DOI: 10.1111/j.1467-8659.2009.01694.x

Eurographics/ IEEE-VGTC Symposium on Visualization 2010
G. Melançon, T. Munzner, and D. Weiskopf
(Guest Editors)

Volume 29 (2010), Number 3

The Perception of Correlation in Scatterplots

Ronald A. Rensink and Gideon Baldridge

University of British Columbia, Vancouver, Canada

Abstract

We present a rigorous way to evaluate the visual perception of correlation in scatterplots, based on classical psychophysical methods originally developed for simple properties such as brightness. Although scatterplots are graphically complex, the quantity they convey is relatively simple. As such, it may be possible to assess the perception of correlation in a similar way.

Scatterplots were each of 5.0° extent, containing 100 points with a bivariate normal distribution. Means were 0.5 of the range of the points, and standard deviations 0.2 of this range. Precision was determined via an adaptive algorithm to find the just noticeable differences (jnds) in correlation, i.e., the difference between two side-by-side scatterplots that could be discriminated 75% of the time. Accuracy was measured by direct estimation, using reference scatterplots with fixed upper and lower values, with a test scatterplot adjusted so that its correlation appeared to be halfway between these. This process was recursively applied to yield several further estimates.

Results of the discrimination tests show jnd(r) = k (1/b – r), where r is the Pearson correlation, and parameters 0 < k, b < 1. Integration yields a subjective estimate of correlation g(r) = ln(1 – br) / ln(1 – b). The values of b found via discrimination closely match those found via direct estimation. As such, it appears that the perception of correlation in a scatterplot is completely described by two related performance curves, specified by two easily-measured parameters.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation / methodology.

1. Introduction

The design of an effective display for data visualization often requires considerable guesswork. For example, to display a particular kind of dataset using a graph, the designer must pick the type of graph, the sizes and types of the graph elements, the scaling of the axes and so on. Often it is not clear which choices are best, and considerable testing must then be done.

This paper investigates this issue for the case of scatterplots. These have been used for over a century as a way to visually represent data [FD05]. Much of their popularity has been due to their ability to allow correlations to be easily perceived by a human viewer (see e.g.,[Cle93, Har99]). But despite the widespread use of scatterplots, relatively little is known about the effect of various design factors on their ability to convey such correlation. It is difficult to compare the relative effectiveness of two designs, or to determine the absolute efficiency of a particular design.

Many careful studies of correlation perception in scatterplots have been carried out (e.g. [BK79, CDM82, LP89, MTF97, Pol60]). Almost all have been based on the direct estimation of the Pearson correlation r, e.g., asking observers to provide a number describing the degree of correlation perceived. (For a review, see [DAAK07]). Several important results have been obtained this way, such as an underestimation of correlation r in the region $.2 < |r| < .6$, and the finding that essentially no correlation is perceived when $|r| < .2$.

However, such results are not enough to constitute a solid foundation. To begin with, the central assumption used in direct estimation—that numbers can consistently be assigned to perceived magnitudes—may be incorrect

[EF00]. Second, even if such estimates do provide a good assessment of *accuracy*, they leave out an important aspect of perception: *precision*. And even if precision could somehow be handled, there still remains the issue of *systematicity*—whether there are general laws describing how both aspects of perception relate to each other. This issue is of more than just theoretical interest: if such regularities exist, performance could be described using relatively few parameters. If so, it might be possible to completely assess a given design using just a few tests.

To examine this issue, we use an approach developed at the beginnings of vision science: assessing how well stimulus properties can be discriminated. This is based on measuring the just noticeable difference (jnd), the difference in properties between two side-by-side stimuli (e.g. squares of differing brightness) that can be discriminated 75% of the time. This was the first step in the development of a rigorous way to study human vision. By adapting this to scatterplots, it may be possible for it to play the same role here.

This approach has several strengths. First, it provides a useful measure of precision, in that the jnds are essentially a measure of this quantity. Note that it can be used as a complement to direct estimation—it need not be a competitor. Also, jnds often have a simple behavior. For example, if p denotes some physical property (e.g. length or brightness), it is often the case that jnd(p) = dp = kp. This is known as *Weber's Law*; it has been found to hold for many physical properties; values of the *Weber fraction* k are typically in the range 0.02-0.08 [CME99].

Under some circumstances a linear jnd can be integrated to yield a psychological estimate P = k log(p), which is known as *Fechner's Law*. This is the case for several physical quantities, such as weight and brightness [CME99]; if it also holds for correlation, it would indicate a direct relationship between precision and accuracy. Moreover, if the psychological estimate is a logarithmic function, it would need only a few parameters for its complete specification. We will therefore test this possibility as well, using direct estimation of the degree of perceived correlation.

It could be objected that this approach is unlikely to work for correlation, in that it is a relatively complex property based on relatively complex stimuli. However, although scatterplots themselves may be complex, the quantity they convey (viz., correlation) need not be so. Indeed, given that correlation can be perceived from scatterplots by almost anyone with just a little training, it may be a relatively simple psychological property. As such, there are at least some grounds for believing that the approach proposed here might succeed.

2. General Methods

Stimuli were scatterplots each of 5° extent vertically and horizontally, containing 100 normally-distributed points. The mean of each was set to 0.5 of its extent, and standard deviation to 0.2. For any target correlation t, the scatterplot correlation r = t ± 0.005.

Each point was created using pseudo-random numbers taken from a gaussian distribution. The x-coordinate was the first number chosen (after appropriate scaling and translation). A y value was then created and transformed using equation 1 to create a correlated pair (x,y'). To avoid points outside the range of the graph, any point greater than 2 standard deviations from the mean was eliminated, and a new point generated to take its place.

$$y' = \frac{\lambda x + (1-\lambda) y}{\sqrt{\lambda^2 + (1-\lambda)^2}}, \text{ where } \lambda = \frac{r - \sqrt{r^2 - r^4}}{2r^2 - 1} \qquad (1)$$

Experiments were run with 20 observers, each seated 57 cm from a screen of extent 30° x 20°. All observers were tested on both discrimination and direct estimation; each observer was tested on all conditions of both tasks. Average age was 24 years. All observers had at least some experience with scatterplots; most had made and used scatterplots on several occasions. Observers were given as much time as needed to complete each task, although it was mentioned that accuracy was important. A small practice run of 50 trials was given to familiarize observers with the task.

2.1. Discrimination

The first test measured precision by determining the sensitivity of observers to differences in correlation. Such assessments are potentially noisy, in that it is the differences that are measured rather than the quantities themselves. (This explains in part why most previous work was based on direct estimation.)

To help combat this noisiness we used a variant of the staircase method commonly used for studies in perception [CME99]. Here, each observer was shown two side-by-side scatterplots—one more highly correlated than the other—and asked to select the one that was more highly correlated (Figure 1). Initial difference of the correlations was 0.1. When a correct answer was given, this was decreased by 0.01, making the task more difficult. For an incorrect answer, it was increased by 0.03, making the task less difficult. To ensure that observers based their responses on the general property of correlation, scatterplots were replaced by new instances each time. This continued until a *just noticeable difference* (jnd) was found, where steady-state accuracy was 75%.

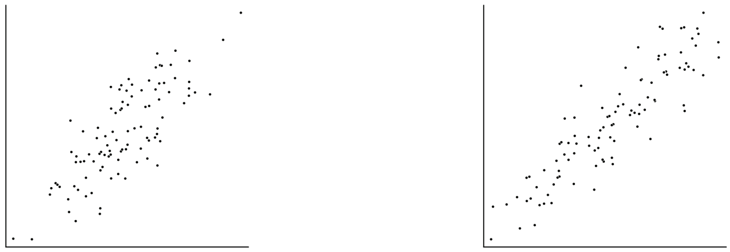

Figure 1: *Example of scatterplots used in the discrimination task. Observers were asked to choose which scatterplot is more highly correlated. In this example, base correlation is 0.8; jnd is from above.*

Performance was measured via a moving window of 24 consecutive trials. This was divided into 3 sub-windows of 8 trials each (Figure 2). For each base correlation an initial set of 24 trials was run. Subsequently, after each trial the average variance within the sub-windows was compared to the variance of the averages of the sub-windows (essentially an F-test). Testing halted when this value reached a sufficiently low level (0.25) or when 50 trials had been run. The average of the sub-windows was then used as the jnd. This proved reasonably effective, yielding results within 36.6 trials on average over all trial, and failing to converge on only 97 of the 380 runs.

The base correlations tested ranged from r = 0 to 0.9, in increments of 0.1. The order was determined by a latin square design [Kir95], which provided counterbalancing across all 20 observers. For each base correlation, differences from both above and below were measured, with one exception: to avoid issues dealing with negative correlation, there was no test from below for r = 0.

2.2. Direct Estimation

The next test measured perception of correlation via direct estimation, the goal being to directly determine the subjective estimate of correlation g(r) as a function of objective correlation r. Traditionally, this has been done by asking for a number describing the degree of correlation perceived in a given scatterplot (e.g [BK79, LMvW08]). Since the use of direct numerical estimates may be problematic, whereas the use of ratios is not [EF00], we used an approach based on bisection. Here, each observer was shown two *reference plots* (one with a high level of correlation, one with a low) along with a *test plot* having a level of correlation between the two reference values. Observers were asked to adjust the correlation of the test plot until it looked like it was exactly halfway between the correlations of the referents. (Figure 3) This was done via keyboard control, with observers free to adjust the correlation however they wished.

To avoid the possibility that observers could somehow base their performance on the number of steps used, each step size was given a random value between 0 and 1/10 of the difference between the reference correlations. As for the discrimination tests, individual scatterplots (both reference and test) were replaced by new instances each time an adjustment was made, requiring observers to base their judgement on the population property of correlation rather any particular feature of any particular instance.

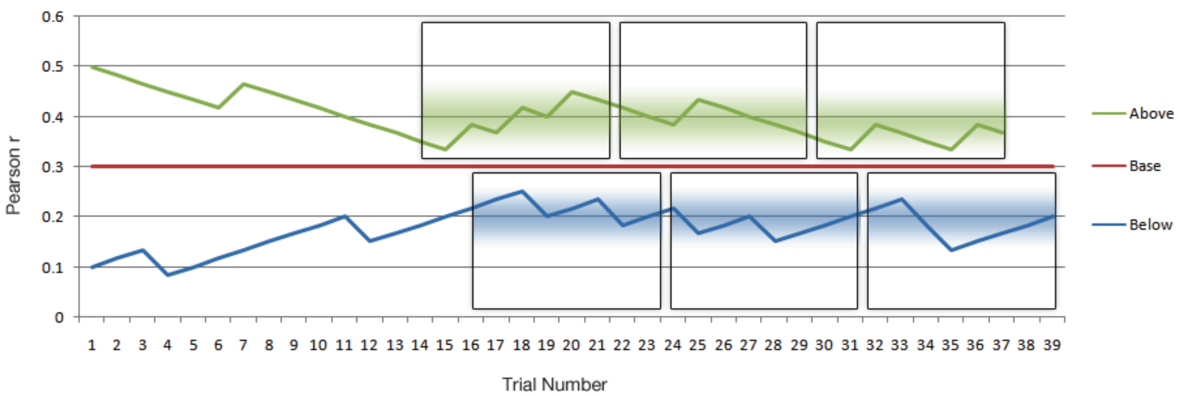

Figure 2: *Schematic of threshold algorithm. The distance from base correlation is adjusted until the variance of the averages of the sub-windows is 0.25 of the average variance within the sub-windows.*

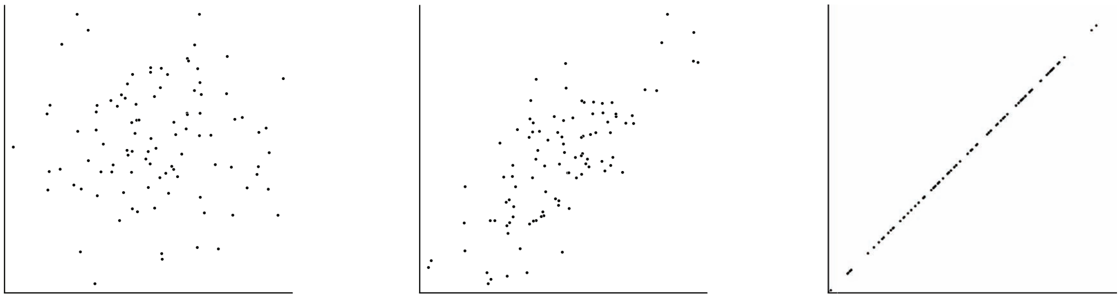

Figure 3: *Example of scatterplots used in the direct estimation task. Observers adjusted the correlation of the central test plot, until its correlation was halfway between those of the two reference plots. Here, adjusted value of the test plot is r=0.74, which corresponds to the subjective midpoint.*

In the first round, observers judged the halfway point between the extremes r = 0 and r = 1. This was done four consecutive times, with the mean of these judgments taken as the value of r for subjective estimate g = 1/2.

The second round applied this method recursively, with each observer again asked to find the value of r that appeared to be halfway between the reference values. Two variants were used. In variant A observers judged the point halfway between g = 0 and 1/2; in variant B they judged the point halfway between g = 1/2 and 1. The order of these was counterbalanced across observers. Again, each judgement was made four consecutive times, with the averages providing the value of r corresponding to subjective estimates g = 1/4 and g = 3/4.

In the third round, this method was again applied to determine the values of r corresponding to the subjective estimates g = 1/8, 3/8, 5/8, and 7/8. The variants at this stage were presented in random order.

After the estimates were measured for each observer, tests were given to ensure that each observer understood the task and tried to do it accurately. First, a screening criterion checked for observers who made adjustments too inconsistently (standard deviation > 0.19) or who had an overall range of estimates < .2. Four observers failed this criterion. They were replaced by new observers who ran all conditions on both the discrimination and direct estimation tasks.

Finally, checks were made on the consistency of the method itself. Each observer re-estimated the point g = 1/2 using as reference pairs their previous estimates of g = 1/8 and 7/8, g = 1/4 and 3/4, and g = 3/8 and 5/8. Re-estimates were also made of g=3/8 using references of 1/8 and 5/8, and of g=5/8 using references of 3/8 and 7/8. These were run in random order.

3. Results

Analysis was based on the average jnd and subjective estimate of all observers. Unless otherwise noted, analysis of fit with empirical data was based on the squares of the errors of the fit, with averages given in terms of root mean square (rms) values.

Unless specified otherwise, comparisons were based on repeated-measures F-tests. (When comparing two quantities, these are equivalent to the use of paired, two-sided t-tests.)

3.1 Discrimination

Average time to make a discrimination between plots was 1.6 seconds. Average jnds for each base correlation -- both from above and below -- are shown in Figure 4. For base correlations of 0.2 or less, the adaptive algorithm encountered a floor effect for jnds from below; these values were omitted from the analysis.

A least-squares fit of the jnd-below data yielded a slope m = -0.20 and y-intercept a = 0.223. The linearity of the data is striking, with R^2 = .978. For jnd-above data over the same range, corresponding values are m = -0.25 and a = 0.282, with R^2 = .964. This is consistent with earlier reports [SH78, CDM82] that precision is greatest at higher correlations. Note that jnd is proportional to the distance of the base correlation from the intersection with the x-axis; in this way, performance can be described in terms of a Weber fraction (–m). Although the size of this fraction is somewhat greater than that for most simple properties, it nevertheless describes jnd to a high degree of accuracy.

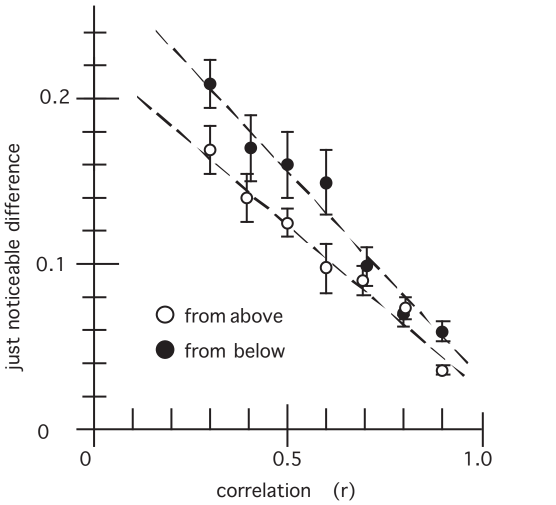

Figure 4: *Jnd as a function of raw correlation r. Error bars denote standard error of the mean*

Although the values for the two kinds of jnd are somewhat similar, they are reliably different (with base correlations as levels, $F(1,6) = 13.4$; $p < ,015$). The existence of a single jnd measure is therefore potentially problematic. To investigate further, the analysis was repeated, but with base correlation r replaced by $r_A = r + 0.5\ jnd(r)$, the average of the two correlations tested. (Jnd-below values were taken to be negative for this.) Slopes and intercepts of both jnd lines are now virtually identical, each with k = -0.22 and a = 0.25 (Figure 5). Thus, the adjusted correlation r_A enables the use of a single well-defined jnd measure.

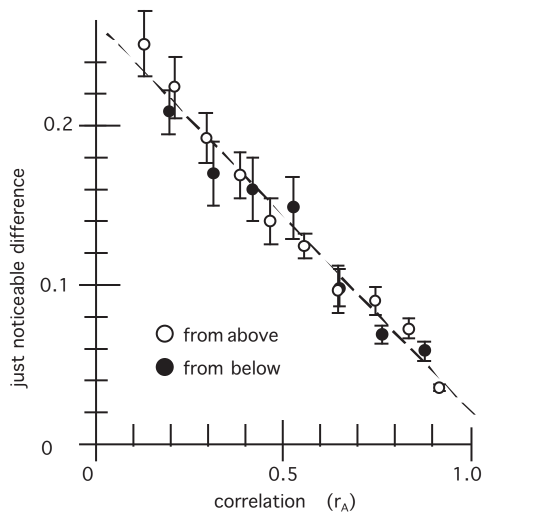

Figure 5: *Jnd as a function of adjusted correlation r_A. Error bars denote standard error of the mean.*

Analysis of all data in terms of r_A shows highly linear behavior over the entire range of correlations (Figure 5); $R^2 = .971$. When raw correlations are used instead, $R^2 = .934$, indicating somewhat less linearity. Therefore, adjusted value r_A will be used here in preference to r as the basis of discrimination analysis.

Given that jnds are proportional to the distance from the intercept of the jnd line with the x-axis, a natural way to describe their behavior is via the formula

$$jnd(r) = k\ (1/b - r_A), \quad 0 < k, b < 1 \qquad (2)$$

where k is the *variability parameter* (or Weber fraction), defined as -m, and b the *offset parameter*, defined as the reciprocal of the intersection of the jnd line with the x-axis. (Defining b this way allows it to have a finite range $0 < b < 1$. It also mitigates the effect of noise in estimates of k, which can cause the intersection point to vary considerably.) For both k and b, smaller values denote better performance, with optimal performance as these values approach zero.

A final refinement is to estimate k and b by minimizing the variance in the estimates of k at each base correlation. More precisely, if for base correlation r_i, $k_i = jnd\ (r_i)\ /\ (1/b - r_{Ai})$ the value of b is that which minimizes the variance of the normalized k_i, i.e., k_i divided by average value (k). This is largely the same as direct least squares, except using ratios rather than absolute differences so that the estimate of k at small r is not as severely affected by noise. Estimation yields k = 0.24 and b = 0.907.

3.2. Direct Estimation

Averages for all observers are shown in Figure 6. Consistent with other reports [CDM82, KM08], severe underestimation of correlation occurs for $0.2 < r < 0.6$. The trend is consistent with two previous proposals: the square of the correlation $g(r) = r^2$ [Pol60, BK79], and the double-power function $g(r) = 1 - (1-r)^{\alpha}(1+r)^{\beta}$, where α and β are free parameters [CDM82]. Both fit the data reasonably well: rms error for the square is 0.03, while for the double-power function it is 0.02.

The consistency checks show few problems with the estimation method. Of the 60 comparisons of original estimates and re-estimates across the 20 observers, only 3 had differences that were significant (i.e., $p < .05$). This lack of effect is unlikely to have resulted from a lack of precision in the method—the average standard error in the estimates of individual observers was only 0.04.

The overall reliability of the data raises the possibility of testing the proposal that accuracy of correlation perception is related to precision. In particular, given that

$$k = jnd(r) / (1/b - r_A) = \Delta r / (1/b - r_A) \qquad (3)$$

it is possible to consider the *Weber assumption*, which postulates that k is proportional to Δg, a unit step in the subjective estimate g of correlation. [CME99]. As such, this can be written

$$\Delta g = C_0 \Delta r / (1/b - r_A) \qquad (4)$$

where C_0 is some constant. As $\Delta r \to 0$, $r_A \to r$, and this becomes

$$dg = C_0 dr / (1/b - r) . \qquad (5)$$

Integration leads to

$$g(r) = -C_0 \ln(1/b - r) + C_1 , \qquad (6)$$

where C_1 is the integration constant. The values of C_0 and C_1 can be determined by imposing the conditions $g(0) = 0$ and $g(1) = 1$, yielding

$$g(r) = \ln(1- br) / \ln(1 - b), \qquad (7)$$

essentially Weber's law for the quantity u = 1-br.

The best fit of this curve to the data is with b = 0.875. This yields in an rms error of 0.02, comparable to that of the other proposals ($F(1,6) = 1.33$; $p > .3$).

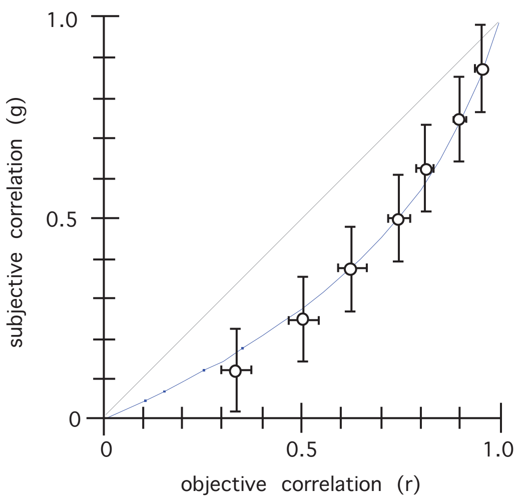

Figure 6: *Results of direct estimation. Vertical error bars show one jnd; horizontal error bars standard error. Curve is g(r) = ln(1–br) / ln(1-b), with b = 0.875.*

Interestingly, the value of the offset parameter b obtained via direct estimation (b = 0.875) is within 4% of the value obtained via discrimination (b = 0.907). The rms difference in estimates using these two values is less than 0.03, with no significant difference in their fits to the data ($F(1,6) = 1.85$; $p > .2$). This further supports the idea that the two *performance curves*—for precision and for accuracy—are systematically related, with precision proportional to the reciprocal of the derivative of the accuracy curve. Moreover, both curves are simple, and are jointly governed by just two numbers: variability parameter k and offset parameter b.

3.3. Individual Variation

The fit and systematicity of these results constitute strong evidence that the precision and accuracy of correlation perception can be described by the functions proposed here. However, it might be argued that this is just a coincidence, and that at least for subjective estimation it is better to stay with the older formulations, since these appear to be equally accurate.

To examine this possibility, the behavior of individual observers was examined. The data from an untrained (and possibly unmotivated) individual is usually too noisy to justify extensive analysis. However, just as individual *data* can be aggregated to reduce the effects of noise, so too can the aggregate of individual *behaviors* be examined for interesting trends, which could help decide which proposal is most suitable.

Consider first the square of the correlation. Because it has no free parameters, it does not give a good fit to individual data: average rms error is 0.132. In contrast, using the log function (and adjusting b) results in an rms error of only 0.057, a reliably better value ($F(1,19) = 25.3$; $p < .0001$).

Next is the double-power function. As with the log function, this can be fit to the results of individual observers. However, despite the presence of a second free parameter, average individual error is 0.052, only marginally different from the error obtained using the single-parameter log function ($F(1,19) = 3.76$; $p = .07$).

A unique aspect of the proposal here is that the subjective estimate g(r) has a close connection to jnd(r), the two sharing the same offset parameter b; no such connection exists in the double-power proposal. To see whether such a relationship might exist in individual behaviors, values of b were calculated for each observer based on both the discrimination and direct estimation data. Average rms difference between the two is 0.29. For the best 50% of observers (defined as those with the least variance in their correlation estimates), this drops to 0.04, and for the best 25%, it is only 0.008. Correlation between the two estimates of b improves similarly: when taken over all observers it is 0.0, for the best 50% it is 0.73, and for the best 25% it is 0.97. Thus, the more capable (and motivated) the observer, the stronger the match between the b values obtained by the two methods. This further supports the proposal that for perception of correlation in scatterplots, precision and accuracy are tightly linked.

4. Conclusions

This study has shown that all important aspects of the perception of correlation in scatterplots—precision as well as accuracy over all correlations—can be described by two related functions governed by two parameters: the variability parameter k and the offset parameter b. In particular, precision is proportional to u = 1 – br, while accuracy is proportional to the logarithm of this quantity. In addition to their systematicity and comprehensiveness, these functions provide a good fit to the data—as good as any existing proposal, and in some cases even better.

4.1 Design Evaluation

According to the results of this study, evaluation of the absolute performance of a given design or comparison of the performance of various design factors (e.g., different dot sizes or densities) requires the determination of only two quantities: k and b. This can be done as follows:

1. Select two or more correlations r_i (e.g., r_1 = 0.4, r_2 = 0.8). Measure the jnd Δ_i for each r_i using the method described in section 2.1 (or equivalent).
2. For each i, let $k_i = \Delta_i / (1/b - (r_i + \Delta_i/2))$.
3. Set b to the value that minimizes the variance of the k_i /k; set k to the average of the k_i.

The performance curves are then obtained by placing k and b into the appropriate functions.

Although two measurements are sufficient in principle, additional accuracy can be gained by using three or more base correlations sufficiently separated (e.g. 0.5, 0.7, 0.9). Using a large number of observers to reduce noise effects will also help. To maximize sensitivity, a within-observer design can be used, where all observers are shown the same designs.

The effects of overall scale, dot size, symbol shape, or any other design factor can be directly tested this way. Evaluation of two competing designs is straightforward: the one with the smallest k and b is best.

Evaluation can also be taken a step further. If p(r) denotes the probability of encountering correlation r in a given task, the average precision of the estimates for a given design is

$$\int p(r)\, \mathrm{jnd}(r)\, dr \tag{8}$$

while average error in accuracy is

$$\int p(r)\, |\, r - g(r)\, |\, dr. \tag{9}$$

An important note: The results here were based on scatterplots with similar horizontal and vertical variance. Since the perception of correlation can vary with these quantities [LAK85], there is a possibility that this method may not apply equally well to all scatterplots. Further testing is needed. However, the effects of any particular design factor can still be evaluated provided that testing is done on scatterplots with equal variances.

4.2 Connection to Perceptual Mechanisms

The goal of this study was to evaluate the overall ability of observers to perceive correlation in scatterplots, and not to investigate the particular mechanisms involved. (Models of these have been proposed elsewhere - e.g., [LiMvW08]) As such, little can be said here about the details of these mechanisms. However, it is worth noting that precision is proportional to the quantity u = 1 – br. This is exactly the same behavior as for the discriminability of several simple physical properties (Weber's law). Similarly, subjective estimation of correlation is well described by the logarithm of u, essentially a form of Fechner's law.

Although the size of the proportionality constant k is larger for correlation than it is for simpler properties, the striking similarity in general form suggests several things. First, the relevant quantity appears to be u. This strongly constrains any model of correlation perception, which must account for why u is based on the distance of br from 1, and why precision is almost exactly proportional to u. In addition, it must be able to explain why the accuracy curve g(r) has a logarithmic form (or equivalently, why the assumption that k is proportional to Δg is valid).

Interestingly, human brain activity during correlation perception increases as correlation is decreased [BHS06], suggesting that the key quantity is the distance from perfect correlation r = 1. The quantity u has this property.

More generally, the results here open up the possibility that the perception of other properties conveyed by graphically complex stimuli (e.g., averages and trends) show a similar log-linear behavior. If so, this would indicate the existence of a class of properties that are relatively simple in terms of perception, yet require several stages of computation for their extraction. It would also suggest that at least some of higher-level visual cognition is based on relatively simple quantities.

4.3 Future Directions

This study is only a first step, showing that the approach developed here is a consistent and useful one. It remains to apply it to various design parameters (e.g., dot sizes, shapes, or colors) to determine how they affect perception. An important variant is to examine scatterplot clouds of different shapes; this could help ascertain how much of correlation perception is based on the shape of the cloud (cf. [MTF97]) and how much on other factors. Issues here include the influence of outliers, and of subsets of data not belonging to the main population.

A key aspect of this approach is that it is general—it is not restricted to scatterplots. It could also be used to evaluate the perception of correlation in bar charts, parallel co-ordinates, or line graphs. In addition, it could be applied to any well-defined property, such as averages or variances.

But a more general point is that the methods developed over the years in vision science can be successfully used to rigorously evaluate visual displays. While a case has been made for closer connections between vision science and visualization (e.g. [CMS99, War04]), this has tended to focus on the perceptual mechanisms engaged by different designs. But methodology may be equally important, providing a solid and systematic foundation for the evaluation of visualization designs. The results of this study provide support for this point of view. It will be interesting to determine the extent to which this kind of connection can ultimately be developed.

Acknowledgements

We would like to thank the reviewers for their comments, and Kyle Melnick and Ben Shear for their feedback. This work was supported by The Boeing Company.

References

[BHS06] BEST, L.A., HUNTER, A.C., STEWART, B.M. Perceiving relationships: A physiological examination of the perception of scatterplots. In D. Barker-Plummer et al (eds.): *Diagrams 2006*, 244-257.

[BK79] BOBKO, P., KARREN, R. The perception of Pearson product moment correlations from bivariate scatterplots. *Personnel Psychology*, 32 (1979), 313-325.

[CDM82] CLEVELAND, W.S., DIACONIS, P., MCGILL, R. Variables on scatterplots look more highly correlated when scales are increased. *Science*, 216 (1982), 1138-1141.

[Cle93] CLEVELAND, W.S. (1993). *Visualizing Data*. Summit NJ: Hobart Press.

[CME99] COREN, S, WARD, L.M., ENNS, J.T. *Sensation and Perception* (5th ed.ition), chapter 2. New York: Harcourt Brace, 1999.

[CMS99] CARD, S.K., MACKINLAY, J.D., SHNEIDERMAN, B. Information visualization. In Card, S.K., Mackinlay, J.D., & Shneiderman, B. (Eds.) *Readings in Information Visualization: Using Vision to Think*, chapter 1. San Francisco: Morgan Kaufman. 1999.

[DAAK07] DOHERTY, M.E., ANDERSON, R.B., ANGOTT, A.M., KLOPFER, D.S. The perception of scatterplors. *Perception & Psychophysics*, 69 (2007), 1261-1272.

[EF00] ELLERMEIER, W., FAULHAMMER, G. Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception & Psychophysics*, 62 (2000), 1505 – 1511.

[FD05] FRIENDLY, M., DENIS, D. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, 41 (2005), 103 – 130.

[Har99] HARRIS, R.L. *Information Graphics: A Comprehensive Illustrated Reference.* Atlanta GA: Management Graphics, 1999.

[Kir95] KIRK, R.E. *Experimental Design: Procedures for the Behavioral Sciences* (3rd edition). Boston: Brooks-Cole, 1995, 37-40.

[KM08] KNOBLAUCH, K., MAHONEY, L. (2008). MLDS: Maximum likelihood difference scaling in R. *Journal of Statistical Software*, 25, 2 (2008), 1-26.

[LAK85] LANE, D.M., ANDERSON, C.A., KELLAM, K.L. Judging the relatedness of variables: The psychophysics of covariation detection. *Journal of Experimental Psychology: Human Perception and Performanc*e, 11 (1985), 640-649.

[LMvW08] LI, J., MARTENS, J-B., VAN WIJK, J.J. Judging correlation from scatterplots and parallel coordinates. *Information Visualization* (2008), doi:10.1057/palgrave.ivs.9500179

[LP89] LAUER, T.W., POST, G.V. Density in scatterplots and the estimation of correlation. *Behaviour & Information Technology*, 8 (1989), *235-244*.

[MTF97] MEYER, J., TAIEB, M., FLASCHER, I. Correlation estimates as perceptual judgments. *Journal of Experimental Psychology, Applied*, 3 (2005), 3-20.

[Pol60] POLLACK, I. Identification of visual correlational scatterplots. *J. Experimental Psychology*, 59 (1960), 351-360.

[SH78] STRAHAN, R.F., HANSEN, C.J. Underestimating correlation from scatterplots. *Applied Psychological Measurement*, 2 (1978), 543-550.

[War04] WARE, C. *Information Visualization: Perception for Design*. (2nd edition) New York: Morgan Kaufmann. 2004.