

統計非専門家の認知を促進するための情報の可視化

A Case Study: Visualizing a Correlation for Non-Experts in Statistics.

西尾 義英^{*1}
Yoshihide Nishio安松 健^{*1}
Ken Yasumatsu徳見 理絵^{*1}
Rie Tokumi^{*1} シナジーマーケティング株式会社
Synergy Marketing, Inc.

This paper aims to explore an effective way to visualize data for non-experts in statistics, reporting analyzed problems and tested solutions in a case where we visualized a correlation between marketing metrics and found out that two sets of bars like the table lens is easier to understand than a popular scatter graph to unseasoned eyes. This case suggests a scatter graph requires high-level ability to see two variables at once and grasp their correlation as well as to look at scattered data points as a group and interpret its meaning.

1. はじめに

情報可視化の研究はグラフのビジュアル要素を解析した[Cleveland 1984]など 1980 年代には行われており、今や分析結果を可視化する適切なグラフの選択[Mackinlay 1986]も難しくなっている。しかし、多くのビジネスパーソンは統計専門家ではないため、企業実務現場におけるグラフを使ったコミュニケーションの場面では、統計専門家としては定石とも言えるグラフが、一般的にはなじみがなく情報伝達の手段として不十分であるという課題がある。

本稿は、ビジネス実務において統計非専門家にとっても認知しやすい可視化の方法を探究することを目的とし、情報の可視化に関する課題を分析、その解決例を報告する。事例としてはマーケティング・メトリクス間の相関関係を可視化するというものを選択した。これは相関関係の可視化には散布図を用いることが定石と考えていたところ、多くのビジネスの現場において散布図による可視化が効果的とは言えないという経験をしてきたことがきっかけである。

1.1 情報可視化とインタラクション

情報可視化は分析を目的とするもの(分析者自身のための可視化)とコミュニケーションを目的とするもの(他者に分析結果を伝達するための可視化)に分かれ、それぞれに適した手法を選択しなければならぬ[Steele 2011]。また情報可視化は静的な表現だけでなくインタラクションとの組み合わせとして捉えることが出来る[Yi 2007]。分析のための可視化については[Rao 1994]などが提案されている。一方コミュニケーションのための情報可視化において、インタラクションを対象とした研究は少なく、例えば[Segel 2010]は既存の手法を分類しているが、コミュニケーションのためのより良いインタラクションはこれからの研究領域である。

1.2 統計非専門家のための情報可視化

本稿では可視化を行った分析者が第三者に伝えるという状況で、可視化したグラフを読者自身が作成したことがないという意味で統計非専門家という言葉を使う。なお、分析の非専門家のための情報可視化をとりあげた研究は[Kwon 2011]などわずかに見られるが、分析者と読者でグラフ作成経験が異なるという前提で認知効果を研究したものは見つけれなかった。

2. 事例報告

2.1 可視化を必要とする読者と状況について

本事例は複数の店舗をチェーン展開するアパレル企業において、店舗毎の会員数前年比と売上前年比に関する分析結果の可視化を実施したものである。これは年度末の全社会議にて報告する分析資料の一部として作成され、この分析結果は、売上金額と会員数は正の相関があることを示すものであった。

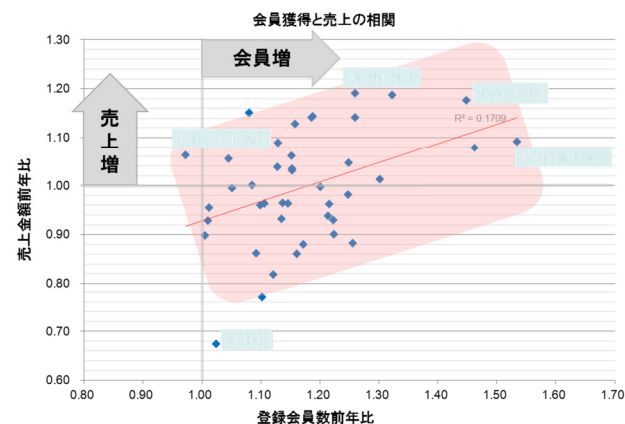
本事例において、読者は、店舗の店長やスタッフなどのセールス担当者やマーケティング担当者などであり、表計算ソフトなどの日常の利用者も一部である。また、可視化の前のデータ分析も我々によるもので、読者とは異なることを付記しておく。

2.2 2種類の可視化とその比較

分析結果から2種類の可視化を行い、3名の読者に見比べてもらった。ここではそれぞれの可視化の手法を説明するとともに、読者の観察内容と、読者へのインタビューを元に可視化手法を比較する。

(1) 散布図によるもの

最初の図は表計算ソフトにより描画した散布図で、X軸に会員数の前年比、Y軸に売上の前年比をとった。回帰直線と決定係数R²も表計算ソフトによる出力である。手作業で一部の店舗名、データの散らばりへの網掛け、前年比 1.00 の位置を強調する補助



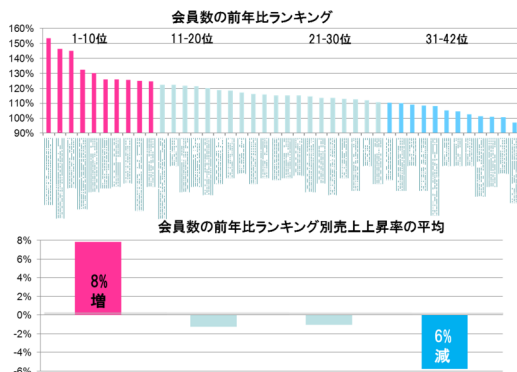
線と矢印を書き入れた。引用にあたり店舗名を伏せるための加工を行っている。

読者の反応としては、全体の傾向よりも傾向から外れた店舗の存在が気になり違和感があるというものであり、ポジティブに評価

連絡先:シナジーマーケティング株式会社 システム開発部
研究開発 G 西尾 義英<nishio.yoshihide@synergy101.jp>

するコメントは得られなかった。なお、2 項目間の相関性と全体の傾向を見せるために加えた回帰直線や網掛けの効果は確認できなかった。

(2) 2 枚の棒グラフに分解したもの



こちらの図は上下 2 つのグラフからなる。上のグラフは店舗別の会員数の前年比を垂直棒グラフで表わしたもので、左から降順に店舗をソートしてある。42 店舗を上位から 4 等分するセグメントに分類し棒グラフを色分けした。その下には店舗セグメント毎に売上の増加率または減少率の平均値を垂直棒グラフで示した。全体としては会員増加率上位 10 店舗の売上前年比が平均 8% 上昇し、会員増加率下位 12 店舗の売上前年比が平均 6% 減少したことを表わしている。散布図と同様、引用にあたって店舗名を伏せる加工を行っている。

読者の評価は、3 名ともに散布図よりこちらの方が見やすく、分かりやすいというものであった。実際にこちらの図は 2 項目間の相関を伝達することに成功した。また、会員数が前年比 20% 以上増加しないと売上増加への影響が見られない、という分析的なコメントが得られた。

3. 考察

3.1 可視化の課題

散布図に対しては「2 軸で見るので混乱する」という意見があった。また、観察からは読者がグラフ中の店舗を表す点を追っていることが分かった。つまり読者の「混乱」は点の座標を X 軸と Y 軸から読み取ろうとする視線移動から来たと推測される。ここで、説明側の意図に反し、個々の点が注目されてしまうということが問題である。我々はこの原因は散布図が比較的難しい可視化であるためと考えた。

散布図が難しいというのはどういふことであろうか。一つは散布図が 2 変数の傾向を同時に図示しており、その関係を読み取るということが、例えば 1 変数の大小関係を読み取るよりも難しいということである。また、複数の店舗を集合として認識し、集合としての特徴を読み取るということは、個々の店舗の特徴を読み取ることよりも難しいと推測される。

3.2 対策

散布図に対比する 2 つ目の可視化は、先に述べた、散布図の難しさについての仮説、すなわち(1)2 変数の傾向を同時に見て関係を読み取る難しさ、および(2)データの集合を認識し、その特徴を読み取る難しさの 2 つを回避するものとなっている。これからその考え方を説明する。

(1) 2 変数の傾向を同時に見て関係を読み取る難しさ

同時に見せるのではなく、順番に伝達することで認知を促進できる。つまり、まず店舗を会員数の前年比降順でソートした図、次に同じ並びの店舗に対して売上の前年比を示す図という 2 枚に分解して並べるという方法をとった。これは the table lens([Rao et al. 1994])が変数間の相関を図示する方法と類似する。

(2) データの集合を認識し、その特徴を読み取る難しさ

当初(1)の方法で店舗毎の売上前年比をそのまま図示したところ、売上前年比のちらばりが大きく、全体の傾向を伝達するにはまだ不十分だと考えた。一方、分析結果から導き出された示唆は「会員がたくさん増加した店舗は売上も増加している」というものであったので、会員数前年比が高い店舗群と低い店舗群の売上前年比を対比させることが出来れば十分であると判断し、図において店舗群を区別することにした。

つまり会員数前年比でソートした店舗を上位から 4 つのセグメントにまとめ、各セグメントにおける売上前年比の平均値を代表値とした可視化を行った。4 つに分割したこと、数学的、認知科学的根拠は無いが、分割数が数個以内に収まること、会員数のギャップが生じる箇所で分割するという意識した結果である。

3.3 想定される反論

我々は散布図の有用性を全面的に否定するものではなく、今回の事例を一般化することにも慎重であるが、散布図の難しさに対する考察には以下の反論が想定される。

(1) 散布図の可視化方法に問題があったのではない

最初の散布図の前に、表計算ソフトによる描画そのままの図も見せたが、両者に対する読者の反応の違いは大きく変わらなかった。可視化方法による問題の可能性は小さいと考えている。

(2) 散布図に十分な数の点があれば個々の点を意識することなく全体の傾向を伝達することができるのではない

仮にそうだとしても、あるデータにおいて散布図より理解されやすい可視化が存在すると主張することは可能である。

(3) 相関関係を読み取れないのはそもそも相関の小さいデータを可視化したからではない

この可能性は否定出来ない。調査した限りでは散布図から相関関係を読み取れる相関の強さについてデータは得られなかった。今後も事例を収集するか、実験によって確認すべきである。

3.4 今後の課題

上述したように本事例で考察した仮説の確かさは更なる研究を要するが、その上で今回の手法は他の場面でも利用可能な汎用性をもっていると考えている。そのためには、事例において店舗セグメントを 4 つとした判断を再現可能にすることが必要である。これを何らかのモデルによって自動化するか、分析者が画面操作でセグメント数を調節出来るような仕組みを検討したい。

参考文献

- [Cleveland 1984] W. S. Cleveland and R. McGill: Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods, *Journal of the American Statistical Association*, 1984.
- [Kwon 2011] B. C. Kwon, B. Fisher, and J. S. Yi: Visual analytic roadblocks for novice investigators, *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2011.
- [Mackinlay 1986] J. D. Mackinlay: Automating the Design of Graphical Presentations of Relational Information, 1986.
- [Rao 1994] R. Rao and S. K. Card: The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus + Context Visualization for Tabular Information, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York: ACM, 1994.
- [Segel 2010] E. Segel and J. Heer: Narrative visualization: telling stories with data., *IEEE transactions on visualization and computer graphics*, 2010.
- [Yi 2007] J. S. Yi, Y. A. Kang, J. Stasko, and J. Jacko: Toward a deeper understanding of the role of interaction in information visualization., *IEEE transactions on visualization and computer graphics*, 2007.
- [Steele 2011] J. Steele, N. Iliinsky 編, 増井 俊之 監訳: Beautiful Visualization, オライリー・ジャパン, 2011