

Multi-Scale Banking to 45°

Jeffrey Heer and Maneesh Agrawala

Abstract—In his text *Visualizing Data*, William Cleveland demonstrates how the aspect ratio of a line chart can affect an analyst's perception of trends in the data. Cleveland proposes an optimization technique for computing the aspect ratio such that the average absolute orientation of line segments in the chart is equal to 45 degrees. This technique, called *banking to 45°*, is designed to maximize the discriminability of the orientations of the line segments in the chart. In this paper, we revisit this classic result and describe two new extensions. First, we propose alternate optimization criteria designed to further improve the visual perception of line segment orientations. Second, we develop *multi-scale banking*, a technique that combines spectral analysis with banking to 45°. Our technique automatically identifies trends at various frequency scales and then generates a banked chart for each of these scales. We demonstrate the utility of our techniques in a range of visualization tools and analysis examples.

Index Terms—Information visualization, banking to 45 degrees, line charts, time-series, sparklines, graphical perception

1 INTRODUCTION

Few visualizations are more common than the line chart, used to communicate changes in value over a contiguous domain, with the slopes of line segments encoding rates of change. Mathematical functions, stock prices, and all varieties of time-series data are plotted in this form. Accordingly, techniques which improve the graphical perception of such displays could be of great practical benefit. Given the human visual system's sensitivity to line orientation [6], it is not surprising that the relative orientation of line segments in a chart can strongly impact an analyst's perception of trends in the data. One way of intentionally manipulating these orientations is through the choice of the *aspect ratio* (width/height) of the chart.

Both in his book *Visualizing Data* [2] and elsewhere [1,3,4], William Cleveland demonstrates how the choice of a line chart's aspect ratio can impact graphical perception. Figure 1 shows plots of average monthly carbon dioxide measurements made at the Mauna Loa observatory, an example used in [2]. The first plot (1a) clearly shows an upward trend in the data. There is an inflection in the curve, indicating that the increase in carbon dioxide is accelerating. The second chart (1b) shows the same data plotted at a wider aspect ratio. The slow onset and quick decay of the yearly oscillations is clearly visible in this plot. However the inflection that was visible in the first chart becomes difficult to see.

Noting that an average orientation of 45° maximizes the discriminability of adjacent line segments, Cleveland introduces a technique called *banking to 45°*, which determines the aspect ratio such that the average orientation of all line segments in a chart is 45 degrees [2,3,4]. While Cleveland's optimization procedure makes it easier to see higher frequency oscillations in the data, it can obscure lower-frequency trends of interest. The aspect ratio used in Figure 1b is the result of banking to 45°. As we have noted, the low-frequency inflection point is difficult to discern in this chart. Cleveland addresses this issue, describing a manual process of fitting smooth regression curves to the data and banking the resulting low frequency curve. Finding interesting trends in the data becomes a tediously iterative trial-and-error process. Users must manually consider each

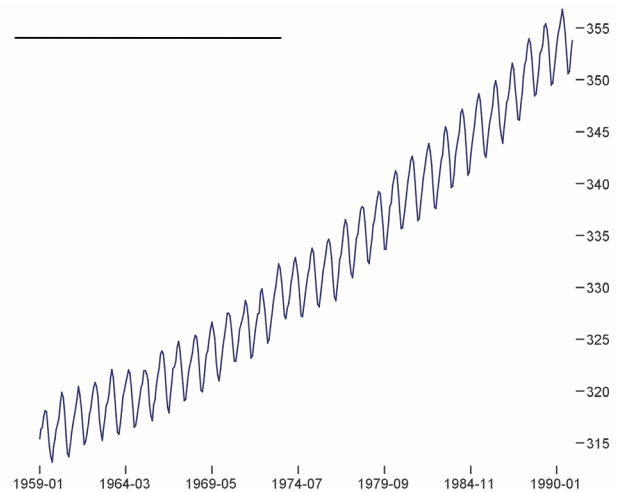


Figure 1a. CO₂ measurements, aspect ratio = 1.17. The x-axis shows time, in monthly increments, while the y-axis shows carbon dioxide measurements taken at the Mauna Loa observatory [2]. Note the bend in the trend of increasing values, suggesting an accelerating increase.

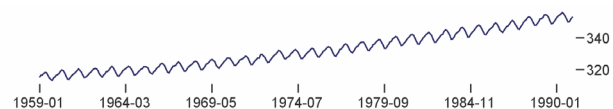


Figure 1b. CO₂ measurements, aspect ratio = 7.87. The wider aspect ratio enables the viewer to see that the ascent of each yearly cycle is more gradual than its decay. However, the bend in the lower-frequency trend is now difficult to see. The choice of aspect ratios for Figures 1a and 1b were automatically determined using multi-scale banking.

smoothness level (*i.e.*, the frequency scale), bank it to 45°, and then visually check for an interesting trend.

In this paper we extend Cleveland's work in two ways. First, we explore alternate optimization criteria for Cleveland's banking procedure. These criteria are designed to find an aspect ratio that further improves the visual perception of line segment orientations. Second, we develop *multi-scale banking*, a technique that combines spectral analysis with banking to 45°. Our technique automatically identifies frequency scales that may be of interest and then generates a banked chart for each of these scales. The aspect ratios for the charts in Figures 1a and 1b were automatically chosen using our multi-scale banking approach. While our technique finds the same aspect ratio as Cleveland's original approach in Figure 1b, it also returns the aspect ratio which reveals the low-frequency inflection point in Figure 1a. We have incorporated multi-scale banking into tools for both static and dynamic visualizations, and present the results of applying the technique on a collection of real-world data sets.

- Jeffrey Heer is with the Computer Science Division of the University of California, Berkeley. E-Mail: jheer@cs.berkeley.edu.
- Maneesh Agrawala is with the Computer Science Division of the University of California, Berkeley. E-Mail: maneesh@cs.berkeley.edu.

Manuscript received 31 March 2006; accepted 1 August 2006; posted online 6 November 2006. For information on obtaining reprints of this article, please send e-mail to: tcvg@computer.org.

2 BANKING TO 45°

Line charts are typically used to present the relationship between two variables. For a given value of variable x , viewers can directly read off the corresponding value of y and vice versa. Viewers can also determine the rate of change of y with x , by judging the orientations of line segments in the chart. As we have seen, the aspect ratio of the chart can significantly affect the visual perception of the orientations of the line segments. The challenge in designing line charts is to choose an aspect ratio that makes it easy for viewers to discriminate line orientations in the chart and thereby determine the rates of change. Cleveland [1,2,3,4] was the first to systematically address this challenge. For completeness we review his techniques and then present several new additions.

2.1 Maximizing Orientation Resolution

Given a set of n line segments, the *orientation resolution* is the range of orientations spanned by the segments and is computed as the maximum orientation minus the minimum orientation. Cleveland et al. [1] conducted human-subject experiments showing that viewers judge the ratio of the slopes of two adjacent line segments most accurately when the orientation resolution between them is maximized. They also show that for two line segments, the resolution is maximized when the absolute value of the *orientation midangle*, the average of the maximum orientation and the minimum orientation, is 45°.

Cleveland et al.'s argument is as follows. Suppose that s_1 and s_2 are the slopes of two line segments (in the original units of the data), and that $s_1 > s_2$. In the physical display space, for a display rectangle of height h , width w and aspect ratio $\alpha = w/h$, the physical slopes become s_1/α and s_2/α . The orientation resolution, computed with respect to the physical slopes of the line segments, is

$$r(\alpha) = \tan^{-1}(s_1/\alpha) - \tan^{-1}(s_2/\alpha),$$

and the orientation midangle is,

$$m(\alpha) = \frac{\tan^{-1}(s_1/\alpha) + \tan^{-1}(s_2/\alpha)}{2}$$

To simplify the formulation let $f = s_2/s_1$ and $\lambda = \alpha^{-1}\sqrt{s_1s_2}$. Then,

$$r(\lambda) = \tan^{-1}(\lambda/\sqrt{f}) - \tan^{-1}(\lambda\sqrt{f})$$

$$m(\lambda) = \frac{\tan^{-1}(\lambda/\sqrt{f}) + \tan^{-1}(\lambda\sqrt{f})}{2}$$

At $\lambda = 1$, $r(\lambda)$ is maximized to a value of $2(45^\circ - \tan^{-1}(\sqrt{f}))$ and $m(\lambda) = 45^\circ$. Thus for two line segments the orientation resolution is maximized when the average angle between them is 45°. Similarly for two line segments of negative slopes the orientation resolution is maximized at a midangle of -45°.

2.2 Median-Absolute-Slope Banking

With more than two line segments it is impossible to choose a single aspect ratio such the midangle between all positively sloped pairs is 45° and all negatively sloped pairs in -45°. Instead, Cleveland et al. proposes choosing the aspect ratio that sets the median absolute slope of the line segments to 1. For a display rectangle of height h , width w and aspect ratio $\alpha = w/h$, to display n line segments with slopes s_i ($i=1, \dots, n$) Cleveland et al. find α such that the median $|s_i/\alpha| = 1$. Thus approximately half the segments will have an absolute slope of greater than 1 and half will have an absolute slope of less than 1.

The procedure for computing α is straightforward. Let $R_y = y_{max} - y_{min}$, the range of y values where y_i range over the set of endpoints, similarly let $R_x = x_{max} - x_{min}$, and let $M = \text{median } |s_i|$. Then setting $\alpha = MR_x/R_y$ satisfies the median-absolute-slope criterion. This closed-form solution is fast and easy to compute.

2.3 Average-Absolute-Orientation Banking

More recently, Cleveland [2,3,4] proposes banking the average orientation to 45° rather than setting the median slope to 1. For aspect ratio α the orientation of a line segment drawn in a chart is given by $\theta_i(\alpha) = \tan^{-1}(s_i/\alpha)$. We must find an aspect ratio such that

$$\sum_i \frac{|\theta_i(\alpha)|}{n} = 45^\circ. \quad (1)$$

Cleveland points out that although there is no closed-form solution to this equation, it is monotone in α . Therefore iterative optimization techniques (*i.e.*, Newton's method) will converge to a solution.

2.4 Weighted Average-Absolute-Orientation Banking

Cleveland further suggests weighting the average absolute orientation by the lengths of the line segments. Denoting the length of the i th line segment by $l_i(\alpha)$, the weighted version of equation 1 becomes

$$\frac{\sum_i |\theta_i(\alpha)| l_i(\alpha)}{\sum_i l_i(\alpha)} = 45^\circ \quad (2)$$

Again we must solve this equation using iterative optimization, but because the equation remains monotone in α it can be solved efficiently. Cleveland asserts that this weighting results in more satisfactory results for the datasets he has tested [4].

It is worth considering the effect of this weighting. In typical line charts, the x -values are uniformly distributed and therefore the length of a line segment is determined by the difference in the y -values of the endpoints. In this case the lines of longest length are also those of greatest absolute slope, and thus of greatest absolute orientation. Weighting by segment length preferentially weights line segments with higher absolute orientations (*i.e.*, closer to vertical), more aggressively optimizing these segments to 45°.

2.5 Banking by Optimizing Orientation Resolution

Cleveland's techniques are based on the hypothesis that maximizing orientation resolution maximizes the viewer's ability to perceive the orientation of line segments. Yet, both median-absolute-slope banking and average-absolute-orientation banking are based on centering either the slopes or orientations of line segments about 1 or 45° respectively. As described in Section 2.1 such centering is optimal for the case of two line segments. For multiple line segments, however, it is unclear whether centering generates the optimal solution. Moreover, the effect of these centering techniques on the overall orientation resolution is indirect.

Therefore, we propose directly optimizing orientation resolution. Given two line segments i and j , the angle between them is $r_{i,j}(\alpha) = |\theta_i(\alpha) - \theta_j(\alpha)|$. Note that $r_{i,j}$ is computed as the smallest acute angle between the line segments, and it therefore always lies between 0° and 90°. For each pair of line segments, the closer the angle difference is to 90°, the easier it is to visually perceive the difference in orientations. Thus, our approach is to find the aspect ratio α that maximizes $r_{i,j}$ between all pairs of line segments.

$$\sum_i \sum_j r_{i,j}^2 \quad (3)$$

In contrast to Cleveland's techniques (Sections 2.2, 2.3 and 2.4), this approach more directly optimizes orientation resolution since it considers the angles between line segments.

In some line charts discriminating the orientations of successive lines is more important than discriminating orientations of all pairs. In this case we can optionally choose to apply equation 3 only to successive pairs of line segments rather than exhaustively applying it to all n^2 pairs of segments. The result is a local, rather than global, optimization of the orientation resolution.

2.6 Average-Absolute-Slope Banking

Cleveland et al. [3] have argued that viewers judge orientation rather than slope when estimating the rate of change of a curve. They point out that our ability to discriminate two orientations θ , $\theta + \delta$ depends only on δ , but discriminating between two slopes s and $s(1+r)$ depends on both s and r . For fixed r and very large or very small s , the slopes become difficult to discriminate. This is why Cleveland formulates his average-absolute-orientation banking techniques specifically with respect to line segment orientations.

As shown previously, slope and orientation are related by the arctangent function. In the case of a single line segment, banking the orientation to 45° and banking the slope to 1 are equivalent operations, as $\tan^{-1}1 = 45^\circ$. The median-absolute-slope procedure considers only a single line segment (the median segment), and because the arctangent is monotone the median slope and median orientation coincide. In the case of multiple line segments, however, the non-linearity of the arctangent function causes the two approaches to diverge. This non-linearity necessitates the use of optimization techniques when computing average-absolute-orientation banking.

Optimizing the average absolute slope to 1 can be computed quickly and easily, but does not produce the same result as optimizing the average absolute orientation. For completeness and comparison, we introduce average-absolute-slope banking here. The procedure for banking the average absolute slope to 1 follows the formulation of median slope banking. As before, $\alpha = MR_x / R_y$, except we now set $M = \text{mean} |s_i|$. As for median-absolute-slope banking the solution for α is closed-form.

2.7 Slopeless Line Culling

All of the banking procedures we have presented consider either the slope or orientation of every line segment in the dataset. An additional modification is to cull “slopeless” lines – those with either zero or infinite slope. Horizontal and vertical lines remain unchanged by variations in aspect ratio, yet contribute to the banking criteria. For example, inclusion of zero-sloped lines can affect the calculation of either the median or average orientation. By removing them from consideration, the remaining line segments can be more closely banked to 45°. This modification can be incorporated into any of the previous banking routines. Figure 2 shows the results of weighted-average-orientation banking with and without slopeless line culling on an illustrative data set. Only when culling is applied do the sloped segments reach an average orientation of 45°.

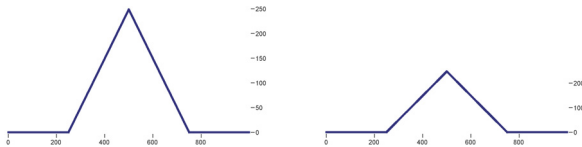


Figure 2. Effects of slopeless line culling. Horizontal lines are typically included when banking by average orientation, resulting in an aspect ratio of 1.97 on the left. Culling these lines in the banking optimization results in an aspect ratio of 4.00 on the right. In this case, the sloped segments are set to 45° after the banking procedure.

2.8 Comparing Banking Approaches

We have applied each banking technique described in the previous sections on four real-world data sets. We compare the results of each of the banking techniques, both with and without the slopeless line culling modification, resulting in a total of 12 banking approaches. The four data sets are: the CO₂ measurements discussed earlier (co2), two months of daily download counts of the prefuse visualization toolkit [5] (prefuse), eight years of share prices for the PRMTX mutual fund (prmtx), and a set of yearly measurements of sunspot activity (sun). These data sets are discussed in greater detail in

Banking Technique	co2	prefuse	prmtx	sun
median-slope (ms)	9.19	8.81	13.70	21.88
median-slope culled (msc)	9.19	8.81	14.55	21.88
average-slope (as)	9.19	12.48	19.66	26.78
average-slope culled (asc)	9.22	12.70	19.99	26.88
average-orient (ao)	7.94	8.99	12.53	18.74
average-orient culled (aoc)	7.98	9.30	12.99	18.88
average-weighted-orient (awo)	9.14	12.42	19.55	26.60
average-weighted-orient culled (awoc)	9.17	12.65	19.90	26.69
global-orient-resolution (gor)	8.02	9.01	12.87	19.26
global-orient-resolution-culled (gorc)	8.05	9.29	13.24	19.37
local-orient-resolution (lor)	6.41	10.33	12.05	16.28
local-orient-resolution-culled (lorc)	6.45	10.59	12.44	16.44

Table 1. Aspect ratio variation across banking techniques. Each column contains the results of banking a single data set using twelve different bankers. Each cell contains the computed aspect ratio. The banking techniques introduced by Cleveland [1,2,3,4] are in bold italics.

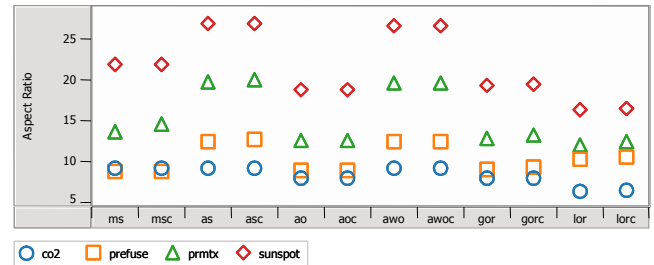


Figure 3. Aspect ratio variation across banking techniques. Culled and non-culled banking techniques show little separation. The (ao) and (gor) techniques produce similar results, as do (as) and (awo). The (ms) technique produces aspect ratios between those of the other two groups.

Section 3.2 and plots of each data set at various aspect ratios are presented in Figures 5-8.

The aspect ratios computed by the banking techniques for each data set are shown in Table 1 and Figure 3. We can see that culling slopeless lines has had little effect. This is to be expected, as the data sets used have few slopeless segments. It is only when the number of slopeless segments accumulates that the culling modification will have a pronounced effect.

A number of similarities and differences between the techniques are also evident. The non-weighted average-absolute-orientation (ao) technique produces results similar to global optimization of the orientation resolution (gor). Yet, (ao) considers a single orientation per line segment, while (gor) considers all pairs of line segments. Therefore (ao) is computationally more efficient and thus may be preferable to (gor).

Similarly we see that the local resolution optimizer (lor) produces consistently lower aspect ratios than the global version. In contrast, the average-weighted-orientation (awo) and average-absolute-slope (as) bankers consistently produce higher aspect ratios. As discussed earlier, this is expected of the weighted-average optimizer, as it preferentially weights segments with higher orientations and thus “flattens” those segments more aggressively. The observed match between (awo) and (as) is particularly interesting. Given the discussion of Section 2.6., it may be surprising that a slope optimizer and an orientation optimizer provide such similar output. The average-slope banker has a closed-form solution and can be computed more efficiently than a weighted-average-orientation banking. Although Cleveland [2,3,4] recommends using the weighted optimizer, these results suggest that similar results may be achieved using a more computationally efficient technique. Finally, the absolute-median-slope banker (ms) consistently calculates aspect

ratios in-between the extremes of the other groups. It can also be computed directly, without need of an optimizer.

What remains unclear is which banking approach is the most perceptually effective. A perceptual argument has been made for optimizing the orientation resolution (suggesting the (ao) method). However, Cleveland recommends the average weighted orientation (well matched in our results by the more efficient (as) method). The (ms) method appears to split the difference between the two. Exploration of additional data sets and further perceptual experiments might be used to resolve the uncertainty.

3 MULTI-SCALE BANKING TO 45°

Each of the banking approaches discussed in the previous section consider the entirety of the data when determining the chart aspect ratio. As a result, local features are accentuated in the banking, sometime obscuring larger-scale trends of interest. For example, the carbon dioxide measurements discussed in the introduction includes an accelerating increase in values. This trend may be difficult to discover if one only saw the banked plot of the data (Figure 1b). As we will see, intermediate scales between local and global trends may also contain patterns of great interest to an analyst. These observations suggest a need for tools which aid analysts by identifying trends and presenting appropriate visualizations. Thus, we introduce *multi-scale banking*, an approach that finds a set of aspect ratios that optimize the display of trends at varying scales.

A rudimentary approach to multi-scale exploration is to manually adjust the aspect ratio and observe the resulting views. Many charting applications allow users to change the aspect ratio by directly resizing the data axes and the display window. Changing the axis range of a fixed display (for example, by using range sliders) also permits manual banking, since the aspect ratio changes as the axis scale is varied. Though these approaches enable manual banking of the data display, existing tools provide very little assistance for choosing the appropriate scale. Analysts must manually test a variety of scales and visually scan the results to find the ones that best show trends in the data. Such an iterative search is excessively tedious, particularly when analyzing large data sets with a high degree of detail.

Cleveland describes a manual, iterative process of trend discovery for multi-scale exploration [2]. He begins by fitting the data using *loess*, a locally weighted regression technique. Loess considers sequential windows of a data set; these local regions are weighted and fit to a polynomial regression curve. The technique requires two parameters: a smoothing parameter α , corresponding to the size of the window, and a parameter λ , specifying the degree of the fitting polynomial (typically 1 or 2). The analyst must manually choose values for these parameters, often by trial and error. Once a desired trend curve has been found in this manner, it can be used as input to a banking routine, to compute an aspect ratio that optimizes the display of the trend. To then explore additional scales, the trend curve is subtracted from the data and the analysis begins anew on the residual data. Cleveland suggests ending the analysis once the residuals are well fit by normally-distributed random values. The goal of our multi-scale banking is to accelerate this analysis process by automatically identifying scales of interest in the data.

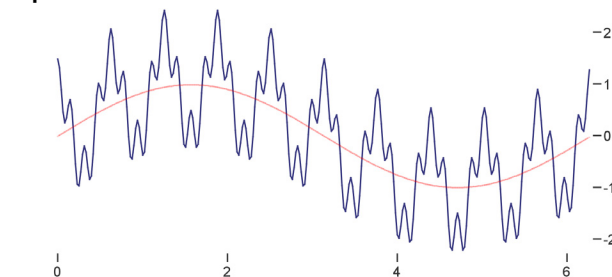
3.1 Method

To identify the scales of interest we analyze a frequency domain representation of the data to look for strong periodic trends. At each scale of interest we low-pass filter the data to eliminate frequencies at higher scales and then bank the data to compute scale-specific aspect ratios.

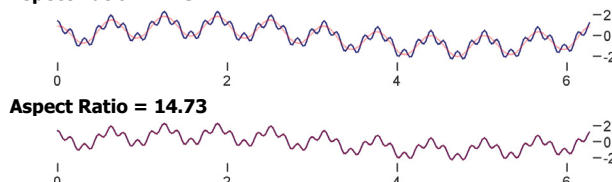
To help illustrate our approach, consider a simple data set created by combining the output of three sinusoidal oscillators, each operating at a different frequency and thus creating a signal with multiple scales of interest. Figure 4 displays the result of applying multi-scale banking to this data set.

$$\sin(x) + \cos(10x) + 0.5 \cos(40x)$$

Aspect Ratio = 2.21



Aspect Ratio = 11.34



Aspect Ratio = 14.73



Power Spectrum



Aspect Ratios



Figure 4. Multi-scale banking of a multi-frequency signal. The data plots at the top of the figure show the data using identified aspect ratios, with banked trend lines shown in red. The power spectrum plot shows the frequency-domain representation of the data, with identified scales of interest indicated by blue bands. The aspect ratio plot shows the banked aspect ratios for each possible lowpass filtering of the data, with the final selection of aspect ratios indicated by blue bands.

We start the analysis by applying a discrete Fourier transform (DFT) to the input data and computing the squared magnitude of each of the resulting Fourier coefficients to form the power spectrum. Points of high-value in the power spectrum correspond to strong frequency components within the input signal. Figure 4 includes the power spectrum for our example data set. Note the prominent spikes for each of the frequency components.

Next, we search for trends of interest in the power spectrum. One approach is to threshold the spectrum and keep all frequency scales for which the power is greater than the threshold. However, we have observed that spectral energy often occurs in “clumps,” sometimes containing local oscillation. By smoothing the power spectrum over small windows we can even out these local regions prior to thresholding. The smoothing is performed by convolving the power spectrum with a Gaussian filter, using a small window width (typically 3) and unit standard deviation. We have also experimented with pre-multiplying the signal with a window function (e.g., Hamming and Hann windows) but have so far noticed little change in the results.

We then threshold the smoothed spectrum and consider all values that exceed the threshold. By default, we use the mean of the smoothed power spectrum as the threshold value. This default has worked well in practice for small to medium size data sets, though it can be increased (or alternatively, the smoothing window can be modified) to reduce the number of scales chosen by our algorithm.

After applying the threshold we are left with a set of candidate scales. These candidate scales correspond to strong frequency components and often form a range of contiguous points in the power spectrum. Proceeding from the lowest to the highest frequency, we retain only the last (highest-frequency) value of each range of consecutive values, thereby capturing the total contribution of that local region of energy. We also ensure that the highest-

frequency component of the entire power spectrum is included, representing the “trend” of the data in its entirety. These values constitute the scales of interest, and are indicated by the blue bands in the power spectrum plot of our example data.

The second phase of a multi-scale banking approach is to create trend curves for each identified scale and bank the resulting curve to arrive at a scale-specific aspect ratio. To form the trend curves we low-pass filter the data, removing all frequencies higher than the current scale of interest. The resulting trend curve is then banked to 45°. Figure 4 shows the resulting banked views for the sample oscillator data, using the median-absolute-slope banking technique (Section 2.3). The trend curves are shown in red.

Although we compute aspect ratios for each scale of interest, in some cases the aspect ratios for two or more distinct scales may be very similar to one another. Plotting charts at each of these similar aspect ratios would result in displays that are nearly identical to one another. Our solution is to cull aspect ratios that are judged to be highly similar. We apply a simple filter that considers each aspect ratio in order of increasing frequency. An aspect ratio is kept only if it differs from the previous aspect ratio by a threshold scale factor. In practice, a scale factor of 1.25 has worked well for most cases.

The final result of the multi-scale banking process is shown in Figure 4. Note that the low frequency scale in the first plot becomes more obscured in the subsequent plots. In addition to the data plots and power spectrum, the figure also shows a graph of aspect ratios for each possible low-pass filtering of the data. The final aspect ratios recommended by our algorithm are indicated as light blue bands on the aspect ratio plot. The aspect ratio for the last identified scale seen in the power spectrum plot was filtered, as it is close in value to the aspect ratio for the previous scale of interest. The aspect ratio chart also suggests the alternative approach of analyzing the aspect ratio curve directly. This possibility is discussed further in Section 4. A pseudocode description of the technique is presented as Algorithm 1.

3.2 Results

We have applied our multi-scale banking technique to a number of real-world data sets. To compute the banking, we used Cleveland’s median-absolute-slope technique with slopeless line culling applied. Results are reported using the same small multiples format of Figure 4. Plots using each computed aspect ratio are shown, including the banked trend curves in red. Each figure includes a plot of the power spectrum, with identified scales indicated by light blue bands. The x-axis of the power spectrum shows the frequency index, indicating the periodicity of the trend (e.g., a value of 8 indicates that the trend repeats 8 times within the data set). A plot of the aspect ratios for each possible low-pass filtering is also shown, with the filtered set of aspect ratios indicated by light blue bands.

3.2.1 Sunspot Cycles

The first data set is a collection of astronomical sunspot observations, a classic example used by Cleveland to illustrate the utility of the banking to 45° technique [4]. The data contains values of the Wolfer number, a measure of the size and number of sunspots observed in a given year, from 1700 to 1987. Results are shown in Figure 5. Spectral analysis finds four scales of interest, at frequency indices 7, 10, 31, and 36, plus the scale of the data in its entirety. Filtering the aspect ratios yields two charts, corresponding to frequency scales 7 and 31.

The first aspect ratio of 3.96 shows a low frequency trend indicating the oscillation of high-points across sunspot cycles. The second aspect ratio of 22.35 shows the oscillation of the individual sunspot cycles, with a period of about 11 years. In this view, it becomes apparent that many of the cycles have a steep onset followed by a more gradual decay.

MULTIBANK(y)

% Returns a set of computed aspect ratios
% y : an array of data values to bank

$w \leftarrow 3$ % Gaussian filter window size
 $\sigma \leftarrow 1$ % Gaussian filter standard deviation
 $\alpha \leftarrow 1.25$ % Scale threshold for aspect ratio

$Y \leftarrow \text{FFT}(y)$ % apply discrete Fourier transform
 $P_i \leftarrow \|Y_i\|^2$ % compute power spectrum
 $Z \leftarrow P \times G(w, \sigma)$ % convolution with Gaussian filter
 $T \leftarrow \text{MEAN}(Z)$ % mean filtered power, use as threshold

% Compute aspect ratios for trends that pass the threshold.
% For consecutive runs, only the last value is retained.

for each $i \in [1 \dots \text{LENGTH}(Z)]$
 if $i = \text{LENGTH}(Z)$ or $(Z_i > T \text{ and } Z_{i+1} < T)$
 $ar \leftarrow ar \cup \text{BANKTO45}(\text{LOWPASS}(y, i))$

$AR \leftarrow \{ ar_i \}$ % aspect ratios, include lowest by default

% Keep aspect ratios that are sufficiently different from one another

for each $i \in [2 \dots \text{LENGTH}(ar)-1]$
 $j \leftarrow \text{LENGTH}(AR)$
 if $\text{MAX}(ar_i, AR_j) / \text{MIN}(ar_i, AR_j) > \alpha$
 $AR \leftarrow AR \cup ar_i$

return AR

Algorithm 1: Multi-scale banking. Computes aspect ratios for charts at multiple scales by using spectral analysis to identify trends of interest. This pseudocode description assumes the availability of an FFT (fast Fourier transform) routine and a routine LOWPASS for low-pass filtering a signal subject to a given periodicity (frequency index). The routine BANKTO45 may use any of the banking routines discussed in Section 2.

3.2.2 Carbon Dioxide Measurements

The next example, also taken from Cleveland, is the previously introduced set of average monthly atmospheric carbon dioxide measurements from the Mauna Loa observatory [2]. Results are shown in Figure 6. Spectral analysis finds two scales of interest, at frequency indices 11 and 33, plus the scale of the data in its entirety. Filtering the aspect ratios yields selected views at scales 11 and 33.

The first aspect ratio of 1.17 shows a general trend of increasing CO₂ concentrations. A slight bend can be observed, indicating the trend is accelerating. The second aspect ratio of 7.87 provides a clearer view of yearly oscillations. Notably, these cycles have a gradual rise and a steeper decline.

3.2.3 Mutual Fund Performance

Figure 7 shows of the results of our technique applied to financial data, in this case daily closing share values of the T. Rowe Price Media and Technology Fund (PRMTX) from 1997 to the present. Spectral analysis finds one scale of interest at frequency index 38, plus the scale of the data in its entirety. There is an interesting lack of dominant frequency components past the first identified scale. Filtering the aspect ratios yields two charts.

The first aspect ratio of 4.23 shows a low frequency oscillation in values, depicting the “dot-com” boom and bust followed by a subsequent recovery. The second aspect ratio of 14.55 allows local variations to be more clearly seen. The second aspect ratio is included as a result of our policy of always including a scale corresponding to the entirety of the data. The corresponding plot is the result of banking the non-filtered data set to 45°.

Sunspot Cycles

Aspect Ratio = 3.96

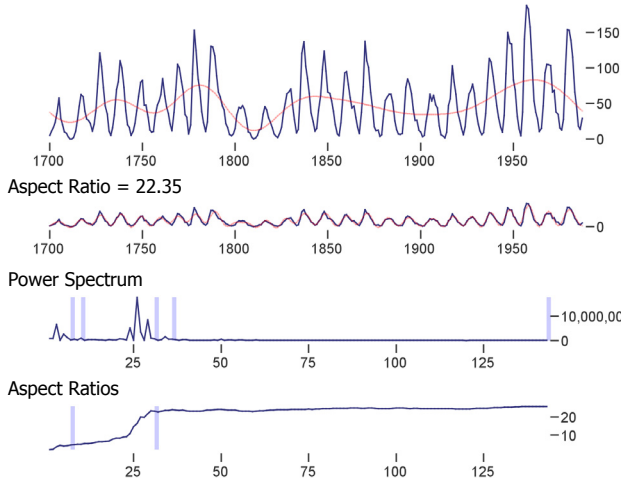


Figure 5. Sunspot observations, 1700-1987. The first plot shows low-frequency oscillations in the maximum values of sunspot cycles. The second plot brings the individual cycles into greater relief.

Carbon Dioxide Measurements

Aspect Ratio = 1.17

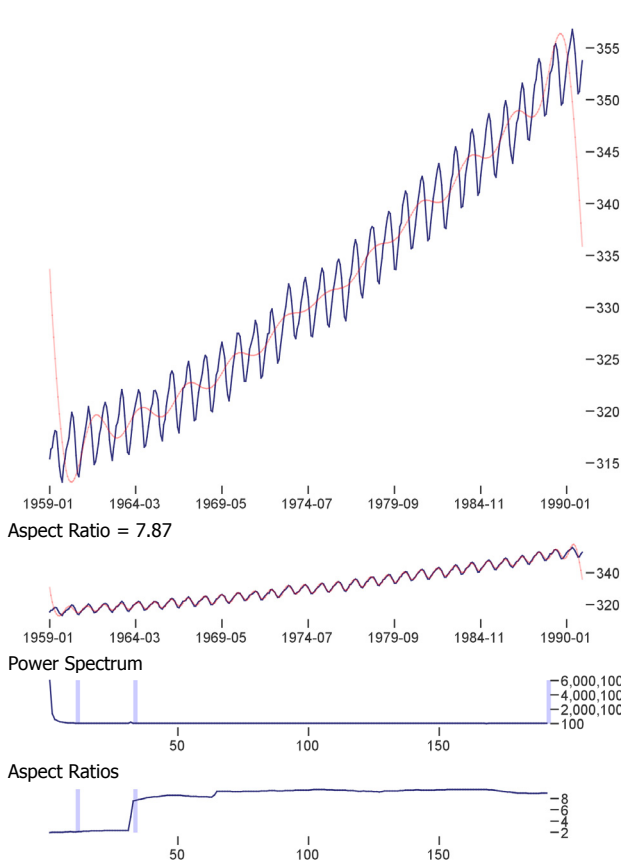


Figure 6. Monthly atmospheric CO₂ measurements. The first plot shows a baseline trend of increasing values, with a slight inflection. The second plot more clearly communicates the yearly oscillations.

Figures 5-8 show the results of applying multi-scale banking to real-world data sets. Data sets are plotted at each computed aspect ratio, with banked trend lines shown in red. The power spectrum plot shows a frequency-domain representation of the data, annotated with potential scales of interest. The aspect ratio plot shows the banked aspect ratios for each possible lowpass filtering of the data, annotated with the final aspect ratios returned by the algorithm.

PRMTX Mutual Fund

Aspect Ratio = 4.23

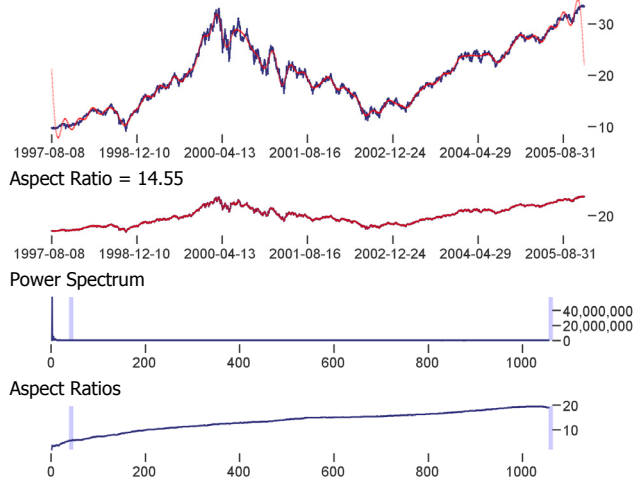


Figure 7. PRMTX mutual fund performance, 1997-2006. The first plot shows the boom and bust of the "dot-com" bubble and subsequent recovery. The second plot affords closer consideration of short-term variations.

Downloads of the prefuse toolkit

Aspect Ratio = 1.44

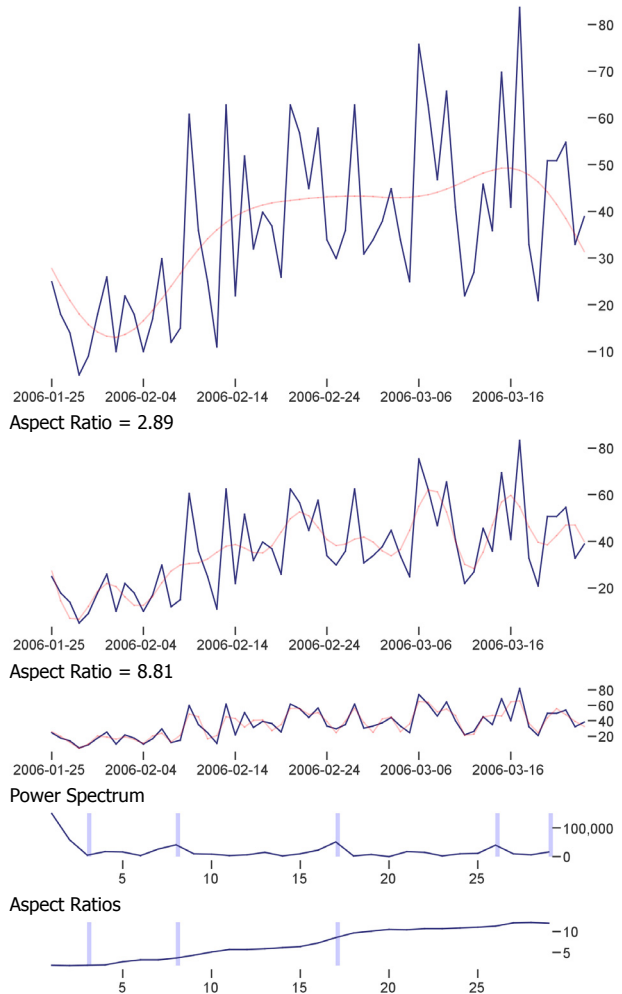


Figure 8. Daily download counts of the prefuse visualization toolkit. The first plot shows a general increase in downloads. The second plot shows weekly variations, including reduced downloads on the weekends. The third plot enables closer inspection of day-to-day spikes and decays.

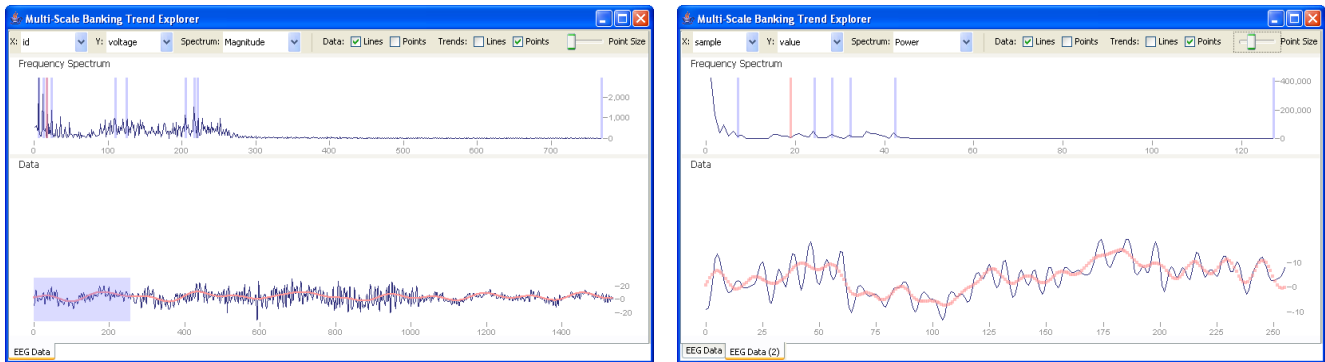


Figure 9. Trend Explorer application, visualizing a segment of electroencephalogram (EEG) readings. An interactive spectrum view shows trends identified by the multi-scale banking algorithm and enables users to select different filtering points, triggering recalculation of the aspect ratio. By selecting a data range in the line chart (shown in the image on the left), users can “zoom” into subsets of data (shown in the image on the right).

3.2.4 Open Source Software Project Tracking

The final data set consists of two months of daily project download counts for *prefuse*, an open source visualization toolkit developed by the authors [5]. A beta release of the toolkit was launched in February 2006, and accordingly we are interested in how downloading patterns may have been affected. Results are shown in Figure 8. Spectral analysis finds four scales of interest, at frequency indices 3, 8, 17, and 26, plus the scale of the data in its entirety. Filtering the aspect ratios yields three charts, corresponding to the first three scales.

The first aspect ratio of 1.44 shows an increase in the baseline rate of downloads, starting with the February 9th release of the beta. The second aspect ratio of 2.89 shows an oscillating pattern that repeats 8 times over the two month period, indicating a weekly pattern. Looking closer, one discovers a trend that one might already suspect: fewer downloads occur on the weekends. The third aspect ratio of 8.81 allows day-to-day oscillations to be better understood. Interestingly, downloads appear to increase immediately following a minor release of the toolkit (occurring on February 20, March 5, and March 20). The download count then witnesses a small initial decline, followed by another increase. Then a steeper decline sets in. This analysis indicates that toolkit users may update on different schedules. For example, some users update immediately while others wait a few days.


3.3 Applications

Multi-scale banking can be fruitfully applied in a number of visualization tools. In this section we discuss several applications we have built which utilize the multi-scale banking technique. Each of the applications was constructed using the *prefuse* visualization toolkit [5].

3.3.1 Small-Multiples Reports

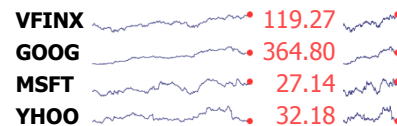
One application of multi-scale banking is for generating descriptive reports of a data set, using small multiples [7] to display trend data at various aspect ratios. These summary reports are useful for gaining a static overview of a data set and also appropriate for use in presentations or printed papers. We have written a simple tool for generating these reports, and have used it to generate the figures shown in the previous sections.

3.3.2 Sparklines

Banking can also be applied to aid the generation of *sparklines*, data-intense, word-sized graphics designed to be incorporated into text [8]. For example, a plot of sunspot data might be included inline , supporting uninterrupted reading without recourse to a separate figure. The vertical extent of a sparkline is fixed by the ascent of the typeface, leaving the horizontal extent of the graphic as the single degree of freedom. As noted by Tufte [8], banking can be applied to optimize the sparkline aspect ratio, thereby

determining the width. We have built a simple tool that takes a data set and typeface as input and generates appropriately banked sparkline images.

The following set of sparklines generated by our tool shows the daily closing share prices of a popular index fund and three well-known technology companies from March 1, 2005 to March 1, 2006. Multi-scale banking was applied separately to each data set. The aspect ratios for the lowest and highest frequency scales for each data set were collected. We separately averaged the aspect ratios for the high frequency scales and the low frequency scales to generate two final aspect ratios: ~ 8.3 for the high frequency scale and ~ 2.9 for the low frequency scale. The high frequency plots on the left afford detailed inspection, while the low frequency plots on the right provide a summary of the major trends. More nuanced approaches to banking multiple lines are discussed in Section 4.



3.3.3 Trend Explorer

We have also applied multi-scale banking within an interactive analysis tool, incorporating the results of automated analysis into manual exploration tasks. Figure 9 shows our “trend explorer” application for interactively exploring line charts at various scales. The lower panel displays the data plot and trend curves while the upper panel provides a frequency-domain representation of the data (users can select between viewing either the power spectrum or the magnitude of the Fourier coefficients). The results of the multi-scale banking routine are included as light blue bands within the spectrum display, highlighting the automatically identified scales of interest. Users can bank the data display to a particular trend by moving a red selector widget within the spectrum display, either by dragging it or by clicking at a desired location. In response, the system filters and banks the data, computing a new aspect ratio and updating the data display. As the user moves the selector near an automatically identified trend of interest, a “snap-to” effect is used to easily bank the display to that specific frequency.

In the default configuration, changes in the aspect ratio result in a rescaling of the data display, maintaining an overview of the whole time series. Both the horizontal and vertical dimensions may change in an effort to maximize the resolution of the overview display subject to the aspect ratio and available display space. The aspect ratio is preserved across window resizing operations. The application also supports a fixed data display mode, where changes in the aspect ratio instead result in a change in the displayed data range. The current range is displayed (and can be manually adjusted) using a range slider. This mode allows users to “zoom-in” to the data to view high-frequency trends at a higher display resolution. Users can perform more nuanced multi-scale analyses by selecting a range of

data in the display and then performing a new multi-scale banking analysis on just that subset, enabling fractal-like exploration of the data. This is particularly useful for diving into regions with interesting trends that do not recur across the data as a whole.

4 CONCLUSION AND FUTURE WORK

In this paper we have investigated techniques for automatic selection of aspect ratios for line chart displays, revisiting Cleveland's technique of *banking to 45°*. An empirical investigation of various banking approaches revealed useful correlations. For example, the data suggest that Cleveland's preferred technique of banking the weighted average absolute orientation to 45° may be well approximated by the more computationally efficient approach of banking the average absolute slope to 1. We also explored techniques combining spectral analysis and banking to automate the selection of aspect ratios at multiple scales. The initial results are encouraging, identifying trends of interest across multiple scales of the data.

It is worth discussing a few issues that arise when applying the presented techniques in practice. In our provided examples we have presented the banked data on a y-axis range determined by the minimum and maximum values in the data. Of course, for quantitative values on a ratio scale it is often more appropriate to start the scale at zero. Fortunately, this does not require any changes to the banking routines, only to the display logic. When the range of values does not include zero, the aspect ratio ar_θ for a zero-based plot that preserves the banked aspect ratio ar_b for the data can be computed as $ar_\theta = ar_b (y_{max} - y_{min}) / y_{max}$.

Another issue to consider is how to properly bank multiple curves plotted on the same display. Example scenarios include parallel coordinates, multi-line plots, and aligned small multiples (as in the sparkline example, §3.3.2). As described by Cleveland, one can bank multiple curves plotted over the same scale by simply including the line segments of each of the curves within a single banking optimization. For non-aligned scales, one can appropriately normalize the data and then perform a batch optimization using the normalized scale. While this approach covers the banking optimization, the spectral analysis component of a multi-scale banking should be run separately for each curve, creating a collection of potential frequency "cut-points". Each individual curve can then be low-pass filtered subject to a cut-point, and the filtered curves can then be aggregated for the banking optimization. An interesting avenue for further research is the task of correlating the different trends identified in each curve. The goal would be to better select "cut-points" for a group of curves being banked together.

Results of multi-scale banking analyses presented in this paper have included plots of the aspect ratios for each possible low-pass filtering of the data set. This raises an additional possibility: forego spectral analysis and instead compute and then analyze the aspect ratio curve directly. For example, "plateaus" in the aspect ratio curve may provide good views of the data. While feasible, the approach suffers from a few limitations. Most notable among these is the computational cost, as the technique requires an exhaustive application of filtering and banking computations. If the data set size is large or a banking routine requiring iterative optimization is used this quickly becomes prohibitively expensive.

While we have used low-pass filtering to create trend curves for identified scales of interest, other approaches are possible. Following the lead of Cleveland, we could subtract lower-frequency trends from the data, and bank the residuals to examine higher-frequency trends in isolation. Equivalently, we could simply high-pass filter the data instead. Applications of band-pass filtering also suggest themselves, for example to examine a local "clump" of spectral energy, as discussed previously. The results of spectral analysis could also be applied to inform alternative trend line generation schemes. Recall that the *loess* regression technique used by Cleveland takes two parameters: a smoothing parameter α and a polynomial-degree parameter λ . As this first parameter describes the size of a local window of consideration, good values of α may

correspond well to the periodicity of frequency components suggested by spectral analysis. We defer detailed explorations of these possibilities to future work.

Finally, it is worthwhile to consider alternative means of identifying scales of interest. We have used spectral analysis, which identifies periodic patterns across the entire data set. Cleveland has used repeated, manual application of locally weighted regression to identify scales of interest. Alternative techniques for isolating trends in the data (such as wavelets or automated regression techniques) are possible, and may prove appropriate dependent on the context of analysis and the type of data under consideration.

ACKNOWLEDGEMENTS

We would like to thank Dan Goldman for asking the questions that led to our initial inquiries about alternate optimization criteria.

REFERENCES

- [1] Cleveland, W. S., M. E. McGill, and R. McGill. The Shape Parameter of a Two-Variable Graph. *Journal of the American Statistical Association*, 83:289-300, 1988.
- [2] Cleveland, W. S. *Visualizing Information*. Hobart Press, 1993.
- [3] Cleveland, W. S. A Model for Studying Display Methods of Statistical Graphs. *Journal of Computational and Statistical Graphics*, 2:323-343, 1993.
- [4] Cleveland, W. S. *The Elements of Graphing Data*, Revised Edition. Hobart Press, 1994.
- [5] Heer, J., S. K. Card, and J. A. Landay. *prefuse: a toolkit for interactive information visualization*. *ACM Human Factors in Computing Systems (CHI)*. 2005.
- [6] Palmer, S. *Vision Science: Photons to Phenomenology*. MIT Press, 1999.
- [7] Tufte, E. *The Visual Display of Quantitative Information*. Graphics Press, 1983.
- [8] Tufte, E. *Beautiful Evidence*. Graphics Press, 2006.