Use boxplots to visualise group differences: Bar charts are inefficient, uninformative, and misleading

Daisung Jang

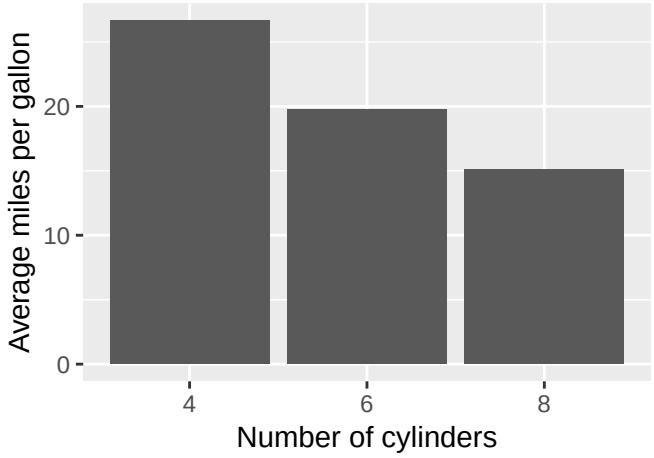
Executive summary

This memo is intended for people who represent group differences in means with bar charts. Bar charts are undesirable for that purpose. If one desires to graphically display group differences in means, the boxplot is the superior graphical display.

Why bar charts are a bad way to display group differences in means

Bar charts are often used to display group differences in means. But for that purpuse are inefficient, uninformative, and misleading.

To demonstrate, I use the *mtcars* dataset to display the same data in bar chart and boxplot form. The *mtcars* dataset comprises 'fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models)'. First, below is a bar chart that plots average miles per gallon against the number of cylinders in an engine:

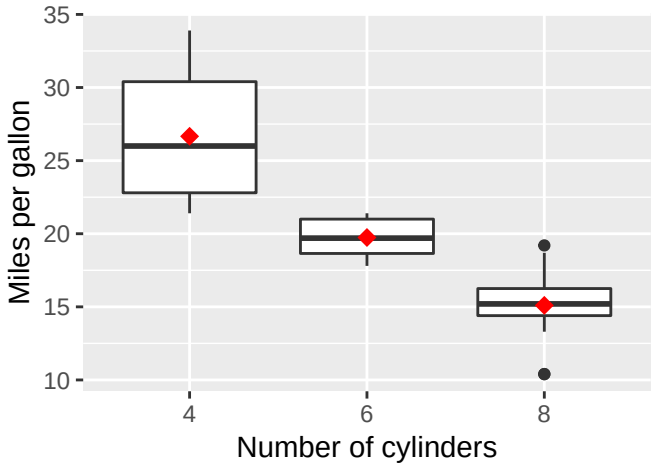


There are three key reasons why such graphs are bad:

- They are inefficient because they waste space. The bars in a bar graph convey just one piece of information—the mean. All space below and above the mean are meaningless.
- They obscure important information about the nature of the data, including the variance, skewness, and potential outliers.
- They are misleading because bar graphs give the impression that a particular condition or level of factor leads to a production of a response up to (or down to) some level. This is simply wrong. Units of analysis within groups do not collectively produce a response up to (or down to) some level, but rather produce a distribution of data. If you intend to provide information about means, a table is more effective.

Boxplots are superior when displaying group differences in means

Boxplots are superior to bar charts because they not only display central tendency, but also convey variance, skewness, and indicate outliers. Below is the same data plotted using boxplots, with red diamonds indicating means.

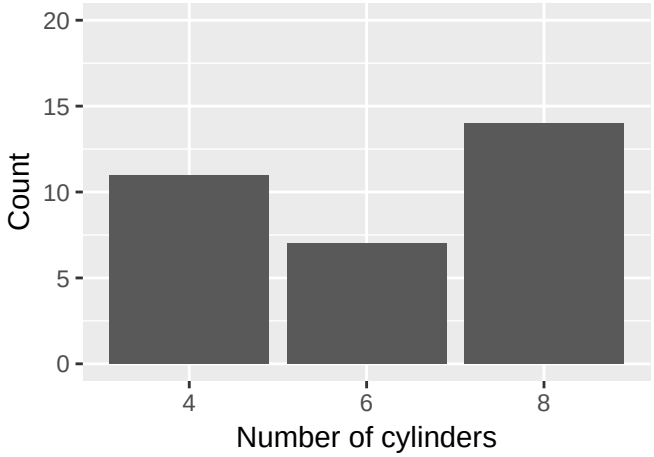


The following are immediately clear:

- In addition to clear differences in means, variabilty within groups also becomes obvious.
- Outliers are visible, and the viewer can form their own thoughts about them.
- Assumption of normality is replaced with a visual check for normality.
- The subjects in the conditions do not appear to collectively reach some level, but produce distributions.

Bar charts are good for graphic representation of accumulation

Bar charts can be used productively to display group differences in counts. Used this way, every increment in a bar chart is meaningful in that they convey an additional data point. Below is a bar chart that shows the number of cars in the *mtcars* dataset with either 4, 6, or 8 cylinders in their engines.



All increments in the bars accurately portray the data and does not mask information. Other kinds of data that are best thought of 'accumulations', such as time elapsed or total volume, are also useful to display in a bar chart format.