

Pictures worthy of a thousand words

Penelope M. A. Brasher, PhD · Rollin F. Brant, PhD

Received: 18 June 2010 / Accepted: 6 July 2010 / Published online: 10 September 2010
© Canadian Anesthesiologists' Society 2010

Everyone has heard the adage, “a picture is worth a thousand words”, and yet, all too frequently, the information in graphical displays submitted to the *Journal* could be summarized in fewer than a dozen words. In this editorial, we discuss the use of bar graphs for displaying data arising from measurements and classifications. We identify the deficiencies of these graphs and propose more informative alternatives.

Bar graphs for measurement (continuous) data

Bar graphs that portray means with error bars for standard errors (or sometimes standard deviations), such as those in Fig. 1A, are ubiquitous in the medical literature. It is not clear how such plots have come into widespread use, as

many statistical packages do not provide them as a standard option. In fact, perhaps surprising to some readers, a majority of statisticians dislike them intensely and refer to them disparagingly as dynamite plots, skyscrapers with TV antennas and pinhead plots. One Department of Biostatistics even has a policy about them:

“Dynamite plots often hide important information. This is particularly true of small or skewed data sets. Researchers are highly discouraged from using them, and department members have the option to decline participation in papers in which the lead author requires the use of these plots.”¹

Such plots have many deficiencies. First, there is a low data-ink ratio.² A great deal of ink (and space!) is used to show only four numbers – two means and two standard errors. Second, there is the problem of accuracy. The reader must “look up” the values by comparing the heights of the bars (lines) with the axis. The four numbers could be reported more accurately and concisely in the text. Third, there is the issue regarding the relevance of the information provided. As pointed out by Cleveland,³ “a standard error of a statistic has value only insofar as it conveys information about *confidence intervals*”. If the intent of the plot is to convey information about sample-to-sample variation of the mean, then the confidence interval should be plotted. The readers should not have to do arithmetic in their heads! This is particularly important for statistics other than the mean, since not all confidence intervals for population parameters are based on the normal approximation.

As well, there is the issue of robustness to outliers, since both the mean and standard deviation (and thereby, the standard error) are prone to perturbation by atypical data values. Finally, of primary interest when comparing groups is the *difference* between the means of populations not the means of individual groups in and of themselves. Thus, a

Editor's Note This editorial is one in a series of editorials and special articles published by the *Journal* and dedicated to enhancing the accuracy and transparency of reported scientific articles. This particular article addresses the subject of graphical displays in the form of bar graphs. It is our hope that the articles in this series will provide practical insights and guidance that will benefit our authors and our readers as well as all those who participate in the scientific enterprise.

P. M. A. Brasher, PhD (✉)
Centre for Clinical Epidemiology and Evaluation, Vancouver
Coastal Health Research Institute, 828 West 10th Avenue,
Vancouver, BC V5Z 1M9, Canada

R. F. Brant, PhD
Department of Statistics, University of British Columbia and
Child and Family Research Institute, Vancouver, BC, Canada

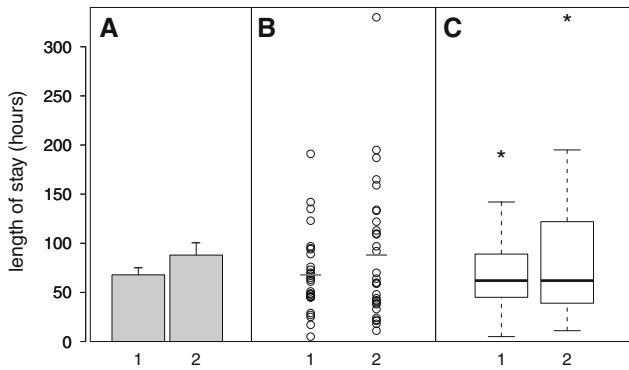


Fig. 1 Length of hospital stay for two patient groups. **A** Average length of stay by group. Lines represent the standard error of the mean (sem). **B** Dotplots of the length of stay by group. The thick bar represents the sample mean of the group. **C** Boxplots of the length of stay by group

confidence interval should be provided for the difference between means and not for the means of individual groups.

Banishing plots, such as Fig. 1A, does not mean that authors should not provide a graphical display relating to the comparison between two groups. Optimally, the authors should provide the reader with a plot of the raw data.

Figure 1B shows the data on which the summary statistics are based. The reader can assess the range and variability of the data, any skewness that might be present, the presence of outliers, and the extent of overlap of the two groups. Figure 1B provides a complete description of the *samples*. If desired, summary statistics can be added to the plot. In this example, since the averages of the two groups are the summary statistics of primary interest, these have been overlaid on the graph of the individual data.

When sample sizes are large enough that superimposed points distort perception, a boxplot is an option that is widely recommended (Fig. 1C). Boxplots (also called box-and-whisker plots) depict the data with five-number

summaries.⁴ The graph provides information about the centre and the spread and skewness of the sample and indicates the presence of unusually large (or small) outlying values.

Bar graphs for categorical data

Bar graphs are used widely to represent tabular summaries for categorical data or for displaying summary statistics across categories.

Figure 2A shows mortality rates categorized by sex (male vs female) and type of community (urban vs rural). In principle, bar graphs provide valid representation of such data, but it often may be difficult to discern patterns. Dot charts (Fig. 2B) are superior to bar charts for conveying patterns that arise in multi-category comparisons,³ and they have a much higher data-ink ratio.

A glance at Fig. 2B reveals that the pattern of the four points is not consistent with increasing age. While the mortality rate for urban males is always greater than that for rural males, the same is not true for females. The same information is not immediately apparent in Fig. 2A. Indeed, to decode this pattern from Fig. 2A requires two scans, first a comparison of rural and urban males for each age group and then the comparison of rural and urban females for each age group.

Another advantage of dot charts over bar charts is that groups can be depicted using different symbols. It is easier to distinguish symbols than to decipher different “hatching” schemes, which are typically used in black and white publications.

In his book, *Statistical Rules of Thumb*,⁵ Van Belle offers some excellent advice, “Always think of an alternative to a bar graph”.

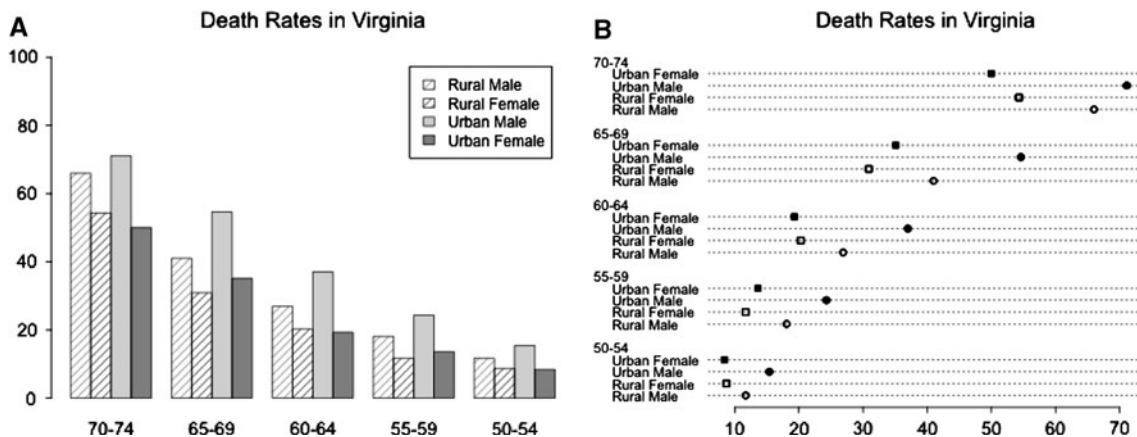


Fig. 2 Mortality rates by sex and place of residence.⁶ **A** Clustered bar chart. **B** Dot chart

In summary

Graphical displays have the potential to convey complex information to the reader quickly and efficiently. There is no formula for determining the type of graphical display to use for particular groupings of data. Authors should graph their data and results in numerous ways to determine the method that conveys the most information to the reader. In so doing, authors should keep in mind Tufte's list of nine "shoulds" for graphical displays. They deserve repeating here.

Graphical displays should:

1. show the data,
2. induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else,
3. avoid distorting what the data have to say,
4. present many numbers in a small space,
5. make a large data set coherent,
6. encourage the eye to compare different pieces of data,
7. reveal the data at several levels of detail, from a broad overview to the fine structure,
8. serve a reasonably clear purpose: description, exploration, tabulation, or decoration, and
9. be closely integrated with the statistical and verbal descriptions of a data set.

Des images qui valent mille mots

Une image vaut mille mots»: tout le monde connaît l'adage. Pourtant, trop souvent encore, les informations soumises sous forme graphique au *Journal* pourraient se résumer en quelques mots. Dans cet éditorial, nous abordons la question de l'utilisation d'histogrammes pour la représentation de données issues de mesures et de classifications. Nous soulignons les faiblesses de ce genre de graphiques et proposons des alternatives mieux adaptées pour présenter ce type d'informations.

Histogrammes pour les données de mesure (continues)

Les histogrammes qui décrivent des moyennes à l'aide de barres d'erreurs pour les erreurs type (ou parfois les écarts

Note de la rédaction Cet éditorial fait partie d'une série d'éditoriaux et d'articles spéciaux publiés par le *Journal* et dont l'objectif est d'améliorer la précision et la transparence des articles scientifiques présentés. Cet article en particulier aborde la question des représentations graphiques sous forme d'histogrammes. Nous espérons que les articles de cette série offriront des perspectives et des conseils pratiques qui profiteront à nos auteurs et à nos lecteurs, ainsi qu'à tous ceux qui prennent part à l'entreprise scientifique.

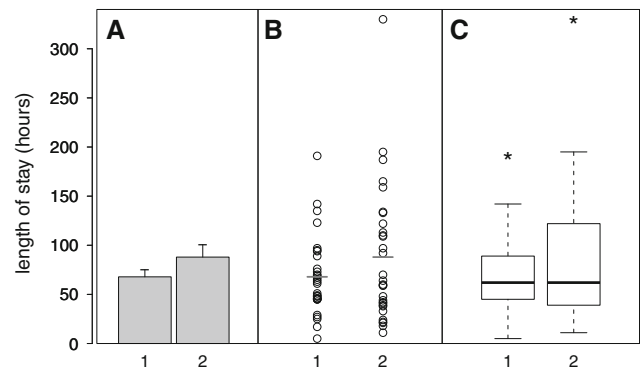


Fig. 1 Durée d'hospitalisation de deux groupes de patients. **A** Durée moyenne d'hospitalisation par groupe. Les lignes représentent l'erreur type de la moyenne (etm). **B** Diagrammes de points de la durée d'hospitalisation par groupe. La bande épaisse représente la moyenne d'échantillon du groupe. **C** Diagrammes de boîtes de la durée d'hospitalisation par groupe

type), comme ceux présentés en figure 1A, sont omniprésents dans la littérature médicale. L'origine de l'utilisation généralisée de tels graphiques est incertaine, ces diagrammes ne faisant pas partie des options de base de nombre de logiciels de statistique. En fait, au risque d'étonner certains de nos lecteurs, la majorité des statisticiens les ont en profonde aversion et les surnomment péjorativement 'graphiques explosifs', 'gratte-ciels munis d'antennes TV' ou encore 'diagrammes en tête d'épingle'. Notre département de biostatistique a même une politique les concernant:

«Les 'graphiques explosifs' occultent souvent des informations importantes. C'est particulièrement le cas pour les ensembles de données de petite taille ou asymétriques. Nous décourageons vivement les chercheurs de les employer, et les membres du département peuvent refuser de participer à un article pour lequel l'auteur principal exige l'utilisation de tels graphiques.»¹

De tels graphiques ont de nombreux défauts. Tout d'abord, le rapport données-encre est médiocre.² Pensez à la quantité d'encre (et d'espace!) nécessaire à illustrer quatre nombres seulement, deux moyennes et deux erreurs type. Ensuite vient le problème de la précision. Le lecteur doit «consulter» les valeurs en comparant la hauteur des bandes (lignes) par rapport à l'axe. Ces quatre chiffres pourraient être rapportés de façon à la fois plus précise et plus succincte dans le corps du texte. Troisièmement se pose le problème de la pertinence des informations présentées. Comme le soulignait Cleveland,³ «l'erreur type d'une statistique n'a de valeur que si elle transmet des informations quant aux intervalles de confiance». Si le but du graphique est de transmettre des informations quant à la variation de la moyenne d'un échantillon à un autre, alors il faut inclure l'intervalle de confiance dans le graphique. Les

lecteurs ne devraient pas avoir besoin de faire du calcul mental! Ce point revêt une importance particulière lorsqu'il s'agit de données statistiques autres que la moyenne, étant donné que tous les intervalles de confiance pour les paramètres de population ne sont pas basés sur une approximation normale.

Il y a en outre le problème de la robustesse des données par rapport aux données aberrantes: en effet, l'écart moyen et l'écart type (et par conséquent, l'erreur type) sont sensibles aux perturbations causées par des valeurs de données atypiques. Enfin, ce qui intéresse le lecteur lorsqu'on compare des groupes, c'est la différence entre les moyennes de population—non pas les moyennes de groupes individuels en et pour soi. Dès lors, il faut fournir un intervalle de confiance qui décrive la différence entre les moyennes et non les moyennes des groupes individuels.

Interdire des graphiques comme celui de la figure 1A ne veut pas pour autant dire que les auteurs ne devraient pas fournir de représentation graphique liée à la comparaison entre deux groupes. Idéalement, les auteurs devraient proposer un graphique représentant les données brutes.

La figure 1B montre les données sur lesquelles se fondent les données statistiques de base. Le lecteur peut apprécier l'étendue et la variabilité des données, toute asymétrie ou déviation possiblement présentes, la présence de données aberrantes, et le recouvrement des données des deux groupes. La figure 1B illustre une description complète des échantillons. Si on le souhaite, on peut ajouter les données statistiques de base au graphique. Dans l'exemple présenté ici, étant donné que les données de base qui nous intéressent sont les moyennes des deux groupes, elles ont été superposées sur le graphique des données individuelles.

Lorsque les tailles d'échantillon sont tellement importantes que la superposition des données en complique la compréhension, le diagramme en boîtes est une des

alternatives vivement recommandées (figure 1C). Les diagrammes en boîtes (également appelés boîtes à moustaches) décrivent les données statistiques de base à l'aide de cinq nombres.⁴ Le graphique fournit des informations concernant le centre, l'étendue et l'asymétrie de l'échantillon et indique la présence de valeurs aberrantes exceptionnellement grandes (ou petites).

Les histogrammes pour les données catégorielles

Les histogrammes sont très utilisés pour représenter des sommaires sous forme de tableaux de données catégorielles, ou pour illustrer des données statistiques de base touchant plusieurs catégories.

La figure 2A illustre les taux de mortalité catégorisés par sexe (homme vs. femme) et selon le type de collectivité (urbaine vs. rurale). En principe, les histogrammes permettent une bonne représentation de telles données, mais il est souvent difficile de discerner clairement les tendances. Les diagrammes en points (figure 2B) sont supérieurs aux histogrammes pour illustrer les tendances qui apparaissent dans des comparaisons de catégories multiples³—et leur rapport données-encre est bien meilleur.

Il suffit d'un regard vers la figure 2B pour voir que la disposition des quatre points n'est pas nécessairement liée à l'âge. Alors que le taux de mortalité des hommes en milieu urbain est toujours plus élevé que celui des hommes vivant en campagne, ce n'est pas le cas chez les femmes. Cette même information n'est pas immédiatement apparente dans la figure 2A. En fait, il faut deux lectures de la figure 2A pour distinguer cette tendance: premièrement, une comparaison des hommes en milieu rural et urbain pour chaque catégorie d'âge, puis la même comparaison pour les femmes.

Un autre avantage des diagrammes de points par rapport aux histogrammes est que les données peuvent être décrites

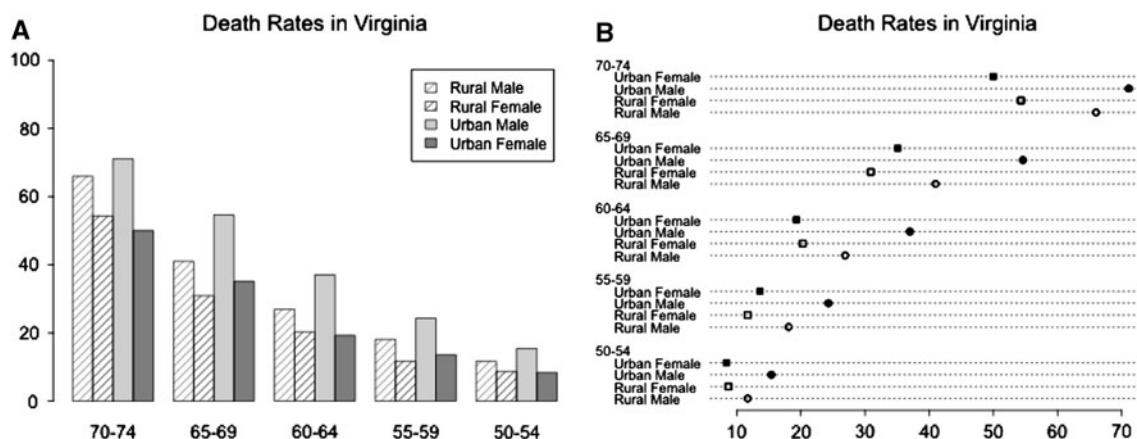


Fig. 2 Taux de mortalité par sexe et lieu de résidence.⁶ A Histogrammes groupés. B Diagramme de points

à l'aide de différents symboles. Il est plus facile de distinguer des symboles que de déchiffrer différents types de hachures—une caractéristique typique des histogrammes dans de nombreuses publications en noir et blanc.

Dans son ouvrage, *Statistical Rules of Thumb*,⁵ Van Belle donne un excellent conseil: «Pensez toujours à une alternative à l'histogramme».

En résumé

Les représentations graphiques peuvent fournir de façon rapide et efficace des informations complexes au lecteur. Il n'existe pas de formule pour déterminer le type de représentation graphique à utiliser pour un ensemble particulier de données. Les auteurs devraient essayer plusieurs représentations graphiques de leurs données afin de déterminer quelle méthode transmet le plus d'informations au lecteur. Ce faisant, les auteurs devraient garder à l'esprit les neuf «règles» de Tufte concernant les représentations graphiques. Ces règles méritent d'être réitérées.

Les représentations graphiques devraient:

10. montrer les données;
11. amener le lecteur à réfléchir sur le fond plutôt que la forme, sur la substance plutôt que la méthodologie, la conception graphique, la technologie de la production graphique, ou autre chose;
12. éviter de déformer ce que les données ont à dire;
13. présenter beaucoup de données dans un espace limité;
14. rendre cohérent un vaste ensemble de données;
15. encourager le lecteur à comparer différentes parties de données;
16. révéler les données à plusieurs niveaux de détail, d'une perspective large à une structure détaillée;
17. avoir un objectif relativement clair: la description, l'exploration, la tabulation ou la décoration; et
18. être bien intégrées aux descriptions statistiques et verbales d'un ensemble de données.

Acknowledgements The authors would like to thank David Johnson for permission to use the data in Fig. 1. This is unpublished data arising from the study described in reference 7. The data in Fig. 2, which appeared originally in reference 6, was provided by the R-project.⁸

Competing interests None declared.

References

1. Department of Biostatistics, Vanderbilt University. Departmental Policy. Available from URL: <http://biostat.mc.vanderbilt.edu/wiki/Main/StatisticalPolicy> (accessed May 2010).
2. Tufte ER. The Visual Display of Quantitative Information. 2nd ed. Cheshire, CT: Graphics Press; 2001.
3. Cleveland WS. The Elements of Graphing Data. Summit, NJ: Hobart Press; 1994.
4. Tukey JW. Exploratory Data Analysis. Reading, MA: Addison-Wesley; 1977.
5. van Belle G. Statistical Rules of Thumb. 2nd ed. Hoboken, NJ: Wiley & Sons, Inc; 2008.
6. Molyneux L, Gilliam SK, Florant LC. Differences in Virginia death rates by color, sex, age, and rural or urban residence. *American Sociological Review* 1947; 12: 525-35.
7. Johnson DW, Adair C, Brant R, Holmwood J, Mitchell I. Differences in admission rates of children with bronchiolitis by pediatric and general emergency departments. *Pediatrics* 2002; 110: e49.
8. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from URL: <http://www.R-project.org> (accessed July 2010).