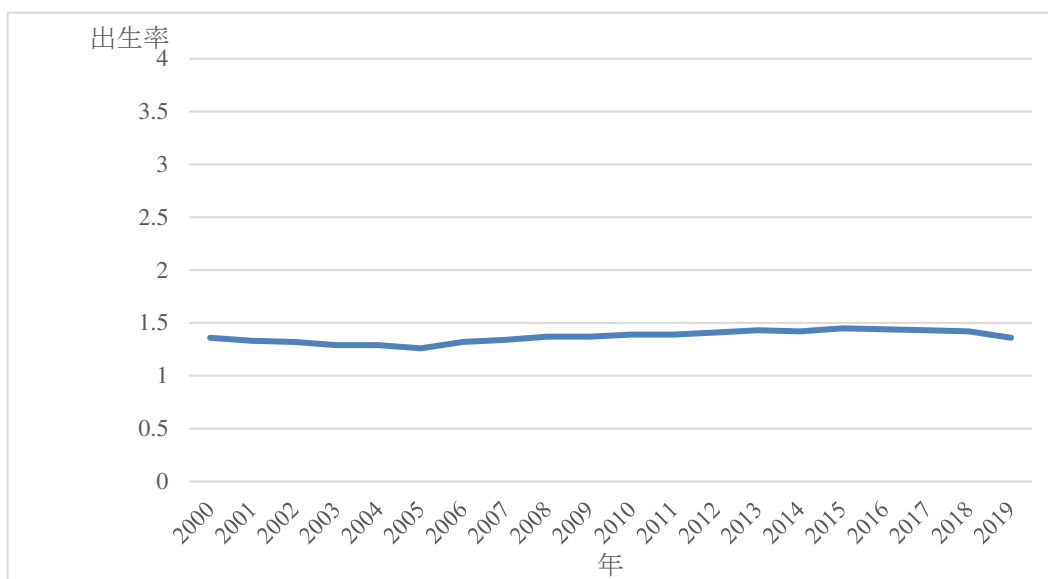


図4 日本の出生率の変化 (2000年以降) A

図4は、数値軸の目盛りを図3と同じにしてある。しかし、これでは年ごとの小さな変化がわからない。そこで、目盛りを変えたのが図5である。また、これには節目の年の出生率も入れてある。

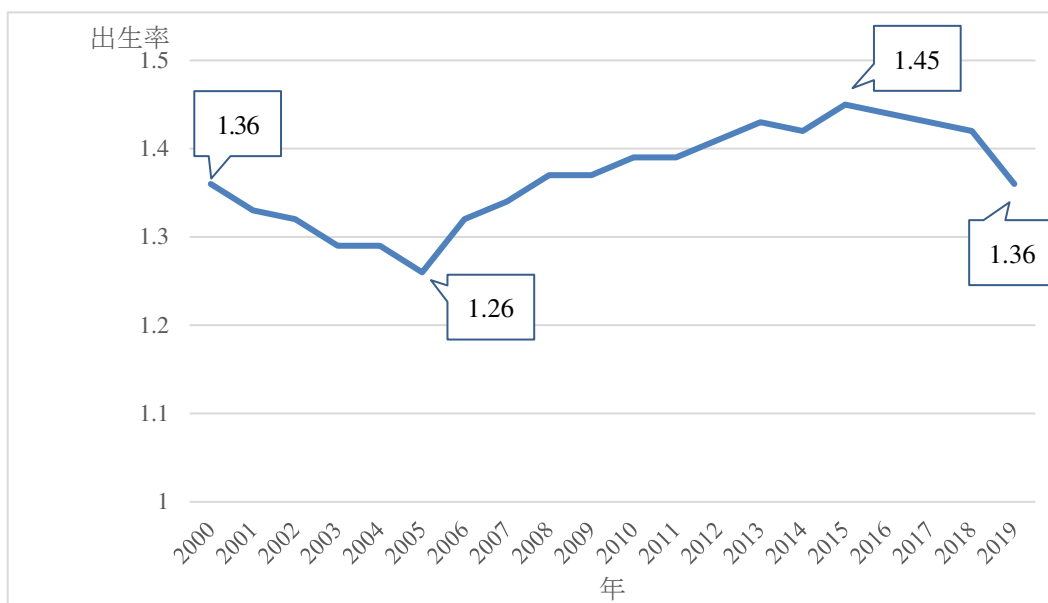


図5 日本の出生率の変化 (2000年以降) B

図5を見ると、出生率は2005年まで低下し、それ以降、少しずつ上昇して2015年には1.45になっている。その後また徐々に低下し、2019年には1.36になったことがわかる。このようにグラフを変えると、小さな変化がわかりやすい。

なお、折れ線グラフでは棒グラフと違い、数値軸を0から始めることは必須ではない。上の図4

のようなグラフでは、細かい変化を見ることができない。折れ線グラフは、あくまで時間的な変化を見ることが目的である。一方、棒グラフでは個々の棒の長さが量を表し、それを比較することが目的なので、0を基準にする必要がある。

ここで、図3から一部を切り出して図5を作ったときのように、時間の幅や目盛りの取り方によって、折れ線グラフから受け取る印象はかなり変わる。この点に留意が必要である。基本としては、まず時間幅をできるだけ長くとり、大きな傾向がつかめるようにする。そして必要に応じ、一部の期間にズームインして、より細かい変化がわかるようにする。このようにすれば、バランスのとれた見方が可能になるだろう（上田、2005）。

ところで、折れ線グラフでは、一つのグラフ中に複数の折れ線を表示することも多い。そのとき、各線が示すものを凡例にすることがよくある。図6は、その一例である。

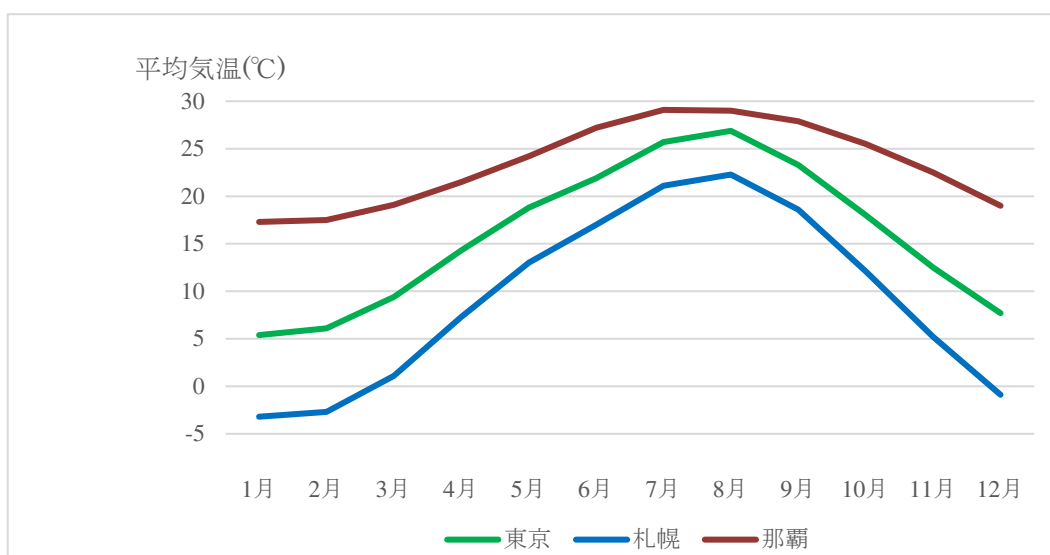


図6 平均気温のグラフ A

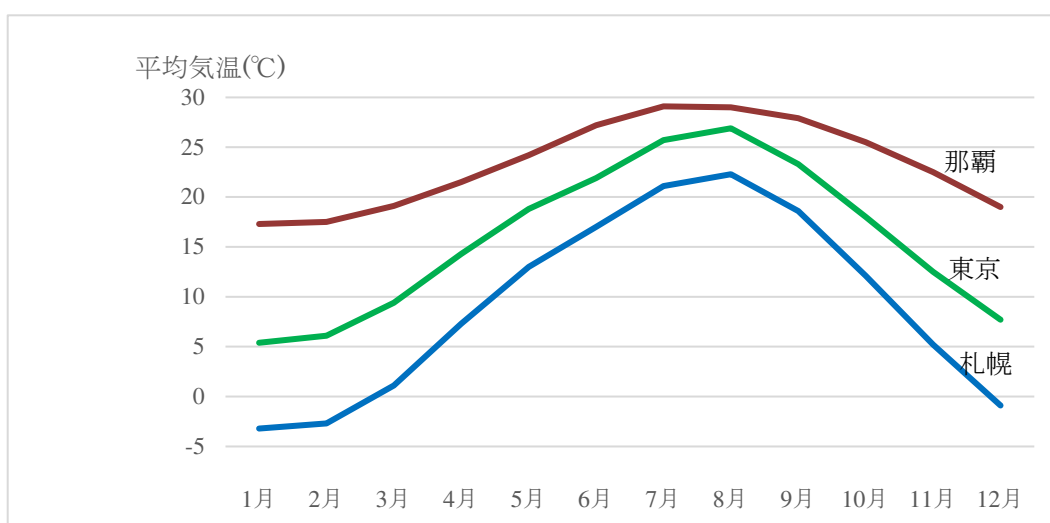


図7 平均気温のグラフ B

図6では、凡例と折れ線を目が行き来せねばならず、見る側に負担となる。線の数が増えれば、なおさらである。

こうした負担を減らすには、図7のように凡例を削除し、各線の表すものを直接示せばよい(関、2007)。阿部(2021)は、このような配慮を「表示の直接性」と呼んでいる。

円グラフ

図8は、円グラフのサンプルである。円グラフは、各要素の構成比を示すときに使われる。

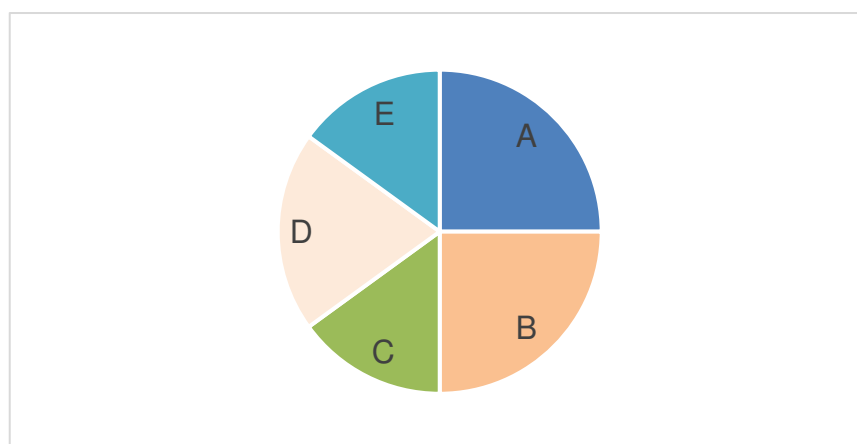


図8 円グラフの例

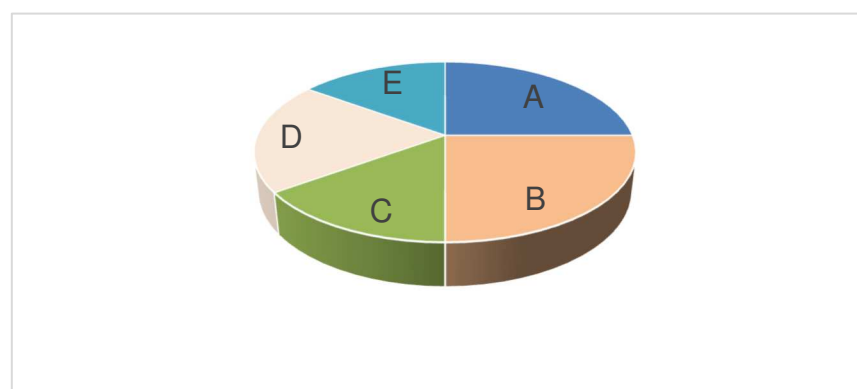


図9 円グラフの例(立体化)

表計算ソフトなどでは、図9のように円グラフを立体化(3D化)したものも作成できる。しかし、これは誤解を生みやすい。図8・図9のAとB、CとEは、同じ比率である。ところが、立体化した図9では、手前のBやCが実際以上に大きく見えてしまう。棒グラフなどでも立体化すれば、同様のおそれがある。したがって、グラフの立体化は原則として避けるべきである。

このことに関連するのが、Tufte(2001)のいう「データインク比」という考え方であろう。彼は、

グラフを含めた図解のわかりやすさの基準になるものとして、図 10 のようなことをあげている。

$$\text{データインク比} = \frac{\text{データインク（データを表すためのインク）}}{\text{図解に使われたインクの総量}}$$

図 10 データインク比

これは、ひとつの図解に使われたインク全体のうち、データそのものを表すために必要なインクが占める比率を指している。したがって、図解のなかに不要な表現が少なければ、データインク比は高くなる。この比率を高めるほど、よいデータ表現になる、というのが彼の考えである。

これに沿えば、グラフの立体化などは不要な要素を増やし、データインク比を低くする。立体化した図 9 よりも、もとの図 8 の方がデータインク比は高く、わかりやすいということになる。

つまり、立体化のような不要な装飾は省き、できるだけシンプルな図解がよい、ということである。これはデータを視覚表現するときに、ひとつの基準になる考え方であろう。

おわりに

データの視覚表現としてのグラフについて、見る側の理解を高め、そして誤解を避けるために必要なポイントをいくつか取り上げてみた。

近年、さまざまな分野で「可視化」「見える化」ということが言われるようになった。そのままでは見えていないものを見えるようにする、ということである。それによって、複雑なことでも理解しやすくなることが期待されている。なぜなら、見えることは理解につながるからである。英語の“see”という単語には、「見る、見える」とともに「わかる、理解する」という意味もある。これは、示唆的であろう。

一目で見てわかることは、視覚表現の大きな利点である。その反面、不適切な表現をすれば、誤った情報（misinformation）が伝わるということにもなる。ときには、意図して虚偽の情報（disinformation）を伝えるために、グラフや図解が用いられることもある。

いずれにせよ、さまざまなデータの重要性が増してきている現在、データの視覚表現の作成法や受け取り方についても、一定のリテラシーが必要になっているといえよう。

引用文献

阿部圭一. 2021. 『よくわかるデータリテラシー』(近代科学社).

関友作. 2007. 「説明と視覚表現」 比留間太白・山本博樹 編『説明の心理学』 80-93 (ナカニシヤ出版) .

Tufte, E. 2001. *The Visual Display of Quantitative Information*. Graphics Press, Conn.

上田尚一. 2005. 『統計グラフのウラ・オモテ』 (講談社) .

Wong, D.M. 2010. *The Wall Street Journal Guide to Information Graphics*. W. W. Norton & Co., NY.