CMU Data Science Club April 2017
Effective Data Visualization

Graphics Checklist

Component	Successful	Unsuccessful
Legible	Image is in a vector or high-resolution bitmap format. Font is large enough to read easily. Data are not hidden or overwhelmed by ticks, axes, or gridlines. Different colors or symbols are easily distinguishable. Aspect ratio shows data clearly.	Image is low-resolution, grainy, or pixelly. Font is too small to read. Data are hidden by other graph elements. Colors or symbols cannot be distinguished. Aspect ratio causes data to be too bunched up or spread out to see patterns easily.
Comprehensible	Graphic has an informative title or caption, axis labels, and (if relevant) legend. Axis ticks are labeled with sensible, round numbers. Graphic axes, legend, colors, etc. are consistent across small multiples (if relevant).	Graphic has no (or unclear) title or caption, axis labels, or legend. Axis ticks are unmarked or are marked at arbitrary, unhelpful numbers. Graphic elements are inconsistent across small multiples.
Informative	Graph clearly highlights any trend or pattern in the data, which is summarized in the title or caption and described in the body text. Interesting differences or comparisons are plotted directly.	Graph highlights no interesting or useful pattern. Pattern is not indicated in title or caption, or not described in body text. Readers have to mentally compute differences instead of seeing them directly.
Statistical Summaries	Data are shown foremost, with statistical summaries overlaid as appropriate. Some measure of statistical precision (e.g. a confidence interval) is shown for any summary statistic.	Summaries (e.g. averages, medians, trend lines) are shown alone, without the underlying data. Summary statistics are shown with no indication of their statistical precision.

Component	Successful	Unsuccessful
Quantitative Comparisons	Quantitative variables use visual encodings high on the Cleveland-McGill ordering. Encodings are used sensibly (bars start at 0; hues are ordered intuitively; etc.). Elements to be compared are as near each other as possible.	Quantitative variables use visual encodings low on the ordering. Encodings are implemented poorly (bars not anchored at 0; arbitrary hues assigned to quantitative/ordinal variable). Elements to be compared are distant.
Grouping and Search	Gestalt and preattentive processing features are chosen to ease task (find important groups, follow lines, etc.) Elements to be compared are aligned, as much as possible. Distinct variables are mapped to separable dimensions. Choice of colors, shapes, etc. is easy to discriminate.	Difficult to find groups, follow lines, etc. Elements to be compared are not aligned. Distinct variables are mapped to integral dimensions (e.g. point width and height). Distinct elements cannot be discriminated.
Cognition	Differences, proportions, or other important derived variables are plotted directly. Items are ranked by variables on which comparisons are to be made.	User must compute differences, etc. mentally. Ranking is arbitrary or unhelpful for analysis (e.g. alphabetical).
Consistency	Meaning of graphical elements is consistent across small multiples. Changes in design are purely data-driven. Visual variables are used only when mapped to data. Semantic associations are used, if possible (e.g. blue = cold, red = hot). More means more (larger size or deeper hue maps to larger value of the variable).	Small multiples are not consistent. Design changes are stylistic or arbitrary (e.g. new colors for the same categories). Superfluous visual variables are shown (3D, shadow, other variables not mapped to data). Semantics are mangled (e.g. 'orange' and 'blue' crab species are not mapped to orange and blue colors). More (stronger encoding) is mapped to less (lower value of data variable).