https://doi.org/10.20378/irbo-51101

The dot plot: A graphical tool for data analysis and presentation

Lukas Sönning
University of Bamberg

Cleveland's (1984) introduction of the dot plot to the scientific community dates back more than 30 years. Its clarity, flexibility, and efficiency make it a useful tool that is applicable to a wide range of descriptive and inferential analyses. Yet, this graph type has not gained the currency it deserves; in fact, it appears to be unknown to most researchers (Jacoby 2006; Keen 2010). This paper presents the dot plot and brings together various extensions that have emerged over the last 30 years. Advantages over alternative chart types are illustrated and design options and recommendations for the display of more complex data sets are discussed. The application of dot plots to quantitative data in linguistics is demonstrated, focusing on examples from corpus linguistics, meta-analysis and statistical modeling. The final sections reflect on important limitations of this display type and refer the reader to software for the implementation of dot plots. An online appendix provides a brief R tutorial as well as templates for Microsoft Excel, which allow for easy production of dot plots by entering data into spreadsheet templates.

1. Introduction

Graphs are indispensable tools in quantitative research since they reveal structure in the data in an effective and accessible way. A functional distinction is often made between graphs for data analysis and data presentation (Fienberg 1979; Schmid 1983). Graphing in data analysis serves to communicate between researcher and data. It is an iterative process and involves drawing many displays to gain different perspectives on a data set (Unwin 2015). Presentation graphs, on the other hand, aim to effectively communicate findings to an audience. To this end, principles of visual perception should guide the choice of graph type and graphical parameter settings to obtain an effective display.

This paper introduces the dot plot (Cleveland 1984), a display method suitable for both data analysis and presentation. It is an (unjustly) underutilized graph type that appears to be unfamiliar to most researchers (Jacoby 2006; Keen 2010). Its conceptual simplicity, however, makes it a versatile tool for many types of statistical analyses. The design of the dot plot is inspired by insights gained from research on visual perception, the aim being an optimization of the decoding of quantitative information. There are also several practical advantages compared to other more widely used chart types, such as the bar chart. It is the aim of this paper to demonstrate the usefulness and added value of the dot plot and argue for its routine usage in quantitative research (for examples of their application in linguistic research see Werner & Fuchs 2016; Krug et al. 2016; Schützler forthcoming).

After an outline of the theoretical background on graphs in scientific research, Section 3 introduces the simple dot plot, including the relevant terminology and a number of extensions for more complex data sets. Next, advantages over alternative chart types are summarized and illustrated. Section 5 discusses design options and gives recommendations on the construction of dot plots. Applications to linguistic data analysis are demonstrated in Section 6, including usage in simple meta-analyses and in the investigation of binary and frequency outcomes in corpus linguistics. The final sections reflect on the limitations of dot plots and discuss their implementation in R and Microsoft Excel. An online appendix includes brief tutorials for dot plots in R and spreadsheets for their implementation in Excel.

2. Theoretical background

The discussion and comparative evaluation of graph types can build on theoretical insights gained across a wide range of disciplines. These include exploratory data analysis (Tukey 1977), experimental research on graphical perception (Cleveland 1993), psychology (Wertheimer 1938) and neuroscience (Kosslyn 2006). This section aims to lay a conceptual and terminological foundation and elaborates on four aspects: (i) the purpose of statistical graphs, (ii) the active process of decoding information from a graph, (iii) a model of graphical perception, and (iv) psychological principles of graph perception and design. Key terms are italicized throughout the paper.

The purpose of graphs

Tukey (1993: 2) concisely states the "true" purpose of graphs: first, graphs are not meant to communicate precise values, but are rather *semi-quantitative*; exact numbers should be provided in tables. Second, graphs are for *comparisons*. As pointed out by Tufte, "at the heart of quantitative reasoning is a single question: *Compared to what?*" (1990: 67, emphasis in original). Third, graphs are for *impact* on the viewer – important information must be easily discernible. In short, the purpose of a graph is to "force" the viewer to make key *comparisons* of interest in a *semi-quantitative* manner. According to Tukey (1993: 3), such *semi-quantitative comparisons* yield statements like "is way above", "is above", "is a little above", "is almost equal to/is almost on", "is a little below", "is below", "is way below".

Decoding information from a graph

In order for such *semi-quantitative comparisons* to be made, the viewer must formulate a conceptual question, a piece of information to be extracted from the graph (Pinker 1990: 94). In other words, not every piece of information can be forced upon the viewer; rather, he or she plays an active role in decoding information from a display. This operation can be conceived of as a two-step process (Ware 2013: 139). First, a *visual query* is formulated, which identifies the problem to be solved or question to be answered. The second step is the *visual search*, the decoding of the display in response to the query, whereby the viewer identifies relevant patterns in the display. The success of a visual display thus also depends on the viewer (and data analyst), who must know where to look and what to look for.

A model for graphical perception

The *visual search* is an active process guided by principles of visual perception. Based on experimental research, Cleveland (1993) proposed a model for graphical perception. It introduces a number of useful terms for the description of displays and the mental operations involved in decoding information. Graphs encode *quantitative* and/or *categorical* variables. Quantitative variables yield values or measurements; categorical variables (binary, nominal and ordinal) assign observations to different groups or categories. The displayed content of a graph can be divided into *physical* and *scale* information. *Physical* information refers to the ink (or pixels) shown, excluding numeric and category labels (i.e. numbers on the axes and labels in the key). Such labels provide *scale* information and assign numbers (in the case of quantitative variables) and labels (for categorical variables) to the *physical* information in the display. According to Cleveland's (1993) model, graphical perception involves two mental operations: (i) *pattern perception*, which refers to the decoding of *physical* information, and (ii) *table look-up*, which refers to the decoding of *scale* information. *Pattern perception* in turn involves three visual operations: (i) *detection*, the recognition of physical elements, (ii) *assembly*, the grouping of elements belonging to the same category, and (iii) *estimation*, the *comparison* of visual elements.

Psychological principles

Pattern perception is governed by general principles of cognition; these help explain how humans decode visual information and thus inform graph construction. Kosslyn (2006) formulates eight psychological principles of effective graph design. These include the audience-oriented principles of *relevance* (show only relevant information) and *appropriate knowledge* (take into consideration the prior knowledge of the audience). Concerned with the visual appearance of the graph are the principles of *salience* (prominent elements receive more attention), *discriminability* (elements have to be sufficiently different to be distinguishable) and *perceptual organization*. The last set of Kosslyn's principles focuses on communication and includes the principles of *compatibility* (form must match content), *informative changes* (changes in form must signal changes in content) and *capacity limitations* (do not overload your audience's working memory). Of particular importance is the principle of perceptual organization, which includes the notion of pre-attentive attributes of stimuli, which affect *detection* and discriminability, and Gestalt laws of perception (Wertheimer 1938; Ware 2013: 181–199). The latter facilitate *assembly* – that is, the selective perception of entities belonging to the same group.

Gestalt laws include the law of *similarity* (similar elements will be grouped together), *proximity* (close elements will be grouped), *good form* (regular or symmetric shapes are perceived as single units) and *connectedness* (linked elements will be grouped; Palmer & Rock 1994).

Theoretical insights into graph design and perception provide a useful foundation for the informed application of statistical graphs in quantitative research. As such, they can guide the choice between different graph types and design options for the display of a particular data set.

3. The dot plot

The dot plot was introduced by Cleveland (1984) as a graphical display of labeled data. Figure 1 shows a simple dot plot of the relative frequency of the 10 most frequent nouns in the *British National Corpus* (BNC; Leech et al. 2001). The horizontal scale encodes a quantitative variable (frequency), the vertical scale a categorical variable (noun); light horizontal lines connect the data points with their labels. Labeled data – that is, numeric values with labels – are common in data analysis. They occur in the form of raw data (e.g. individual measurements or counts in a corpus), summary statistics (e.g. measures of central tendency/location and dispersion/spread, percentages or other effect sizes) and model parameters (e.g. regression coefficients and information criteria). Dot plots can therefore be put to use in a wide range of descriptive and inferential analyses.

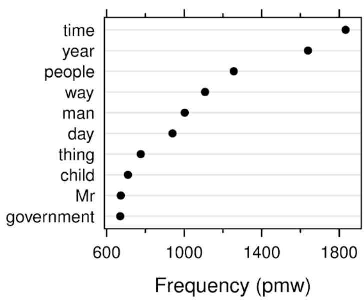

Figure 1. A simple dot plot showing the relative frequency of the 10 most frequent nouns in the BNC; data from Leech et al. (2001)

Simple dot plots can be extended. Figure 2 illustrates a number of additional features and defines the relevant terminology (largely borrowed from Cleveland 1994: 21–22). The data are from a study comparing British (BrE) and American English (AmE) newspaper texts regarding the preference for (orthographically) regular verb forms (e.g. *learned* vs. *learnt*) in simple past and past participle contexts (Levin 2009).

The *main panel* compares two groups (BrE vs. AmE) using different *plotting symbols*. These are *superposed* – that is, plotted on the same line – and labeled in the *key* at the top (arranged to match the major pattern in the plot). *Data labels* on the vertical scale list the verbs, which are ordered by relative frequency in BrE, increasing from bottom to top. The (optional) *appended panel* on the right expresses the *comparison* between BrE and AmE directly by plotting the differences. A *reference line* marks zero, which signals no difference, a relevant reference value. *Tick marks* point outward and are also drawn at the top to facilitate *table look-up*. Difference estimates are indicated by filled circles and include *error bars* showing 95% confidence intervals as a measure of statistical uncertainty. Error bars are explained in the *scale label*.

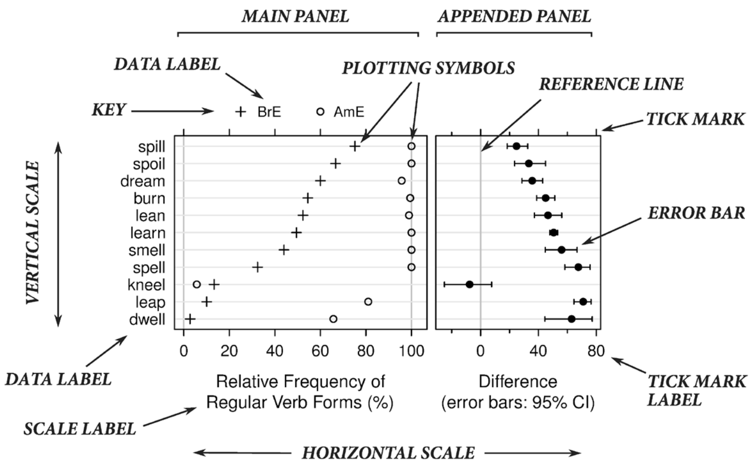

Figure 2. Elements of the dot plot: Terminology and style of presentation borrow heavily from Cleveland (1994); data from Levin (2009)

4. Advantages

The most common graphical display of labeled data is the bar chart, which has three variants: the simple, grouped, and stacked bar chart. It can be replaced by the dot plot in many of its established uses, which often produces a more effective display. This section discusses advantages of dot plots over bar charts.

Aesthetic minimalism

One of the principles of graphical design outlined by Tufte (2001) is the minimization of redundant visual information. Redundancy is expressed with the data-ink ratio, the ratio of “the non-erasable core of a graphic” to “the total ink used to print the graphic” (2001: 93). Using a single prominent symbol to show the data, the dot plot avoids superfluous visual elements. While there is no empirical evidence for the superiority of a high data-ink ratio (Spence 1990; Gillan & Richman 1994; Siegrist 1996), eliminating redundant ink yields a less cluttered graph and thus clear vision. Especially in multivariate displays this is an advantage over grouped or stacked bar charts. Figure 3 shows two variants of a grouped bar chart of Levin’s (2009) results, both of which produce a more cluttered display compared to Figure 2.

Horizontal format

By convention, quantities are often plotted vertically. A horizontal orientation, however, yields four practical advantages: (i) the data labels are shown horizontally and are thus easy to read; (ii) long data labels do not require abbreviations or rotation (cf. Figure 3), which may slow down or even interfere with the decoding of information; (iii) the display can be extended comfortably to show a large number of data points (cf. Figure 9); (iv) the amount of (vertical) space needed for the graph can be reduced without affecting the resolution of the display. The horizontal format, however, yields a number of important limitations of this display method (see Section 7).

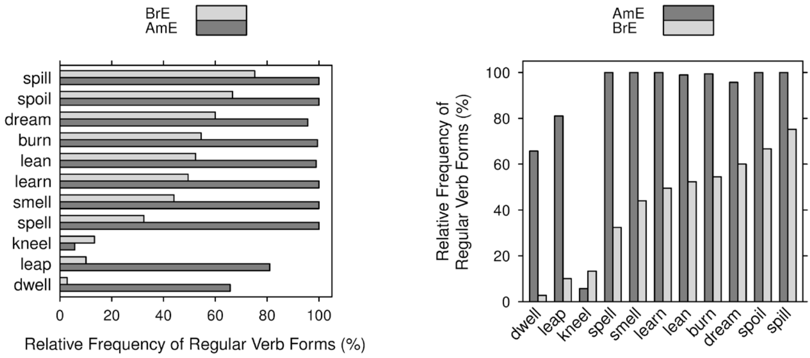

Figure 3. The data from the main panel in Figure 2 shown as a horizontal and vertical grouped bar chart

Resolution

Dot plots offer several benefits in terms of resolution. The issue of axis scaling – that is, whether scaling to zero is necessary – has received much attention in the literature. In science, there appears to be consensus that excluding zero from the scale is often desirable. "Zooming in" by rescaling can greatly enhance the resolution of a display and thus facilitate perception of the variation in the data (Tukey 1977: 51; Cleveland 1994: 92; Wainer 1997, 2009). On the other hand, it has also been argued that such rescaling is inherently misleading (e.g. Huff 1954; Krämer 2001). However, this partly depends on the type of display chosen (Robbins 2005). Bar charts use position (end of bar) as well as size (length and area) to encode numeric values. Without a baseline of measurement, the length of a bar encodes meaningless information and indeed provides misleading visual cues by exaggerating actual differences. Dot plots only use position; the distance to the left side of the graph is further de-emphasized by drawing light horizontal lines across the graph.

The resolution of a graph can be greatly affected by skew, where – due to a few outliers – most data points are crammed into a small part of the graph. Two remedies are data transformation and the use of a scale break. In contrast to conventional scale breaks (i.e. two short parallel lines intersecting the axis), a full scale break divides the graph into two panels, each with a full frame and its own scale (Cleveland 1984). The visually salient discontinuity arguably discourages pattern or continuity perception across the break. The more frequently applied strategy, however, is data transformation. Logarithms are a particularly useful tool when the data are skewed towards large values or when relative differences are of interest. When graphing on a logarithmic scale, dots should be used; the length of bars would provide meaningless information since a log scale has no logical baseline or origin. Figure 4 illustrates the use of a full scale break and a logarithmic transformation to the display of the 10 most frequent verbs in the BNC (Leech et al. 2001). Due to the dominance of the primary verbs (*be, do, have*), even the log transformation does not contribute much to our assessment of the variability among the lower-frequency verbs. In this case, the use of a full scale break helps.

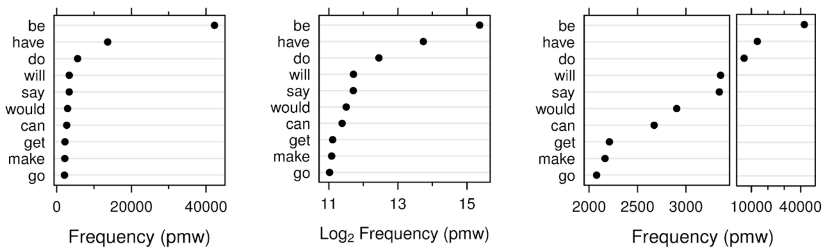

Figure 4. The 10 most frequent verbs in the BNC shown on the original scale, a $\log_2$ scale and using a full scale break; data from Leech et al. (2001)

Error bars

In many cases it is useful for point estimates to include information on statistical variation (Wilkinson & Task Force on Statistical Inference 1999). Such measures are typically indicated with error bars, which may denote different types of information (e.g. standard deviations, standard errors, a confidence interval or a percentile interval). An advantage of dot plots over bar charts is the fact that they produce a more effective presentation of error bars (Cleveland & McGill 1984; Schnell 1994; Wainer 2009). Figure 5 provides three displays of the difference scores that are shown in the appended panel of Figure 2. The principle of *discriminability* and the Gestalt law of *similarity* facilitate *detection* and *assembly* in the plot on the left. Point and interval estimates are visually discriminable, which makes it easy to focus on one type of estimate while mentally muting the other. In bar charts this is more difficult due to the similarity of geometric elements (right-angled linear segments with the same orientation). Minimization of ink adds *salience* to error bars and point estimates, which further facilitates *comparison* and assessment of the variability between verbs. The estimates for each verb are also more easily perceived as a single visual unit due to their point and axis symmetry (Gestalt law of *good form*).

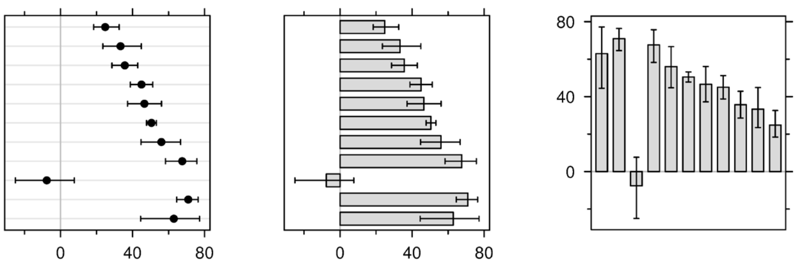

Figure 5. Error bars: Dot plot vs. horizontal and vertical bar chart

Interval scales

Quantitative variables are divided into interval- and ratio-scaled measures, depending on whether there is an absolute zero. While ratio variables only take on non-negative numbers, an interval scale allows positive and negative quantities (e.g. difference scores and correlation coefficients). Bar charts are ill-suited for interval scales, especially if positive and negative values occur in the same plot (cf. Figure 5). As pointed out by McElreath (2016: 203), the only information added by bars – at the expense of a more cluttered display – is "which way to zero". Moreover, error bars extending beyond zero yield an odd appearance (cf. Figure 5). The lengths of bars also encourage ratio comparisons ("A is about twice as large as B"), which may not be warranted on interval scales (e.g. in the case of correlation coefficients).

Pattern perception

When the plotted categories are grouped, the dot plot usually outperforms the divided and grouped bar chart. A comparison of Figures 2 and 3 shows that grouped bar charts quickly become cluttered, which interferes with *pattern perception* (Robbins 2005). In dot plots, successful superposition facilitates Gestalt-like perception of the groups – that is ,they can be visually assembled while mentally filtering out the other elements (Cleveland 1994). Bertin (1983: 67) calls this "selective perception", noting that "[t]he eye must be able to isolate all the elements of [a] category, disregard all the other signs, and perceive the image formed by the given category".

Estimation

Experimental research into graphical perception has identified a number of elementary perceptual tasks that are used to visually extract quantitative information from a display. Examples of such tasks are position, length, angle and area judgments. The visual decoding of dot plots involves position judgments along a common scale. This elementary perceptual task produces more accurate *estimation* than length or angle judgments, which are used in decoding bar charts and pie charts, respectively (Cleveland & McGill 1984). However, performance differences between position, length and angle judgments may be relatively small (Carswell 1992).

In sum, several arguments suggest that the widely used bar chart can in many cases be constructively replaced by a dot plot.

5. Design

There are different options for the design and extension of dot plots. This section illustrates a number of add-ons and discusses construction principles aiming to optimize the resulting display. While such fine-tuning is particularly important for presentation graphs, most of the following considerations are also relevant for the use of dot plots in data analysis.

Order

Location in the two-dimensional space is a powerful visual cue and can be attended to easily and selectively (Kubovy 1981). Dot plots should thus make use of the vertical dimension by (re-)ordering categories or groups in specific ways. If the categories have no logical arrangement, data-based ordering (often according to value or size) facilitates information processing and reveals additional structure in the data (Bertin 1983; Schmid 1983; Wainer 1997). This also applies to multipanel displays and the use of *superposition,* where different options for ordering exist. Ordered symbols are more easily perceived as belonging to the same group (Gestalt law of *proximity* and *good form*). The data analyst should try different arrangements to foreground different gestalts and comparisons, which is likely to uncover different aspects and patterns in the data.

Multipanel conditioning and superposition

Additional categorical variables can be incorporated into dot plots by means of superposition or juxtaposition. In essence, these are different plotting strategies for the comparison of (sub-)groups. While superposed groups are shown in the same panel (cf. Figure 2), juxtaposition involves the use of multiple panels to plot subsets of the data (cf. Figure 11). In general, multipanel conditioning is a powerful method for the display of multivariate data sets (Becker et al. 1996). It is important to note that the two strategies are complementary approaches to the display of multivariate data sets. In general, however, superposition facilitates comparisons *between* groups; juxtaposition, on the other hand, strives for clear vision and allows for better comparison *within* groups. When the number of groups is small, superposition may be more effective than multipanel conditioning (Sarkar 2008).

Plotting symbols

The choice of plotting symbols should allow for good visual detection and assembly. In a simple dot plot, filled circles (●) are recommended since they are salient and combine well with error bars. If two groups are compared in the same panel, the choice of plotting symbols depends on whether overplotting occurs. When there is no overplotting, filled and empty circles (● ○) are a good choice. In the case of overplotting, empty circles and crosses (○ +) allow for better pattern perception (Cleveland 1994). Their distinct pre-attentive attributes ensure excellent texture discrimination (Malik & Perona 1990). Ease of detection and assembly allow the viewer to focus on one group while backgrounding the other. Moreover, salient filled circles (●) can then be used in appended panels to directly show key comparisons (cf. Figure 2). Empty circles and crosses (○ +) may also serve as iconic symbols, signaling presence/absence of a certain attribute (cf. Figure 8). In addition, other symbols may make sense. Letters, for instance, make it easier to remember the groups or categories shown, saving time that would otherwise have been spent looking back and forth between key and data (cf. Figure 8). However, the set of symbols should still be sufficiently discriminable (on the discriminability of graphemes see Lewandowsky & Spence 1989).

Appended panels

A particularly useful add-on for dot plots are appended panels (cf. Amit et al. 2008; see also Heiberger & Holland 2015: 566). Despite its superficial similarity, this plotting strategy is conceptually different from multipanel conditioning. Appended panels do not display a different subset of the data, but rather add more information on the data set plotted in the main panel. While there are many possible uses for appended panels, they seem particularly valuable for directly showing focused comparisons between two groups in the main panel. Such comparisons can be expressed using various types of effect sizes such as difference or ratio measures. Alignment along a common scale makes it much easier to compare effects across categories on the y-axis (e.g. the different verbs in Figure 2). Since different effect size measures may offer different perspectives on the same comparison, it may make sense to append more than one panel (cf. Figure 8). Inferential information can be added to effect size estimates in the form of confidence intervals, which indicate the degree of uncertainty associated with the estimates shown (see Figures 2, 9 and 10).

Error bars

Several options exist for the design of error bars, differing in the way interval limits are marked and as to how many intervals are shown for each point estimate. Figure 6 shows several variants. The most widely used type of error bar is single-tiered with the upper and lower limit marked by crossbars. The use of crossbars has met with criticism since it draws attention to the endpoints of the interval. At any rate it appears reasonable to limit crossbar length to the diameter of the plotting symbol of the point estimates. Error bars may also display several intervals for the same estimate. Cleveland (1994), for instance, suggested two-tiered error bars for showing different confidence levels. Outer tiers may also be used to add interval limits that are adjusted for multiple comparisons (Tukey 1993; cf. Figure 9). As illustrated in Figure 6, inner intervals can be delimited using crossbars (cf. Cleveland 1985: 226), line width (cf. Gelman & Hill 2007: 497) or shading (cf. Harrell 2015: 282). While this appears to be a matter of taste, the use of different line types (more specifically, solid and dashed lines) should be avoided as dash patterns may lead to minor inaccuracies in the boundary locations of the inner and outer tiers (see Kastellec & Leoni 2007: 759 for an example).

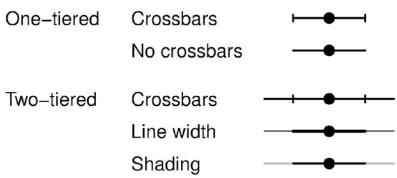

Figure 6. Design options for one-tiered and two-tiered error bars

Dodging

When adding error bars to panels with superposed plotting symbols, overlap is an issue. A simple strategy is to use vertical displacement (see middle panel of Figure 11). In Wickham's (2013) *ggplot2* package for R, this strategy is called "dodging". Plotting symbols and error bars are relocated above and below the light horizontal line, which avoids error bar overlap while still preserving assembly and pattern perception.

Logarithms

Logarithms are a very useful tool for data analysis and visualization. Plotting on a log scale shows relative rather than absolute differences. While logarithms can be expressed using different log bases, it is important to note that the choice of base does not affect the *physical* information in the plot: the same pattern occurs regardless of whether log base 2, *e* or 10 is used (these are typical choices). What changes is the *scale* information, that is, the tick mark labels. The base should be chosen to facilitate *table look-up*. This includes recovering the original values of the points plotted and, more importantly, making comparisons, that is, *estimating* the relative difference between two points plotted. The viewer can be assisted in making these judgments by adding original units to the tick marks at the top (cf. Figure 9).

Lines and color

If more than two groups are superposed in the same panel, it becomes increasingly difficult for the viewer to *detect* and *assemble* groups. *Pattern perception* may then be facilitated by using color or linking the points with lines (Gestalt law of *connectedness*). With the addition of lines, the graphical display approaches the fuzzy category boundary to parallel coordinate plots (see Unwin 2015: 99–130) and line plots (sometimes called interaction plots). The use of lines is further discussed in Section 7.

6. Applications

This section will illustrate the application of dot plots to different types of descriptive and inferential analyses of linguistic data, demonstrating most of the design options discussed above.

Meta-analysis

The term meta-analysis refers to a set of techniques for combining evidence from different studies on the same or similar issues (Cumming 2012). Graphs play an important role in meta-analysis. A frequently employed display type is the forest plot (Lewis & Clarke 2001), which allows the researcher to visually assess effect estimates and confidence intervals reported in the literature. It thus provides a graphical synthesis of the empirical evidence available (see Borenstein et al. 2009 for many examples). Proper meta-analyses also condense the evidence into a single effect size estimate with a (usually much narrower) confidence interval. A simple visual summary, however, is a useful starting point since it allows for a contextualization of new findings, yielding a more solid basis for their interpretation (recall Tufte's quote on quantitative reasoning; Section 2). Forest plots are in fact very similar to dot plots but include a few additional features such as the variation of the size of plotting symbols to signal the degree of uncertainty associated with a particular point estimate (see Lewis & Clarke 2001; Cumming 2012).

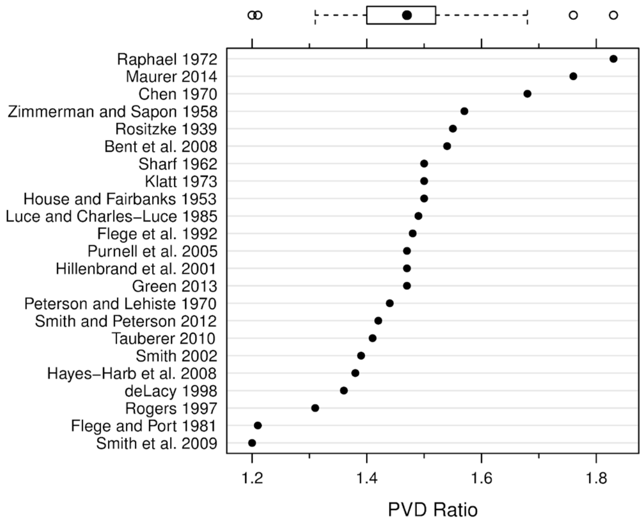

Figure 7. A visual summary and comparison of the results from different studies on the same phenomenon (PVD effect in American English)

Figure 7 summarizes empirical estimates of the PVD (preceding vowel duration) effect in American English. Minimal pairs differing in final obstruent voicing (e.g. *bad-bat*, *peas-peace*) are primarily distinguished by the duration of the preceding vowel, which is longer before voiced obstruents (*bad*, *peas*). This PVD effect can be expressed as a duration ratio. Figure 7 shows the estimates obtained in 23 studies in increasing order from bottom to top. A boxplot has been added to show the distribution of the values. The literature is in fairly good agreement that the ratio ranges somewhere between 1.4 and 1.5. A few studies have reported particularly high or low values, which would prompt us to study their methods section in more detail to identify possible confounding variables.

Figure 8 shows another application of the dot plot to a simple research synthesis. It gives an overview of the empirical evidence on the voice onset time (VOT) of voiced and voiceless stops in American English. VOT, the duration of the interval between stop release and onset of voicing, is an acoustic correlate of the voicing distinction in initial stops. The main panel shows the measurements reported in each study, ordered by the overall average VOT, increasing from bottom to top. Letters (more precisely: IPA symbols) serve as plotting symbols, which facilitates *table look-up*. A reference line is included at zero, an important reference value for these data. The box plots above the main panel compare the distribution of voiced and voiceless consonants. While there is little variation across studies regarding voiced stops, VOT measurements for voiceless stops differ drastically. Further, it is obvious that VOT varies systematically with place of articulation: velar stops /k,g/ show the largest, bilabial stops /p,b/ the smallest values.

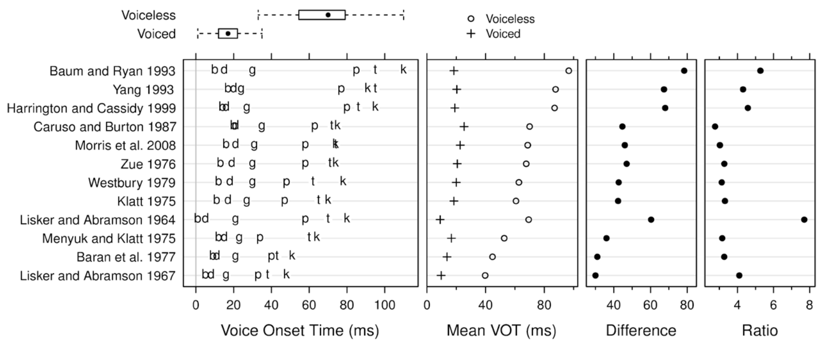

Figure 8. A visual summary of the results of different studies on the same phenomenon (VOT in American English stops)

Three appended panels provide different perspectives on the data in the main panel. The second panel shows the average VOT for voiced and voiceless stops. The plotting symbols indicate the presence (+) or absence (○) of voicing (i.e. the feature [±voice]). The difference in variability between the two categories is even more apparent in this display. To directly express the effect of [±voice] on VOT, difference or ratio scores may be used. These facilitate the comparison across studies and are shown in the appended panels further on the right. To increase resolution, the x-axis does not include zero in these two rightmost panels. It is clear that difference measures covary with overall average VOT, which in turn mostly reflects the VOT in voiceless stops. Ratio measures appear to somewhat control for this effect and may thus be the preferred measure for comparing results across studies. The panels on the right also force the viewer to note that one study clearly sticks out – a finding in need of an explanation.

Corpus data analysis

There are two types of data that frequently arise in corpus linguistics: binary and frequency outcomes. While frequency data reflect the number of occurrences of an event (e.g. word) during a period of observation (e.g. a text or a corpus), binary data stem from variables comprising two categories (or levels), such as regular vs. irregular verb form. Characteristically, corpus-based studies involve two types of comparisons. Commonly, researchers contrast (sub-)corpora representing different varieties of language (such as spoken vs. written) or populations of speakers (e.g. native vs. non-native). On the other hand, it is also typical to investigate several items (lexemes or constructions of any kind). We may therefore distinguish between the comparison of groups and items.

Figure 9 shows an application of the dot plot to the analysis of corpus frequencies (counts). The data are from Granger & Paquot (2008), a study on verb usage in learner and expert academic writing. Counts from two corpora representing non-native and native speaker academic writing were compared. The plot shows the "top 50 underused" verbs in learner academic writing, which were selected based on the likelihood ratio test statistic. There are two types of comparisons: between groups (learners vs. experts) and items (verbs).

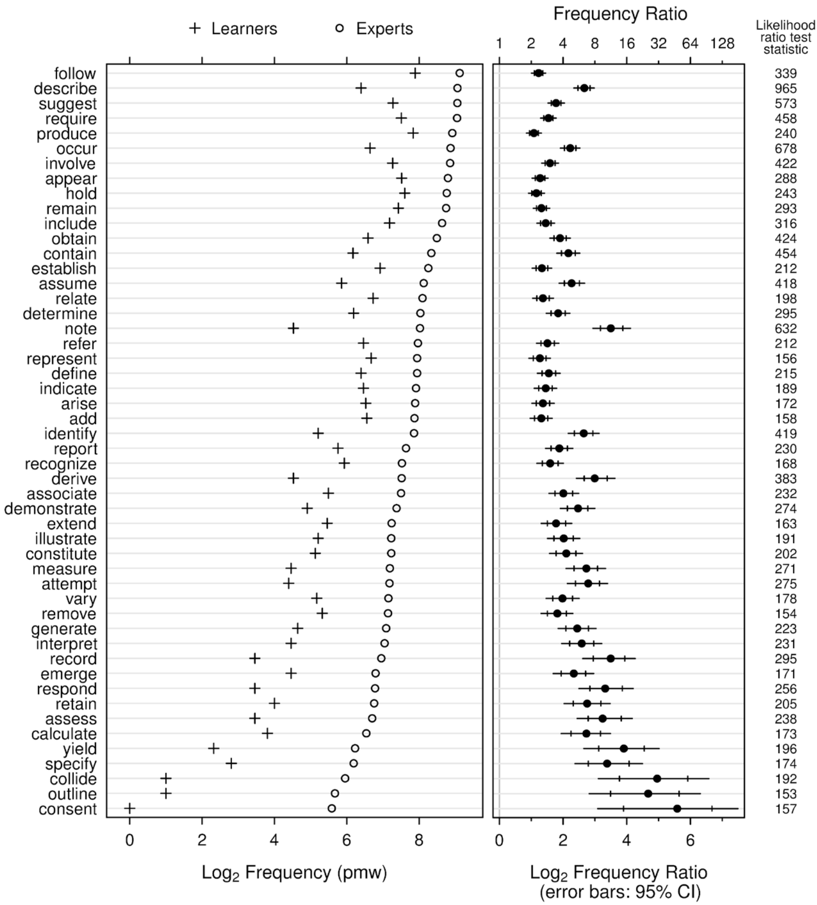

Figure 9. Corpus frequencies: Underrepresented verbs in learner academic writing; data from Granger & Paquot (2008). Inner error bars show individual 95% CIs; outer error bars show 95% CIs adjusted for multiple comparisons

The main panel shows the relative frequency estimates (per million words) for the verbs, which are ordered by their frequency in expert academic writing. Frequency is shown on a $\log_2$ scale, which ranges from 0 (2^0 = 1 pmw) to about 9 (2^9 = 512 pmw). The appended panel shows the $\log_2$ frequency ratio (more precisely: the logarithm (to base 2) of the ratio of the absolute frequencies), which expresses the degree of underuse in learner writing. These ratio scores are shown on a logarithmic scale, and also translated to the original ratio measures by adding the respective tick mark labels at the top. Most verbs are around 2 to 8 times more likely to occur in native speaker expert writing. A handful of verbs are severely underrepresented in learner writing (e.g. *collide, outline, consent*). Information on statistical uncertainty is added in the form of two-tiered error bars. While the inner tiers (delimited by crossbars) show individual 95% confidence intervals, the outer tiers show 95% CIs adjusted for multiple comparisons using the Bonferroni procedure (i.e., showing $1 - \alpha/50 = 99.9\%$ CIs). The likelihood ratio test statistics, which are added at the right margin of the plot, show that comparisons based on such measures may miss important information in the data (such as the underrepresentation of *collide, outline* and *consent*, which have relatively low test statistics due their sparse occurrence, especially in learner writing).

Figure 10 shows an application of the dot plot to the analysis of binary outcomes. The data are from a study by Mondorf (2009) on the variation between synthetic and analytic comparative forms of adjectives in British and American English newspaper writing. Adjectives may form the comparative synthetically with an inflectional suffix (*prouder, purer*) or analytically with *more* (*more proud, more pure*). Mondorf (2009) investigated differences between the varieties in the preference for a particular formation strategy in a number of monosyllabic adjectives. This study thus also involves two types of comparison: between groups (BrE vs. AmE) and items (adjectives).

The results for 15 adjectives are shown in Figure 10. The main panel plots the percentage of analytic comparatives; items are ordered by their relative frequency in AmE, increasing from bottom to top. The appended panel shows the difference in relative frequency between AmE and BrE. There appears to be a bipartition into adjectives with predominantly analytic comparatives (at the top) and those preferring a synthetic form (towards the bottom). Except for *free* and *true,* AmE always shows a stronger tendency towards analytic comparatives. The absolute difference in relative frequency typically ranges from 0 to 20%, with *sour* being a notable outlier (a difference of almost 60% in absolute terms). Reference lines mark the limits of 0 and 100 in the main panel and the reference value of 0 in the appended panel, which denotes equal distribution of synthetic/analytic forms in the two varieties.

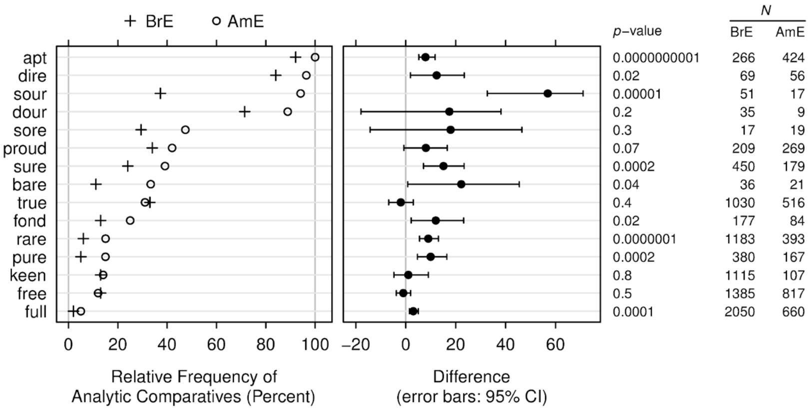

Figure 10. Analysis of a binary outcome: Synthetic vs. analytic comparatives in British vs. American English; data from Mondorf (2009)

P-values from corresponding likelihood ratio tests are added at the right margin; they are in good agreement with the 95% confidence intervals shown in the appended panel. The *p*-values have been rounded up to one significant digit, which produces a semi-graphic representation similar to a stem-and-leaf display (Tukey 1977). Note the unusually large difference between British and American English for *sour*. This extreme divergence is not directly reflected in its *p*-value since this type of measure conflates effect magnitude and sample size. The absolute counts for each adjective are shown to the right of the plot. These correlate with the widths of the confidence interval. Clearly, the corpus contained relatively few tokens of *sourer/more sour*. Thus, while *p*-values collapse effect and sample size into a single measure, effect sizes with confidence intervals allow the researchers to compare and interpret both measures, effect magnitude and statistical uncertainty. This is a strong argument for the preference of confidence intervals over p-values (see also Gardner & Altman 2000).

Application in statistical modeling

Graphical methods play an important role in statistical modeling. Especially in multivariate models, it is difficult for the analyst (as well as the audience) to make sense of tables of coefficients, which are the default output of most statistical software. Indeed, as noted by McElreath, statistical models have "terrible people skills" (2016: 232). Among many other graph types, dot plots have emerged as a particularly suitable aid in model understanding and comparison. As such, model-derived quantities that are translated into graphical form include regression coefficients, test statistics and information criteria. Rather than discuss particular examples of dot plots in statistical modeling, this section will hint at a range of applications and include textbook references for further study.

One strategy is to plot regression coefficients with their associated measures of statistical uncertainty (see Kastellec & Leoni 2007: 765; McElreath 2016: 375, 401). This strategy foregrounds the effect of the predictor and prompts the viewer to compare coefficients rather than *p*-values (or asterisks). If input variables differ in level of measurement and scale, this raises the issue of comparability. For least squares regression, corrective actions include the standardization of regression coefficients (Fox 2016: 100–102; see also Gelman 2008). In logistic regression models, dot plots can be applied for the comparison of predictors based on regression coefficients (Gelman & Hill 2007: 306), odds ratios (Harrell 2015: 282), test statistics (Harrell 2015: 280), and average predictive comparisons (Gelman & Hill 2007: 466–473).

Further comparisons in statistical modeling involve quantities derived from different models for the same data. Such differences may arise from the number of predictors included (Gelman et al. 2013: 423; McElreath 2016: 202) or the fitting of mathematically and/or conceptually different models (Gelman & Hill 2007: 202, 473; Gelman et al. 2013: 400). Further, graphical displays may be used to compare subgroups (Gelman & Hill 2007: 338) or serve as an aid to model comparison using information criteria (McElreath 2016: 199).

7. Limitations

Like other graph types, dot plots have limitations that need to be considered when choosing between different chart types.

Unfamiliarity

One obstacle the dot plot faces is its unfamiliarity to most viewers, which may violate the principle of *appropriate knowledge*. Recognition of the graph type is a critical step in the processing and comprehension of a graphical display. As Kosslyn (1985: 507) notes, "if one has never seen a display type before, it is a problem to be solved – not a display to be read". While the use of dots to encode numerical values in simple displays should pose no problems, more elaborate constructions including superposition and multipanel conditioning may be more demanding for certain audiences. The use of dot plots thus requires reflection on the "graphicacy" (Keen 2010) of the intended audience as well as the time available for graph comprehension (principle of *capacity limitations*). The limitation of unfamiliarity, however, does not apply to the application of dot plots in data analysis.

Cognitive fit

A limitation that applies to the application of the dot plot in data analysis and presentation is the issue of cognitive fit between graph and data (Vessey 1991): the type of display chosen should be compatible with the type of information shown (principle of *compatibility*). Since dot plots show categories on the vertical axis, they are ill-suited for depicting independent variables that are by convention shown on a horizontal axis. Examples are time series and quasi-time differences, such as time trends or variables reflecting age groups, developmental stages or pre- and post-test scores. Figure 11 shows results from an experimental study on plural overregularization in English children's production of irregular plural nouns, for instance **mouses* instead of *mice* (Ramscar et al. 2013). The researchers hypothesized that training on regular plurals would lead to an increase in overregularization in younger children, but to a decrease in older children. The degree of overregularization was measured before (pretest) and after training (posttest) on a scale from -1 (no overregularization) to 1 (overregularization). The bar chart in Figure 11 (left panel) resembles the graph used by the authors to display their results. Due to the interval scale, a bar chart is less suitable (the origin at -0.6 is arbitrary). The dot plot in the middle is a first attempt at producing a more satisfactory display (Sönning 2014) but fails to clearly communicate the experimental results. Since these data involve change over time (as a result of training), the two time points (pretest and posttest) should be shown on the horizontal axis (principle of *compatibility*), which could be conveniently achieved by a line plot, for instance (see right panel of Figure 11). Line plots have a further advantage over other chart types: *assembly* can be greatly assisted by the use of lines (Gestalt law of *connectedness*) and *table look-up* is facilitated by direct labeling of these lines (Gestalt law of *proximity*). As a result, there is no need for a key, which accelerates the decoding of information (Milroy & Poulton 1978; Parker 1983, cited in Pinker 1990: 114).

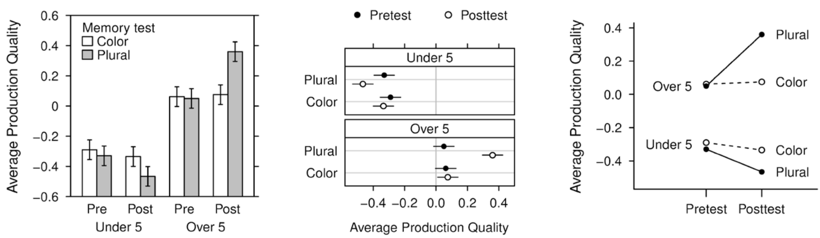

Figure 11. Graph types and cognitive fit: Pretest and posttest scores in two age groups (under 5 vs. over 5) and two tests (color vs. regular plural) shown as a grouped bar chart, a dot plot and a line plot; data from Ramscar et al. (2013)

8. Software

The plots in this paper were drawn in *R* (R Core Development Team 2016) using the packages *lattice* (Sarkar 2008) and *latticeExtra* (Sarkar & Andrews 2016). There is a short tutorial on the construction of dot plots using lattice in the online appendix (www.bit.ly/malt-dotplot-lattice). This appendix also includes dot plot templates for Microsoft Excel, which enable the user to easily construct dot plots (including the use of superposition and appended panels) by copy-and-pasting their data into spreadsheet templates (www.bit.ly/malt-dotplot-excel). Of course, using a template means that there are fewer design options than in *R*. A short instruction manual is also provided online (www.bit.ly/excel-dotplot-instructions).

9. Conclusion

In this contribution, I have argued that the dot plot is a flexible tool for visualizing different types of numeric values with descriptive labels: raw data, frequencies, descriptive measures and model parameters. It is able to replace the bar chart in most of its established uses and likely to produce a more effective display of the data. This paper has demonstrated advantages of the dot plot, illustrated principles for its design and extension to multivariate data sets and exemplified their application to quantitative data in linguistics. Dot plots are a useful tool for data visualization. They should be used more often.

References

Amit, Ohad, Richard M. Heiberger & Peter W. Lane. 2007. Graphical approaches to the analysis of safety data from clinical trials. *Pharmaceutical Statistics* 7. 20–35.

Becker, Richard A., William J. Cleveland & Ming-Jen Shyu. 1996. The visual design and control of trellis displays. *Journal of Computational and Graphical Statistics* 5. 123–155.

Bertin, Jacques. 1983. *Semiology of graphics*. Madison: University of Wisconsin Press.

Borenstein, Michael, Larry V. Hedges, Julian P.T. Higgins & Hannah R. Rothstein. 2009. *Introduction to meta-analysis*. Chichester: Wiley.

Carswell, C. Melody. 1992. Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors* 34(5). 535–554.

Cleveland, William S. 1984. Graphical methods for data presentation: Full scale breaks, dot charts, and multibased logging. *The American Statistician* 38(4). 270–280.

Cleveland, William S. 1993. A model for studying display methods of statistical graphics. *Journal of Computational and Statistical Graphics* 3. 323–343.

Cleveland, William S. 1994. *The elements of graphing data*. Summit: Hobart Press.

Cleveland, William S. & Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association* 79. 531–554.

Cumming, Geoff. 2012. *Understanding the new statistics: Effect sizes, confidence intervals and meta-analysis*. New York: Routledge.

Fienberg, Stephen E. 1979. Graphical methods in statistics. *American Statistician* 33(4). 165–178.

Fox, John. 2016. *Applied regression analysis and generalized linear models*. Thousand Oaks: Sage.

Gardner, Martin J. & Douglas G. Altman. 2000. Confidence intervals rather than P values. In Douglas G. Altman, David Machin, Trevor N. Bryant & Martin J. Gardner (eds.), *Statistics with confidence*, 15–27. London: BMJ Books.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27. 2865–2873.

Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical linear models*. Cambridge: Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari & Donald B. Rubin. 2013. *Bayesian data analysis*. Boca Raton: CRC Press.

Gillan, Douglas J. & Edward H. Richman. 1994. Minimalism and the syntax of graphs. *Human Factors* 36(4). 619–644.

Granger, Sylviane & Magali Paquot. 2008. Lexical verbs in academic discourse: A corpus-driven study of learner use. In Maggie Charles, Diane Pecorani & Susan Hunston (eds.), *Academic writing: At the interface of corpus and discourse*, 193–214. New York: Continuum.

Harrell, Frank E. Jr. 2015. *Regression modeling strategies*. New York: Springer.

Heiberger, Richard M. & Burt Holland. 2015. *Statistical analysis and data display: An intermediate course with examples in R*. New York: Springer.

Huff, Darrell. 1954. *How to lie with statistics*. New York: W. W. Norton.

Jacoby, William G. 2006. The dot plot: A graphical display for labeled quantitative values. *The Political Methodologist* 14(1). 6–14.

Kastellec, Jonathan P. & Eduardo L. Leoni. 2007. Using graphs instead of tables in political science. *Perspectives on Politics* 5. 755–771.

Keen, Kevin J. 2010. *Graphics for statistics and data analysis with R*. Boca Raton: CRC Press.

Kosslyn, Stephen M. 1985. Graphics and human information processing: A review of five books. *Journal of the American Statistical Association* 80. 499–512.

Kosslyn, Stephen M. 2006. *Graph design for the eye and mind*. Oxford: Oxford University Press.

Krämer, Walter. 2001. *So lügt man mit Statistik*. München: Piper.

Krug, Manfred, Ole Schützler & Valentin Werner. 2016. Patterns of linguistic globalization: Integrating typological profiles and questionnaire data. In Olga Timofeeva, Anne-Christine Gardner, Alpo Honkapohja & Sarah Chevalier (eds.), *New Approaches to English Linguistics: Building Bridges*, 35–66. Amsterdam: Benjamins.

Kubovy, Michael. 1981. Concurrent pitch segregation and the theory of indispensable attributes. In Michael Kubovy & James R. Pomerantz (eds.), *Perceptual organization*, 55–98. Hillsdale: Erlbaum.

Leech, Geoffrey, Paul Rayson & Andrew Wilson. 2001. *Word frequencies in written and spoken English*. London: Longman.

Levin, Magnus. 2009. The formation of the preterite and the past participle. In Günter Rohdenburg & Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*, 60–85. Cambridge: Cambridge University Press.

Lewandowsky, Stephan & Ian Spence. 1989. Discriminating strata in scatterplots. *Journal of the American Statistical Association* 84. 682–688.

Lewis, Steff & Mike Clarke. 2001. Forest plots: Trying to see the wood and the trees. *British Medical Journal* 322. 1479–1480.

Malik, Jitendra & Pietro Perona. 1990. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A* 7. 923–932.

McElreath, Richard. 2016. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press.

Milroy, Robert & E. Christopher Poulton. 1978. Labelling graphs for improved reading speed. *Ergonomics* 21. 55–61.

Mondorf, Britta. 2009. Synthetic and analytic comparatives. In Günter Rohdenburg & Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*, 86–107. Cambridge: Cambridge University Press.

Palmer, Stephen E. & Irvin Rock. 1994. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review* 1. 29–55.

Pinker, Steven. 1990. A theory of graph comprehension. In Roy Freedle (ed.), *Artificial intelligence and the future of testing*, 73–126. Hillsdale: Erlbaum.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Ramscar, Michael, Melody Dye & Stewart M. McCauley. 2013. Error and expectation in language learning: The curious absence of mouses in adult speech. *Language* 89(4). 760–793.

Robbins, Naomi B. 2005. *Creating more effective graphs*. Hoboken: Wiley.

Sarkar, Deepayan. 2008. *Lattice: Multivariate data visualization with R*. New York: Springer.

Sarkar, Deepayan & Felix Andrews. 2016. *latticeExtra: Extra Graphical Utilities Based on Lattice*. R package version 0.6-28. https://cran.r-project.org/web/packages/latticeExtra/latticeExtra.pdf (30 November, 2016.)

Schmid, Calvin F. 1983. *Statistical graphics: Design principles and practices*. New York: Wiley.

Schnell, Rainer. 1994. *Graphisch gestützte Datenanalyse*. München: Oldenbourg.

Schützler, Ole. Forthcoming. A corpus-based study of concessive conjunctions in three L1-varieties of English. In Isabelle Buchstaller & Beat Siebenhaar (eds.), *Language variation – European Perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8)*, Leipzig, May 2015. Amsterdam: Benjamins.

Siegrist, Michael. 1996. The use or misuse of three-dimensional graphs to represent lower-dimensional data. *Behaviour and Information Technology* 15. 96–100.

Sönning, Lukas. 2014. The dot plot: A fine tool for data visualization. Paper presented at *Advances in Visual Methods for Linguistics* (AVML). University of Tübingen.

Spence, Ian. 1990. Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance* 16. 683–692.

Tufte, Edward R. 1990. *Envisioning information*. Cheshire: Graphics Press.

Tufte, Edward R. 2001. *The visual display of quantitative information*. Cheshire: Graphics Press.

Tukey, John W. 1977. *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Tukey, John W. 1993. Graphic comparisons of several linked aspects: Alternatives and suggested principles. *Journal of Computational and Graphical Statistics* 2. 1–33.

Unwin, Antony. 2015. *Graphical data analysis with R*. Boca Raton: CRC Press.

Vessey, Iris. 1991. Cognitive fit: A theory-based analysis of the graphs vs. tables literature. *Decision Sciences* 22. 219–241.

Wainer, Howard. 1997. *Visual revelations*. Mahwah: Erlbaum.

Wainer, Howard. 2009. *Picturing the uncertain world*. Princeton: Princeton University Press.

Ware, Colin. 2013. *Information visualization: Perception for design*. Amsterdam: Elsevier.

Werner, Valentin & Robert Fuchs. 2016. The present perfect in Nigerian English. *English Language and Linguistics* First View.

Wertheimer, Max. 1938. Laws of organization in perceptual forms. In Willis D. Ellis (ed.), *A source book of Gestalt psychology*, 71–88. London: Routledge & Kegan Paul.

Wickham, Hadley. 2013. *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Wilkinson, Leland & Task Force on Statistical Inference. 1999. Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* 54. 594–604.