At a Glance: Pixel Approximate Entropy as a Measure of Line Chart Complexity
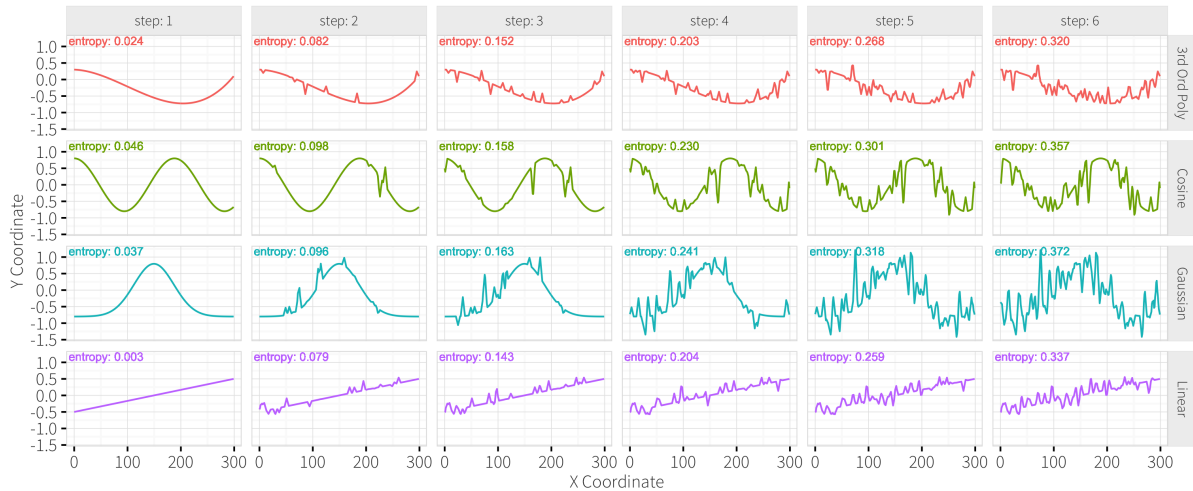

Fig. 1: Visualization of four line charts: 3rd order polynomial, cosine, gaussian, and linear (top to bottom). Noise is incrementally increased to each line chart from left to right in intervals of proposed approximate entropy measure (shown in upper left of each chart). This paper demonstrates that approximate entropy can be used as a metric for the perceptual complexity of line charts.

Abstract— When inspecting information visualizations under time critical settings, such as emergency response or monitoring the heart rate in a surgery room, the user only has a small amount of time to view the visualization "at a glance". In these settings, it is important to provide a quantitative measure of the visualization to understand whether or not the visualization is too "complex" to accurately judge at a glance. This paper proposes Pixel Approximate Entropy (PAE), which adapts the approximate entropy statistical measure commonly used to quantify regularity and unpredictability in time-series data, as a measure of visual complexity for line charts. We show that PAE is correlated with user-perceived chart complexity, and that increased chart PAE correlates with reduced judgement accuracy. 'We also find that the correlation between PAE values and participants' judgment increases when the user has less time to examine the line charts.

Index Terms—Visualization, Graphical Perception, Entropy, At-a-glance

---

1 INTRODUCTION

Information visualization has traditionally approached research from two directions: how to generate visualization forms, and how those forms are perceived by an end user. The latter branch of research has focused primarily on low-level perceptual questions, such as how a single data point is read and to what degree of accuracy. While this work has been useful and has supported an improved understanding of visualization theory, there remains a missing link between this very bottom-up approach to perception and the top-down approach to generation of visualization forms. In order to connect the two sides of visualization research, a higher level approach to the problem of perception is needed.

In almost all cases, especially visualization of larger, more complicated data, the first impression of a visualization centers on its overall shape and trend. Research in psychology has shown that during the initial glance of a visual stimulus, people perceive the higher-level spatial

and functional components of the natural scene [30, 29]. Extending this notion, we posit that a similar mechanism applies to the perception of visualizations. Many real world settings—e.g., emergencies or viewing quickly changing stock prices—are time critical and rely on the user to make judgements based on glances at a visualization.

Although it is anecdotally clear that more complex or noisy visualizations are more challenging to perceive [44], it is unclear how to measure this complexity outside of performing comprehensive user studies. Developing a measure of visualization complexity that is coherent with perceived complexity can have impact on a number of visualization domains. For example, designers can use the complexity measure to design more effective visualizations, especially for time-sensitive decision making tasks. Similarly, visualization recommendation engines such as the "Show Me" feature in Tableau [41] can take advantage of the complexity measure to detect when a visualization may be too complex and suggest an alternate design. To this end, there are a number of desirable characteristics of visual complexity measure:

1. *Correlated with perceived complexity:* The measure should correlate with user perception of chart complexity.
2. *Correlated with noise-levels:* The measure should correspondingly increase when more noise is introduced, since it is computed analytically (rather than through human measurements).
3. *Predictive of perceptual accuracy:* The measure should correlate

with the user's ability to accurately discern patterns in the chart.

4. *Simple:* The measure should be a single understandable value applicable to arbitrary 1D line charts.

5. *Widely Applicable:* The measure should exhibit the above characteristics across many types of line charts without the need for specialized tuning.

In this paper, we examine how people perceive, at a glance, the complexity of a line chart visualization, where we consider "at a glance" to be 200ms or less, approximately the amount of time required to visually process a scene [50]. We explore the use of approximate entropy [46, 45], a statistical measure used to quantify the amount of regularity and unpredictability in time-series data, as a measure to quantify perceptual complexity.

To this end, we conducted experiments to answer several research questions. Is approximate entropy an effective measure of time-series visualization complexity? Do users perceive higher approximate entropy visualizations as more complex and lower approximate entropy visualizations as simpler? As the approximate entropy of a visualization increases, does user accuracy in performing visual comparison tasks decrease? When users are given less time to study charts in order to complete a simple identification task, does the approximate entropy measure become more correlated with user accuracy?

We first ran an analytical experiment to compare approximate entropy with synthetically generated noise levels in visualizations. We then ran four perceptual experiments on Amazon Mechanical Turk [2]: the first asks users to select the most or least complex visualization from a line-up of visualizations with different approximate entropy measures. The result of the study confirms that users perceive visualizations with higher (lower) approximate entropy as more (less) complex. The second measures how the visual complexity (as defined by our entropy measure) of a chart affects the user's ability to detect changes in the chart. The third experiment is similar to the second, but studies the user's ability to identify basic shapes in charts of varying complexity. We find that visual complexity has a significant and large effect on judgement accuracy for both tasks, and that there is a threshold beyond which judgement accuracy degrades to random chance. The last experiment measures the interaction between the amount of time the user has to view a chart (the glance time) with their ability to perform the change detection task in from the second experiment. As the glance time is reduced to < 200ms, chart complexity becomes more highly correlated with judgement accuracy.

Our first perceptual experiment provides the basis for using approximate entropy to measure perceived complexity, and the subsequent perceptual experiments are based on real-world use cases. Understanding how people perceive visualization of complex data within a brief period of time can impact real-world usage in multiple ways. In disaster response scenarios, relief workers have limited time to examine data, and need to quickly get a gist of the available information. Similarly, in many real-time health care monitoring tasks, the typical visualization assumption of time to examine data in detail does not hold true. In all of these situations, the data seen at a glance is the only data the user sees. Understanding how this data is perceived is vital for design and evaluation of such cases, and we describe possible applications of this measure in the Discussion section.

2 RELATED WORK

The use of Approximate Entropy as a measure of perceptual complexity is related to research in psychophysics, perceptual psychology, and information visualization (infovis).

At a Glance Perception: The idea of at a glance perception has been widely studied in perceptual psychology. A well studied area is how people can recognize natural scenes in a short amount of time. Tasks such as rapid scene categorization and object recognition have been found to rely on a broad focus. Greene and Oliva found that in rapid scene categorization, people interpret a scene based on global, ecological properties that describe its spatial and functional aspects rather than by breaking it into objects [30, 29]. Similarly, Biederman et al. found that a person's ability to detect an object in a scene is dependent on specific relations between the object and scene as opposed to

specific characteristics of the object itself [8]. In contrast, data visualization tasks involve recognizing and decoding visually encoded trends and data values. This paper can be viewed as an initial extension of these ideas towards a potential measure of visual complexity for viewing data visualizations at a glance.

Perception of Salient Features: One method for quantifying perception of an image or visualization could be with a method capable of identifying its most salient features, or a measure quantifying the busy-ness of the image. Rosenholtz's work on "visual clutter", for example, seeks to quantify the amount of clutter in natural image displays. Notably, one of the measures from this work, Subband Entropy, uses entropy to quantify the redundancy in a natural image display and was found to be a reasonable measure of visual clutter [58]. Measuring clutter is related to measuring "visual complexity" in natural images [42], where a pattern is described as complex if the parts are difficult to identify or separate from each other.

In the visualization community, research has sought to quantify the salient features of a visualization. Scagnostics is one such example by Wilkinson et al., where multiple metrics were proposed to categorize the perception of scatterplots [68, 43]. This work has led to a number of advances, including research that extend the concept of Scagnostics to parallel coordinates [22], pixel-based displays [59], and generalized techniques for dimension reduction of high-dimensional data [7]. However, scagnostics primarily focuses on identifying meaningful relations to visualize with scatter plots, while pixel-based diagnostics focus on intensity-based visualizations like Jigsaw Maps and Pixel Bar Charts. Our work is similar in spirit to these prior work, but instead we focus on the perception of 1D line charts and the quantification of the visual complexity of these visualizations.

Other work has been done on visual analysis and simplification of time series data. Heer et al. measure the effect of chart size and layering on speed and accuracy in visual comparison tasks [33]. ASAP uses kurtosis to guide time-series smoothing to preserve trends and anomalies while reducing cyclic patterns and noise [57]. Numerous measures, such as *L1 Local Orientation Resolution*, have been developed to select aspect ratios for line charts [66]. Our complexity measure is complimentary to these approaches and can help guide selection of visualization parameters or smoothing.

Perception of Visual Marks and Visual Forms: The study of at a glance perception has led to much cross pollination between infovis and perceptual psychology. Substantial work from perceptual psychology suggests that modeling a holistic shape envelope could capture a fundamental aspect of perceptual encodings of visualizations. At the first glance of an image, users tend to focus on the "big picture" [38], which should produce a far more compact representation of the most salient information in the image [25]. For line charts, this initial big picture is likely to be the holistic shape envelope that surrounds the values.

Infovis research has studied visualization perception for short glance times. However, the emphasis has been on pre-attentive processing of properties of visual marks [32] (e.g., color, orientation, etc.). Fewer works focus on perception of higher-level visualization forms. Szafir et al. recently measured how people can quickly perceive summary statistics (for instance centroid, or density) from visualizations such as scatterplots [62]. They investigated four visual statistical tasks: identification of sets of values, summation across values, segmentation of collections, and estimation of structure. Related works show that some visual tasks—such as correlation from scatterplots [52], mean size of homogenous elements in an array [18]—occur within the pre-attentive processing phase (<200ms).

Another related area is the interplay of glance time and a person's processing of visual information. Based on findings from cognitive science, it may be that shorter glance times are difficult for users because they need to make use of short-term memory to perform visual analysis [39]. Short-term memory is limited [20], decays over time [14], and is expensive to use. For instance, in an experiment conducted by Ballard et al., subjects *serialized* their tasks in order to avoid using short-term memory [5]. This relationship between latency and memory is consistent with other recent papers in the visualization
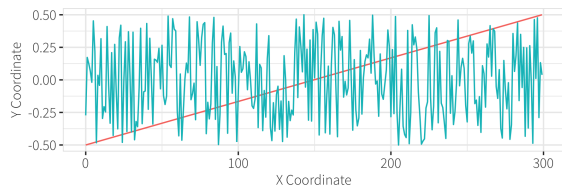
Fig. 2: Linear line and its shuffled variant will both have the same level of entropy based on Shannon Entropy. However the linear line is intuitively less complex than its shuffled variant.

community related to the measure of memorability [40, 11]. We study the interaction effects between glance time and the user's ability to perform basic visual judgements.

Use of Entropy in Visualization: There has been a long history of the use of entropy in the field of visualization. A recent book by Chen et al. summarizes a wide range of applications of entropy and information theory in visualization [15]. Further, scientific visualization research has leveraged entropy to e.g., allocate computational resources [65], choose rendering properties [64], and select compression techniques [63].

In the field of information visualization, similar to scientific visualization, entropy has been used to measure the amount of noise and information in the data. For example, Chen and Jaenicke use entropy to detect the amount of information in a visualization, and ways to optimize the design of visualization [16]. Dasgupta et al. use entropy to measure visual uncertainty to preserve privacy in parallel coordinates [21]. Biswas et al. use entropy to measure importance and model relationships of variables in a multivariate dataset [9]. Lastly, Karloff and Shirley use entropy to determine an optimal summary tree for large node-weighted rooted trees [37].

However, these works primarily use entropy to measure data quality, rather than as a proxy for visualization *perception*. In fact, Chen et al. note that while related, information theory is about efficient communication rather than perceptual and cognitive processes [15], and that compressing the data in a visualization into the minimum set of bits may not result in the best visual representation. In contrast, this paper establishes, quantifies, and explores the relationship between information theoretic notions of randomness (approximate entropy) with graphical perception.

Recent work by Rensink et al. theorized that the perception of correlation in scatterplots can be explained by measuring the entropy of the data points in the scatterplot [54]. Although our study of approximate entropy focuses on the perception of complexity of line charts, we share the same research goal of evaluating entropy as a possible perceptual complexity measure.

3 A LINE-CHART COMPLEXITY MEASURE

Since entropy has traditionally been used to measure the amount of "disorder" in data and has been utilized in perceptual psychology as a proxy of "visual clutter", it is reasonable to apply the same concept to measure the complexity of a line-chart visualization. However, the classic measure of entropy is computed over an unordered set of values, and does not work for ordered visualizations. Figure 2 illustrates an example—both the linear line (red) and its shuffled variant (blue) have the exact same entropy measure.

We therefore surveyed the signal processing, information theory, and statistical modeling literature and identified 8 candidate entropy measures for line charts. We then selected a subset based on the five desirable characteristics of a complexity measure described in the Introduction. These were selected to described their scope and ease of application, as well as relation to user perception, rather than reliant on any particular notion of complexity. For instance, we do not want to assume that "more jagged" shapes are more complex.

In reviewing possible measures, we encountered a wide variety of complexity measures with correspondingly varying definitions of complexity for a given chart. Our goal when conducting this review was not to attempt to define complexity itself, but to identify a mea-

sure that approximates how well users will be able to perceive and use a chart. To this end, we only imposed two analytical constraints based on the first and fourth desired criteria: that the measure should identify more noisy signals as more complex, since noisy signals are anecdotally more difficult to perceive [44], and that the measure computes a single scalar number.

- *Signal to Noise Ratio*: SNR is a signal processing method that measures the relative power of the desired signal to the noise overlaying that signal.
- *Auto Correlation*: A method commonly used in Signal Processing that quantifies the amount of repetitions in a time series by measuring the correlation of the time series with a lagged version. Since a more random time series will have less repetition, this approximates the amount of randomness in the time series.
- *Fourier Analysis*: Fourier Analysis [12] is also a common method in signal processing. It transforms the data into the frequency domain, making it possible to quantify the extent to which the data is made up of different frequencies. Since more random data should result in higher frequency changes, this can also be used to approximate randomness by measuring the high frequency components of the signal.
- *Approximate Entropy*: Approximate Entropy [45, 46, 36] is a robust statistical measure of repetitiveness. It is based on randomness statistics for chaotic functions from Information Theory. Like Auto Correlation, it measures how much components of the signal repeat, but derives an entropy measure for time series data.
- *Sample Entropy and related statistics*: These modifications of Approximate Entropy remove possible biases [56, 69, 17].
- *Multiscale Entropy*: This applies approximate or sample Entropy multiple scales to analyze how a signal may be more or less chaotic at varying scales [19].
- *Flattened Signal Length*: This method can be understood as 'stretching' the data until it becomes flat, intuitively, more complex charts will result in longer flattened lines [6].
- *Sequential Modeling*: Measuring the ability of a Hidden Markov Model to predict the signal. This is a classic statistical modeling method for sequential data. Intuitively, a signal that cannot be easily modeled will be more random [4].

Although we started with these measures, we found that most were not appropriate, or did not satisfy the desirable criteria listed in the Introduction. Fourier analysis and auto-correlation did not analytically correlate closely with added noise to a given line chart (e.g., those shown in Figure 1). Hidden markov models require tuning a number of hyperparameters—such as dimensionality and the initial state—that vary across charts. Similarly, SNR requires a pre-existing model for the desired signal that is unlikely to exist in a real world application. Flattened Signal Length does not account for repetition in the data, thus it can, for example, assign a sinusoid a higher complexity than random data, although the random data visually appears more complex.

Approximate, Sample, Fuzzy, and Multiscale Entropy are all similar, in that Sample Entropy and Fuzzy Entropy are both bias corrected versions of Approximate Entropy, and Multiscale Entropy applies Approximate/Sample Entropy at multiple scales. Multiscale Entropy violates the goal of a *simple* measure because it generates measurements for each scale. We found that in practice, Approximate Entropy and Sample Entropy tend to be very close in value, however, Sample Entropy is sometimes not defined for low entropy charts. We therefore selected Approximate Entropy as the the candidate measure for study in this paper.

3.1 Approximate Entropy

Approximate entropy is a family of system parameters and related statistics developed by Pincus to measure changes in system complexity [46]. In particular, the statistic is designed to be effective at distinguishing complexity in low dimensional systems when only relatively few (tens to low thousands) points are available. This property makes
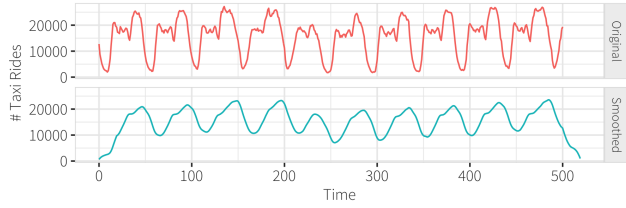
Fig. 3: Example of time series taxi usage in New York City and a smoothed variant.
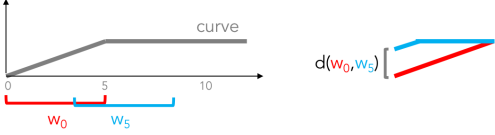

Fig. 4: Example of a curve (grey) and two windows (w_0 and w_5) on the left. The right side shows that the distance function $d(w_i, w_j)$ is simply the maximum difference between aligned pairs of y values.

it an effective measure for line chart visualizations, which are low dimensional and often contain hundreds of pixels (at most one point per pixel).

Approximate entropy quantifies the unpredictability of changes in a series of points. Intuitively, a series of points with more repeated patterns is easier to predict. Approximate entropy reflects the probability that such similar patterns will not be repeated. A line chart without any repetitions, such as stochastic noise will have a very high entropy. In contrast, taxi cab demand in New York City (Figure 3, top), which exhibits very large changes between high and low demand periods, will tend to have a lower entropy because the demand follows a regular daily and weekly pattern. Notice though that the smaller variations in the original plot cause it to have larger entropy than the smoothed version (Figure 3, bottom).

Conceptually approximate entropy is computed using a sliding window approach. Given N samples of a continuous curve, each window of size m is compared with every other window of the same size. If there are many pairs of similar windows, then there is more regularity in the curve and the score should be lower.

Formally, let y_i be the ith sample of the input curve, and $w_i^m = [y_k|k \in [i, i+m]]$ be the ith window of size m. Thus, there are a total of $W = N - m + 1$ possible windows of size m for a line with N values. Figure 4 illustrates a grey curve and two examples windows w_0^5 and w_5^5 of size $m = 5$. We define the distance between two windows as the maximum difference between aligned values:

$$d(w_i^m, w_j^m) = max_{k \in [0,m]}|y_{i+k} - y_{j+k}| \tag{1}$$

Then, the similarity $S_i^m(r)$ for a given window w_i is defined by the percentage of windows whose distance is below a threshold r[1]:

$$S_i^m(r) = \frac{|\{w_k | k \neq i \wedge d(w_k^m, w_i^m) < r\}|}{W} \tag{2}$$

We can now combine the similarity scores for all possible windows into $\Phi^m(r)$ by summing their log transforms. This can be understood as taking the log probability that windows will be closer than r. Note that there is a Φ for each user-defined window size and threshold.

$$\Phi^m(r) = \frac{1}{W} \sum_{i=1}^{W} log(S_i^m(r)) \tag{3}$$

Pincus [46] originally defined the approximate entropy to be the difference $\Phi^m(r) - \Phi^{m+1}(r)$ as the number of samples from the curve N increases to ∞. Intuitively, this difference measures the increased probability that sequences will be greater than r apart when the sequences length increases by one.

$$E(m,r) = lim_{N \to \infty}[\Phi^m(r) - \Phi^{m+1}(r)] \tag{4}$$

---
[1] r is in the same units as y.

However, sampling infinite points is not realistic, and thus approximate entropy is estimated using a fixed N:

$$E(m,r,N) = [\Phi^m(r) - \Phi^{m+1}(r)] \tag{5}$$

This is a natural fit for visualizations, where we can set N to the width of the chart in pixels.

3.2 Pixel Approximate Entropy

We define the *pixel approximate entropy* as the approximate entropy of a line chart visualization. To do so, we use the following procedure:
1. Scale the dataset by mapping its values into the visual domain as positional variables, so that the x and y data values are in terms of pixels.
2. Construct a vector $Y = [y_i | i \in [0,N]]$ where y_i is the pixel y-coordinate for the curve's i^{th} pixel along the x coordinate. N is the pixel width of the chart.
3. Compute $E(m,r,|Y|)$.

Pixel approximate entropy (PAE) calculates approximate entropy based on pixels in the visualization itself. The benefit is that it is independent of the data complexity, and provides a consistent range of entropy values for a given chart resolution.

Finally, note that we have chosen to compute PAE as a single global complexity measure based on the positionally encoded data in the line chart. Alternative measures that e.g., capture both local and global complexities, or account for additional visual encodings, are promising extensions of this work.

3.3 Examples of Pixel Approximate Entropy

To provide a sense of PAE values for different types of line charts, Figure 1 illustrates four base visualized curves and their PAE values (the text in the upper left of each plot shows the PAE measure). We show a third order polynomial (top), cosine function (2nd row), Gaussian distribution (3rd row), and a linear line (bottom). We chose these curves because they are representative of common visualized data in practice. The linear line is the simplest shape that users commonly encounter and we chose it for its simplicity. The gaussian distribution is arguably the most well recognized distribution, and models natural phenomena such as sizes of living tissue (e.g., length, height, weight), stock distributions [10], intelligence [35], and other societal and scientific data. The third order polynomial represents a more complex pattern that is commonly used in scientific models such as thermodynamics [61] and kinematics. Finally, the cosine function represents cyclic patterns such as heart beats, and temperature over time.

Each column in the figure, going from left to right, adds more random noise to the curve making it more "complex". We can see that the curves become seemingly more random, however the overall shapes are still evident.

To demonstrate how Pixel Approximate Entropy works in practice, we provide several examples of its behavior. Figure 5 shows how altering the scale of a chart effects its entropy measure. Increasing the relative height of the chart will increase the entropy, while increasing the width will decrease entropy. These scaling effects can be understood as either increasing the relative noisiness of the chart (for height), or having a smoothing effect by stretching the chart out (for width).

Figure 6 depicts PAE for several real world data sets. In practice PAE can be interpreted as the amount of space on the chart that is taken up by *unpredictable* data. Charts of data that exhibit less noise, such as the S&P 500 stock pricing data (Figure 6a) and NYC taxi ride volume data (Figure 6c,) have relatively low PAE. Charts with more irregular, noisy data, such as the S&P 500 trade volume data and EEG seizure data in Figures 6b and 6d, have much higher PAE.

4 EXPERIMENTS

This section presents four experiments to evaluate Pixel Approximate Entropy as a visual complexity measure consistent with the desired criteria in the Introduction. We translated these criteria into the following hypotheses. Each hypothesis corresponds to one experiment, and we describe the hypotheses in more detail in the corresponding experiment subsection:

(a) Original Data, Entropy: 0.196

(b) Width ×2, Entropy: 0.0906

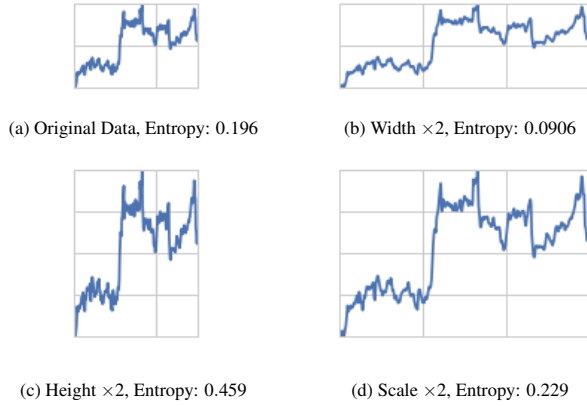(c) Height ×2, Entropy: 0.459

(d) Scale ×2, Entropy: 0.229

Fig. 5: Scaling effects on Pixel Approximate Entropy. Increasing the height of the chart will increase PAE, while increasing chart width will decrease PAE.
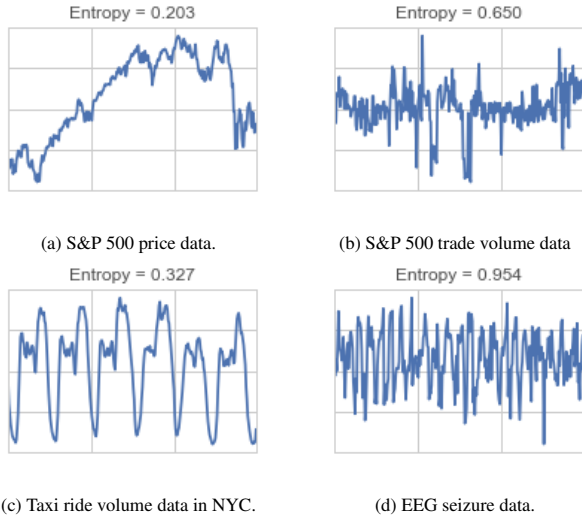

(a) S&P 500 price data.

(b) S&P 500 trade volume data

(c) Taxi ride volume data in NYC.

(d) EEG seizure data.

Fig. 6: Examples of Pixel Approximate Entropy applied to real world data. More visually noisy data, such as stock trading volume and EEG data collected during a seizure, has higher approximate PAE.

- **H1:** There is a statistically significant correlation between PAE and the amount of noise added to the chart.
- **H2:** PAE is an effective measure of perceived complexity, such that there is a statistically significant correlation between participants' perception of complexity and the line chart's PAE.
- **H3:** Varying chart PAE affects participants' ability to perform the visual task of identifying changes in a line chart.
- **H4:** Varying chart PAE affects participants' ability to perform the visual task of identifying the base function of a chart.
- **H5:** Reducing the amount of time participants are given to study line charts, and the PAE of a line chart "at a glance" affects comparison accuracy on charts.

Experiment 1 verifies that PAE is correlated with the amount of noise added to chart (**H1**). We then present four user studies that use both controllable synthetic charts as well as charts from real-world medical and financial datasets to evaluate the user's ability to perform perceptual tasks at varying PAE levels. Experiment 2 uses the Line-Up [67] protocol to test PAE's correlation with perceived complexity (**H2**). Experiment 3 and 4 test the effect of PAE on two visual comparison tasks—matching identical charts and identifying the underlying shape in a chart (**H3,4**). Experiment 5 studies the interaction between glance time and PAE when matching identical charts (**H5**).

4.1 Experimental Setup Overview

We now describe the shared experiment setups.

Noise Generation: For the synthetic data used in our experiments, we systematically introduce noise to control the PAE value of a chart. To do so, we iteratively add noise to a baseline chart until its measured PAE reaches the desired value. Figure 7 depicts how triangle noise is added to a given curve. We sample the magnitude of the noise Δ_y from a uniform distribution $U(-\sigma, \sigma)$, where σ is the standard deviation of the data, and add that value to the y_i value of a random x coordinate x_i. Thus we set $y' = y + \Delta_y$. We then linearly interpolate y' with the value of the curve at $x_i \pm \Delta_x$.
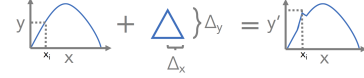

Fig. 7: Noise Δ_y is added at a point along the input curve and interpolated by Δ_x to each side (e.g., add triangle to a region on curve).

Approximate Entropy Parameter Selection: Approximate entropy is parameterized by m and r (see Equation 5), which we set to $m = 2$ and $r = 20$ in our experiments. We selected these values by synthetically generating a training set of line charts with varying amounts of generated noise and found values that maximized the average PAE-to-noise level correlation across the training set.

Specifically, we started with four basic curves, added varying amounts of noise (see Figure 1 for examples), and rendered the results at different visualization resolutions ($\{100 \times 150, 200 \times 300, 400 \times 600\}$). We swept the m and r parameters and found that the $m = 2, r = 20$ parameter settings were most robust across the visualizations. For consistency, all charts are 300×200 pixels.

User Study Setup: We used Amazon Mechanical Turk [2] to recruit participants for the user studies. In the default setup, participants were given a consent form, a training exercise that introduced the task, a brief qualification test, and a demographics survey at its completion. Participants were paid per task at an estimated rate of $\geq \$8.00$ per hour, and all assignments included a 1.00 USD bonus for completing all tasks. For experiments 3 through 5, which involve simulated visual tasks, participants were given an overall time limit based on the assumption they would spend at most 20 seconds per task, with an additional 3 minutes to read the instructions and complete survey. This time limit was intended to keep participants focused while allowing for the possibility of minimal interruptions. In practice, most participants completed each task in less than 4 seconds.

Due to the simplicity of the task in Experiment 2, the qualification task was not included, and participants were simply paid a bonus based on the number of correct answers. Workers were required to have an approval rating of at least 80% and be residents of the United States. We selected an approval rating less than 95% to engage workers that would attempt to perform the tasks quickly and without too much effort—thereby simulating a routine visual task.
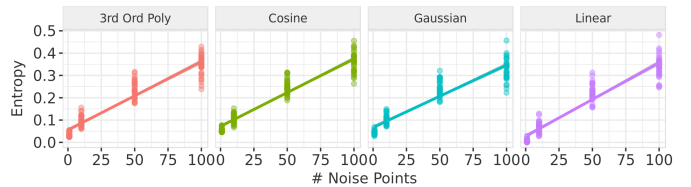
4.2 Experiment 1: PAE and Synthetic Noise

Fig. 8: Correlation between noise level and entropy measure

Our experiments rely on controlled experiments that carefully vary the "complexity" of a chart by adding or removing noise as described above. Since we use PAE as a proxy for complexity, this experiment first establishes the correspondence between the amount of noise added to different basic line chart functions, and the resulting chart's PAE. We select basic functions that tend to appear in many different types of charts: linear, Gaussian distribution, cosine, and 3rd order polynomial.

Figure 8 plots the relationship between noise and PAE. Additionally, for each base function, we test correlation through linear regression between the independent (amount of noise) and dependent (PAE) variables. We evaluate the regression with a t-test against the null hypothesis that there is no correlation, and by the coefficient of determination R^2. To summarize the t-test and R^2 for each base function: 3rd Order Polynomial ($t = 37.3$, $p < .0001$, $R^2 = 0.91$), Linear ($t = 37.8$, $p < .0001$, $R^2 = 0.91$), Cosine ($t = 36$, $p < .0001$, $R^2 = 0.90$), Gaussian ($t = 31.7$, $p < .0001$, $R^2 = 0.88$).

Based on these findings, we accept **H1**, that the PAE of a chart correlates closely with the amount of noise added to the chart. We use this result in the synthetic data generation used in other experiments, in which we add noise to control the PAE of a given visualization.

4.3   Experiment 2: PAE and Perceived Vis Complexity

Does PAE correlate with user perceived notions of visualization complexity? In this experiment, we test the hypothesis that PAE is positively correlated with perceived line chart complexity. To understand whether PAE is effective beyond synthetically added noise, we run two studies—one using synthetic charts and one using charts drawn from medical and stock datasets. We hypothesize that participants' identification of the most and least complex chart from a line up will correlate with (**H2.1**) the chart's PAE, and (**H2.2**) the underlying base function of the chart.

User Tasks: We use two user tasks: one for charts that vary PAE using synthetically generated noise, and one for charts of real data that naturally have varying PAE levels. For both, the training page informed the participant that she would be shown 20 (for synthetic noise, or 16 for real data) sets of similar charts, and asked to pick the most or least complex chart in a given set. It also showed a figure of three example charts and labeled the most and least complex. Participants then completed the 20 (or 16) judgments. Each judgment consisted of selecting the most or least complex chart out of a set of eight charts with varying PAE. Because we are interested in discovering whether or not a correlation exists between perceived complexity and PAE we use the LineUp method from [67] for this experiment.

The synthetic noise tasks generate charts based on five base functions: linear, cosine, gaussian, third order polynomial, and S&P 500 stock data. For the sample of stock data, we use a 300 sample window of S&P 500 daily closing prices that has the median PAE taken from a dataset of S&P 500 stocks over a 15 year period. Each judgment set consists of eight charts from the same base function perturbed with noise to meet a target PAE from 0.1 to 0.8 (in steps of 0.1), and arranged in a random order. For example, the set shown in Figure 9 shows the judgment for the cosine function.

The real data task used data from three datasets (S&P 500 historical stock price and volume data, the MIT-BIH Arrhythmia Database, and the Bern-Barcelona EEG Database [3, 28]), and randomly selected time intervals of the data such that the resulting charts had PAE from 0.1 to 0.8 (in steps of 0.1); they are arranged in random order.

Complexity of Graphs
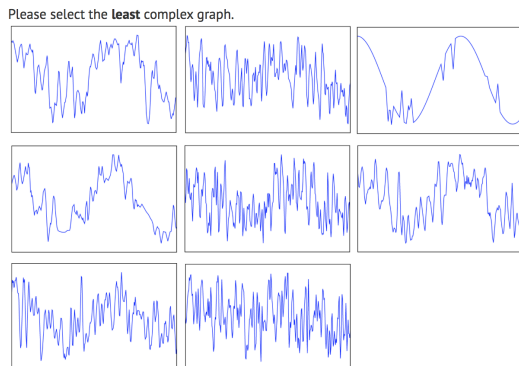
Please select the **least** complex graph.

Fig. 9: Example judgement set for cosine function in Experiment 2.

Each participant was shown the same eight chart judgment set for each baseline function or dataset four times. Two times the participant was asked to select the most complex chart in the set, and two times she was asked to select the least complex. The ordering of judgment sets and position of the target chart was randomized. We use the two judgments of the least complex chart as the attention check to prevent Mechanical Turk participants from gaming the system [26]. If a participant does not select the same chart as least complex during the experiment, we remove the participant's data from consideration.

Participants for the synthetic noise (real data) task made 20 (16) judgments. 50 participants attempted the task; after dropping those that failed the attention check, 33 (25) remained. Of these, 45% (28%) were female; 76% (76%) were between the ages of 25 and 49; 55% (48%) held high school degrees and 36% (52%) held a Bachelor's degree; and 85% (80%) spend upwards of 30 hours per week on a computer. Notably, 33% (52%) ranked themselves as having intermediate expertise in statistical visualizations and 42% (28%) identified themselves as novices in the field.

Results and Statistical Analysis: To test our two hypotheses, we take guidance from [24] and perform a binomial logistic regression, where correctness is the outcome variable, and PAE and the baseline function are explanatory variables. Correctness is defined as a binary variable based on whether or not a participant selected the chart with the least PAE as the least complex, and vice versa for most complex.

The result of the binomial logistic regression (and a follow-up Wald test for baseline function, because it is categorical) indicates a significant association between PAE and correctness ($Z = -4.78, p < 0.001$), and a significant association between baseline function and correctness ($\chi^2(4) = 10.6, p < 0.05$) for the synthetic data. Similarly, for the real data there is a significant association between PAE and correctness ($Z = -5.61, p < 0.001$), but no significant association between base function and correctness.

Discussion of Results: We find that PAE might be used to approximate perceived chart complexity. The binomial logistic regression shows that for synthetic and real data PAE is significantly associated with what charts participants select as most or least complex (**H2.1**). This trend holds for charts from synthetic and real data, and suggests that PAE is a reasonable proxy for perceived complexity of line charts in practice. For synthetic data, we find that underlying base function is significantly associated with what charts participants select as most or least complex (**H2.2**).

In analyzing results, we notice that participants typically displayed better accuracy in identifying simple charts than complex ones. This suggests that small differences in PAE are easier to spot in charts with lower PAE than with higher PAE. Figure 10 shows this case for a gaussian. Increasing the PAE by $\Delta = 0.125$ to a chart with low PAE (0.015) is easier to discern than adding Δ to a higher PAE chart (0.315). This suggests a phenomena akin to just noticeable difference (JND) for PAE, which we explore in later experiments.
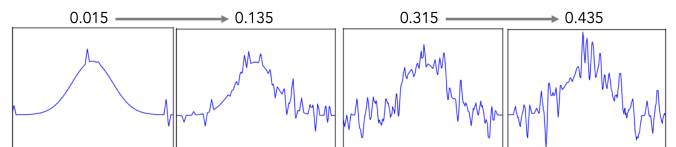

Fig. 10: Gaussian function with 0.015, 0.135, 0.315, 0.435 added PAE.

4.4   Experiment 3: Find-the-Difference Task

Do varying levels of chart PAE affect the ability to *perform* visual comparison judgements? In this experiment, we show participants one chart for a short duration (the length of a saccade), hide it for some time using a mask, and then show a copy of the initial chart alongside another chart with higher or lower PAE. The participant is asked to select the initial chart from the set of two. Participants were given the following direction: "*This HIT consists of a series of graph comparisons. For each comparison, you will first be shown a original chart for a fraction of a second. You then be shown two similar charts and*
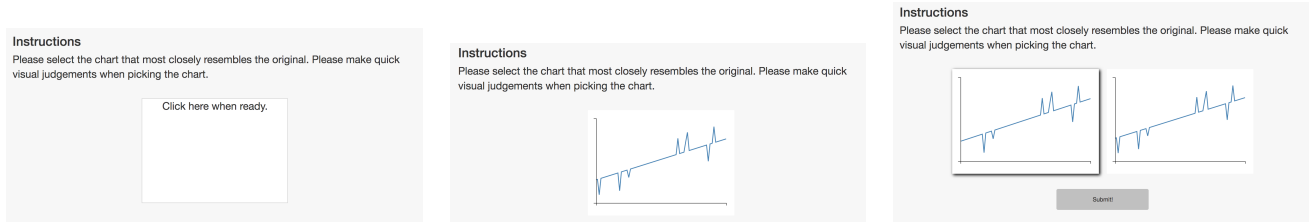
(a) The participant clicks the button to indicate that she is ready to make a judgment.

(b) An initial chart is flashed for the *glance time*, and then hidden and replaced with a mask during the *pause time*.

(c) After the *pause time*, we show two options: the initial chart and the initial chart with PAE added/removed. The order of the charts are randomized.

Fig. 11: Experiment 3 screenshots.

asked to pick the one which most closely resembles the original flashed chart. Please make quick visual judgements and only spend a few seconds when picking the chart.*" The charts are all generated by adding noise to a base function as specified in the noise generation section. We hypothesized that:

- **(H3.1)** Task accuracy is affected by the PAE of the initial chart

- **(H3.2)** Task accuracy is affected by the magnitude of the PAE difference between the initial and alternative chart.

- **(H3.3)** Task accuracy is affected by the sign of the PAE difference (the alternative chart has higher or lower PAE than the initial)

- **(H3.4)** Task accuracy is not affected by the underlying chart type (linear, cosine, polynomial, gaussian).

The setup for Experiment 3 was inspired by Just Noticeable Difference (JND) studies that show two slightly different charts side by side and asking the user to pick the higher chart (for some measure of "higher"). By using a staircase protocol that incrementally increases or reduces the differences between the two charts, researchers can find the JND where users are accurate less than 75% of the time [55, 31]. Figure 11 shows screenshots of the experiment.
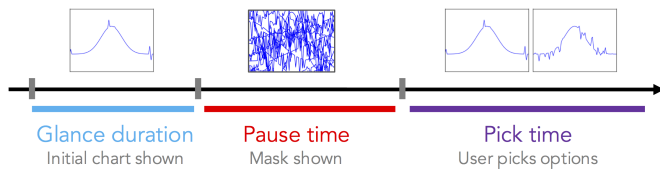

Fig. 12: Timeline of the at a glance task. We first show the initial chart for *glance time*, then hide the chart and don't show anything for a *pause time*, and then show the participant options to choose from.

Each judgment follows the process illustrated in Figure 12. Participants first view an initial chart that flashes for a glance time of 200ms, the duration of one saccade, or slightly longer than the basic stage of correlation perception of 150ms proposed by Rensink et al., and the amount of time required for eye movements to be optimally guided [53, 50]. The initial chart is replaced with a mask that is shown for 200ms of pause time. Masking is used to interrupt the perceptual processing so that user responses are due to cognitive, rather than low-level perceptual pattern matching [51, 60, 27, 13, 47, 48]. The mask, shown in Figure 12, is designed to be similar in visual structure as a line chart, but does not convey any information related to the user task. After the pause time, the participant is shown the initial chart and the initial chart with Δ_y more or less PAE, and asked to choose the initial chart. The order of the two charts is randomized to reduce learning effects.

Our protocol differs from classic JND protocols in that it focuses on the ability to identify a change in chart complexity after the initial glance time. Further, since our goal was not to find this just noticeable threshold (although that is a direction for followup work), we did not directly follow the staircase protocol. We instead sweep through different chart PAE values to confirm that task accuracy is affected by chart PAE.

Based on pilots, the initial chart has low (0.045), medium (0.09), and high (0.18) PAE. The comparison chart differed in PAE by Δ_y of

± 0.015, ± 0.03, ± 0.06, ± 0.09, and ± 0.12. We evaluate these conditions for all four chart types (line, cosine, gaussian, and third order polynomial), and both increasing and decreasing PAE, resulting in a $3 \times 5 \times 2 \times 4 = 120$ factorial design.

Results and Statistical Analysis: There were 52 participants, who took on average 7 minutes and 5 seconds to complete the tasks in the experiment, or approximately 3.5 seconds to complete each comparison task. The participants had the following demographics: 63% male; 65% between 25 and 39 years old; 58% held Bachelor's or Master's degrees; and per-week computer usage was fairly even from 21 to > 60 hrs. Additionally, 67% ranked themselves as low-intermediate, intermediate, or high intermediate level visualization users.

In data collection we recorded the correctness of a participant's response for each judgment, and used this as the binary response variable in our analysis. To test our four hypotheses we performed a binomial logistic regression where we used the base function, initial chart PAE, magnitude of PAE difference, and the sign of the difference, to predict task accuracy. We also used a a follow-up Wald test for baseline function, because it is categorical. We found that all independent variables are predictive of task accuracy. Their correlation and significant scores are summarized as: **(H3.1)** Entropy of the initial chart ($Z = 2.42, p < 0.05$); **(H3.2)** Magnitude of PAE difference between the two charts ($Z = -5.08, p < 0.001$); **(H3.3)** Sign of the PAE difference between the two charts ($Z = -8.06, p < 0.001$); and **(H3.4)** Baseline function ($\chi^2(3) = 9.32, p < 0.05$).

These results imply that the initial chart's PAE, magnitude and sign of Δ_y, and chart type have an effect on participant's accuracy in identifying the initial chart in the gallery. Figure 13 illustrates a clear effect of initial PAE on accuracy. Higher initial PAE (blue) systematically reduces accuracy as compared to low initial PAE (red), especially for the linear chart when Δ_y (x-axis) is small. Accuracy also increases as the magnitude $|\Delta_y|$ increases, meaning it is easier to identify large changes in visual complexity. Across all conditions, we find that judgment accuracy converges to 0.5 (random chance) as Δ_y decreases towards 0.
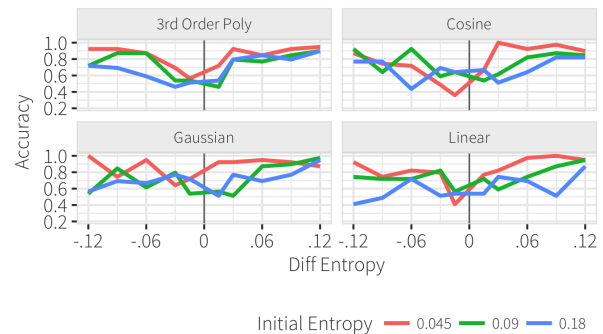

Fig. 13: Average accuracy as Δ_y of PAE varies. Each facet shows a different chart type, and each line is a different initial PAE.

Interestingly, the sign of Δ_y affects task accuracy. To investigate this, we grouped participant judgments by whether Δ_y is positive (the alternative chart has higher PAE) or negative (vice versa). Figure 14 plots the 95% bootstrapped confidence intervals, faceted by the magnitude of the PAE difference $|\Delta_y|$. When $|\Delta_y|$ is low (≤ 0.01), participants are more accurate when the change is positive than negative.

This happens because participants preferentially choose the lower PAE chart, which is correct when Δ_y is positive, and incorrect when negative. As the magnitude of the PAE difference increases, the accuracy for both signs increas, however the bias also persists.
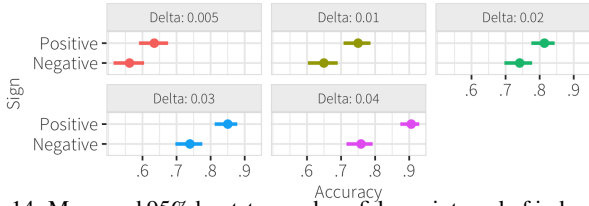

Fig. 14: Mean and 95% bootstrapped confidence interval of judgement accuracy for increasing (positive) or reducing (negative) the PAE of the alternative chart options. Each facet is a different Δ_y magnitude.

Discussion of Results: Based on the results, we accept all four hypotheses: **H3.1**, **H3.2**, **H3.3**, and **H3.4**. We find a clear trend towards 50% accuracy (random guess) as the difference in PAE between the two charts decreases, irrespective of other conditions. Additionally, as the initial PAE increases (the chart is more complex), so too does the necessary change in PAE in order to accurately differentiate the charts. This supports our finding in Experiment H2, that participants are worse at distinguishing between changes for high PAE charts than they are for low PAE charts or when the difference in PAE is large.

Further, given two charts, participants have a systematic bias towards choosing the chart with lower PAE. We do not have an explanation of why this might be, but hypothesize that it might be because the lower PAE charts appear closer to the "shape envelope" of the initial chart, especially when viewed at a glance. However, further studies are needed to evaluate this conjecture.
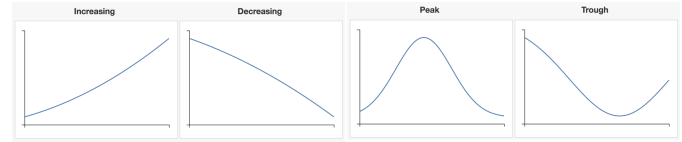
4.5 Experiment 4: Shape Identification Task

This experiment uses the same procedure as experiment 3, but for a different visual judgment task: identifying the overall shape of a chart. Many applications of visualizations involve users attempting to identify patterns in potentially noisy data, and we designed this experiment to simulate this use case. To this end, we provided the participants with examples of 4 underlying base functions (shapes) to study. Then, after adding enough noise to a base function to reach a desired PAE level, users are asked to identify which of the 4 underlying functions the noisy chart represents. The hypothesis is that the PAE of the chart affects the ability to accurately identify the underlying function of the chart (**H4**). Note that in the extreme, there can be so much noise that the charts are quantitatively identical irrespective of the initial base function, and the accuracy should converge to random guessing.

Participant Tasks: Participants first view an initial chart that flashes for a *glance time* of 500 ms ($\approx 2 - 4$ fixations). We allowed a slightly longer glance time since identifying the underlying shape of a chart is more difficult than the previous task of matching charts. The participant is then asked which of the following four shapes is most representative of the chart: increasing trend, decreasing trend, peak, and trough (Figure 15). They are also asked to self-report their answer certainty, from 0 for 'guess' to 4 for 'certain'. Chart order was randomized to reduce learning effects. Participants were asked to make quick judgments, and we limited their time to complete the experiment 30 minutes maximum. For each base function, we showed users 5 versions at each of four PAE levels (0.2, 0.4, 0.8, 1.2), resulting in a $4 \times 4 \times 5 = 80$ factorial design.

Results and Analysis: There were 47 total participants, who took on average 8 minutes to complete the tasks in the experiment, and approximately 3.6 seconds to complete each comparison task. The participants had the following demographics: 62% fell between the ages of 25 and 39, computer usage was spread evenly, although 88% used computers for at least 20 hours per week, 57% held Bachelor's or more advanced degrees, 76% were male, and 79% ranked themselves as some level of intermediate user with statistical visualizations.

We performed a binomial logistic regression using PAE as the independent variable to predict task correctness as the dependent variable.

(a) The four basic chart shapes users needed to identify.
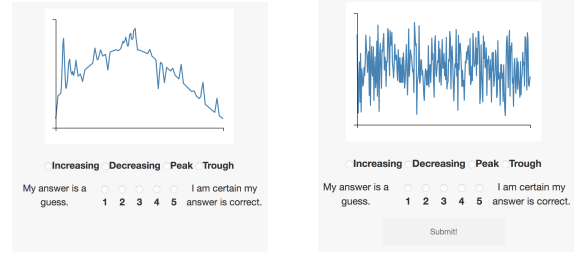
(b) Peak shaped chart with 0.2 PAE. (c) Increasing trend chart with 1.2 PAE.

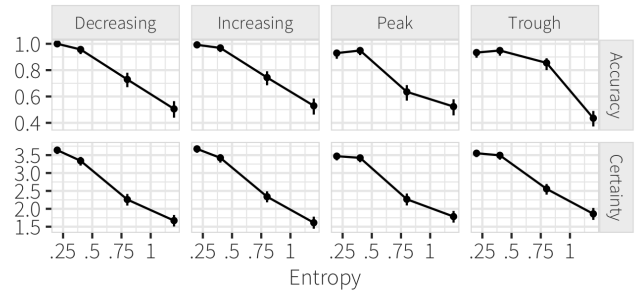Fig. 15: Screenshots depicting Experiment 4.

Fig. 16: Chart shape identification accuracy and certainty as entropy changes. Each facet shows a different chart type.

We found a significant correlation ($Z = 25.75, p < 0.001$), which supports **H4**. Figure 16 shows that users were able to correctly identify the underlying shape of a chart with close to 100% accuracy when the PAE (x-axis) of the chart, with the added noise, is low. However, user accuracy drops to $\approx 70\%$ for a PAE of 0.8, and to $\approx 50\%$ for a PAE of 1.2. Further, we find that answer certainty consistently decreases as the PAE increases.

Discussion: For the shape identification task, we found that there is a clear trend towards lower accuracy in distinguishing chart shapes as the entropy increases, indicating that users find it more difficult to perceive meaningful differences in charts with high entropy. We expect that there is a maximum PAE where charts are quantitatively the same irrespective of the base function, and users resort to random guessing. We do not believe our experiment approached this limit, because the lowest accuracy was still higher than random chance (25%), and the lowest average certainty was ≈ 1.5 rather than 0. Looking further, this study primarily added high frequency random noise to base functions that are low frequency. We speculate that adding lower frequency noise, such as alternative base functions, may reduce identification accuracy at PAE levels.

4.6 Experiment 5: Glance and Pause Time

We now turn to the interaction between glance (and pause) time with PAE on the user's ability to perform visual judgments. We hypothesize that the ability to perceive differences in charts is affected by (**H5**) the length of glance time. We use the same task as Experiment 3, but vary the initial chart's glance time. We ran a separate user study that varied the pause time between the inital chart and the two chart options; we found that there was no discernable difference, nor any statistically significant effect, due to the pause time. We thus omit the details due to space constraints.

Participant Tasks: We fixed the baseline function type to linear; varied initial PAE levels $\in \{0.045, 0.09, 0.18\}$; varied $\Delta_y \in \{0.015, 0.06,$

0.24}; and showed the two chart options immediately after the mask is hidden (pause time is 0). We varied glance time $\in$ {50ms, 100ms, 200ms, 2s} to allow for different numbers of saccades: no more than one fixation to enough time to study the visualization. The study was a $3 \times 3 \times 2 \times 4 = 72$ factorial design.

Results, Analysis, and Discussion: There were 48 participants, who took on average 8 minutes and 21 seconds to complete the tasks in the experiment, or approximately 3.5 seconds to complete each comparison task. There were 81% between the ages of 25 and 39, and 56% classifying themselves as intermediate visualization users. We performed a binomial logistic regression, with glance time as the independent variable to predict response accuracy. Glance time ($Z = -4.01, p < 0.001$) was significantly correlated with accuracy and supports **H5**. Figure 17 shows that as the initial PAE level increases (left to right facets), glance time (line) affects judgement accuracy more. As before, increasing $|\Delta_y|$ makes the judgement easier, and increases accuracy. We find that glance time has a significant effect on the accuracy of identifying differences in the initial chart. At short glance times (20ms), users are nearly unable to discern small changes in PAE or when the initial PAE is high. Although longer glance times show higher accuracy, the task is still challenging when the initial PAE is high. We speculate that further increasing the glance time (e.g., to seconds or minutes) may improve accuracy further, but there may be a limit to the accuracy when the initial entropy is high.
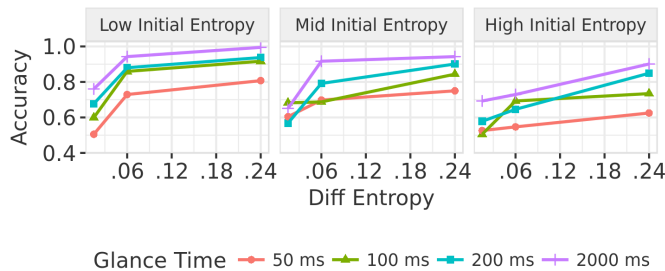

Fig. 17: Acc vs Δ_y for glance time (lines) and initial PAE (facets).

5 DISCUSSION

Our findings on PAE and the perception of line chart complexity have implications for visualization design. They help bridge the gap between research on low-level perception and high-level visualization, and provide a user (as opposed to data) centered measure of visualization complexity. We describe several possible applications for a quantifiable visual complexity measure:

Highlighting Changes: PAE helps quantify changes in visualization complexity, and may be used when it is important to ensure that the user understands changes and comparisons. If the visualization is too complex to perceive certain changes, they may need to be emphasized to the user. Similarly, the systemic bias users experience in differentiating charts of increasing or decreasing complexity implies that visualizations may want to more emphasize changes that increase, rather than decrease, complexity. Further, if PAE is high, users may have difficulty identifying the underlying function in the visualization—if such a detail is important, designers should specifically draw attention to such differences. Finally, the glance time findings can inform the design of displays that rapidly show or change charts. Users need more time to understand more complex charts, and PAE can inform designers to be aware of the user's exposure time depending on chart complexity, and perhaps highlight the key changes.

Large Dataset Visualization: The results can also apply to interactive and approximate visualizations of very large data sets. Quantifying the level of complexity at which a user can identify changes can help designers of approximate visualizations [23, 34, 49, 1] determine when further sampling or other computation will no longer yield perceptible differences. Similarly, designers of interactive visualizations might develop optimizations based on this same phenomena.

Visualization Parameter Selection: PAE might help guide visualization parameter selection (e.g., aspect ratios, layering). For exam-

ple, the choice of horizon graph [33] height and layering may be informed by measuring the resulting chart complexity using PAE. Likewise, PAE could inform aspect ratio selection methods to select a ratio that takes visual complexity into account.

Guided Smoothing: Guided Smoothing: Finally, PAE can be used to indicate when, and to what extent, smoothing or other simplifying methods may be applicable to given chart. For example, PAE could be used with a smoothing method like ASAP to determine when ASAP should be applied to make a chart easier to read, or alternately, PAE could be incorporated into an iterative smoothing method to detect when the chart has simplified enough to be easily readable [57].

6 CONCLUSIONS AND FUTURE WORK

This paper studied the perception of complexity in line chart visualizations. We derive a new measure for visual complexity based on approximate entropy, Pixel Approximate Entropy. We conduct analytical and user experiments to validate its suitability as a complexity measure. In particular, we look at using PAE as a measure of visual complexity; users' ability to perceive minute differences in complexity; and the effect of time on a users' ability to perceive differences in complexity. We performed four sets of experiments and found that PAE correlates closely with noise, that as PAE of a chart increases so too does perceived complexity of the chart; users are better able to perform visual comparison tasks when a chart's PAE is low than when it is high; and users have more difficulty with visual comparison on charts with high PAE when they have less time to view a chart.

There are a number of ways we intend to extend our findings. The first is to design and perform a formal stair case study to better understand whether the JND limits for PAE follow Weber's law. Staircase studies have been used to identify other perceptual features that follow Weber's law, for example, Harrison et al. used a large scale staircase study to measure the JND limits for perception of correlation across 9 visualization types [31]. We intend to perform a similar study to investigate perception of complexity and its relationship with PAE across different one dimensional visualzations. PAE could also be applied to predicting user's perception of correlation in line charts. Rensink et al's studies on scatter plots suggests that users in fact rely on perceived entropy to estimate correlation, whether PAE has similar relationship in 1D visualizations is worth investigating [54].

The second possible extension for our work is to modify PAE to work with other marks and visual encodings. We chose to study line charts in this work because the pixel values rendered in a line chart naturally mapped to a vector of pixel values that could be used to measure PAE. However, line charts are simply one of many possible ways to visually encode data and directly quantifying the visual complexity of other types of visualizations, such as bar charts, scatter plots, and pie charts, remains an open problem.

We also intend to investigate the potential applications of PAE to designing more effective visualizations. As noted in the discussion, PAE has potential applications in designing charts that communicate changes more effectively, efficiently visualizing large data sets, and guiding visual parameter selection and smoothing or other simplifying operations. Testing these applications will require developing visualization systems that integrate PAE and conducting extensive user studies. Ultimately, by developing a quantifiable measure of visual complexity, we hope to contribute to systems that can guide visualization designers in making more readily understandable charts and automatically generate readable charts on their own.

REFERENCES

[1] D. Alabi and E. Wu. Pfunk-h: Approximate query processing using perceptual models. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, page 10. ACM, 2016.

[2] Amazon Mechanical Turk. http://mturk.com. Accessed June 15, 2017.

[3] S. Andrzejak, Ralph G, Kaspar, and C. Rummel. Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients. *Phys. Rev.*, 86, Oct 2012.

[4] J. D. Arias-Londoño and J. I. Godino-Llorente. Entropies from markov models as complexity measures of embedded attractors. *Entropy*, 17(6):3595–3620, 2015.

[5] D. H. Ballard, M. M. Hayhoe, and J. B. Pelz. Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1):66–80, 1995.

[6] G. E. Batista, X. Wang, and E. J. Keogh. A complexity-invariant distance measure for time series. In *Proceedings of the 2011 SIAM international conference on data mining*, pages 699–710. SIAM, 2011.

[7] E. Bertini, A. Tatu, and D. Keim. Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2203–2212, 2011.

[8] I. Biederman, R. J. Mezzanotte, and J. C. Rabinowitz. Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14:143–177, 1982.

[9] A. Biswas, S. Dutta, H.-W. Shen, and J. Woodring. An information-aware framework for exploring multivariate data sets. In *IEEE Transactions on Visualization and Computer Graphics*, volume 19, pages 2683–2692, 2013.

[10] F. Black and M. Scholes. Taxes and the pricing of options. *The Journal of Finance*, 1976.

[11] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *TVCG*, 2013.

[12] R. N. Bracewell and R. N. Bracewell. *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York, 1986.

[13] N. Broers, M. C. Potter, and M. R. Nieuwenstein. Enhanced recognition of memorable pictures in ultra-fast rsvp. *Psychonomic bulletin & review*, 2018.

[14] J. Brown. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958.

[15] M. Chen, M. Feixas, I. Viola, A. Bardera, H. Shen, and M. Sbert. *Information Theory Tools for Visualization*. AK Peters Visualization Series. CRC Press, 2016.

[16] M. Chen and H. Jaenicke. An information-theoretic framework for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1206–1215, 2010.

[17] W. Chen, J. Zhuang, W. Yu, and Z. Wang. Measuring complexity using fuzzyen, apen, and sampen. *Medical Engineering & Physics*, 31(1):61 – 68, 2009.

[18] S. C. Chong and A. Treisman. Representation of statistical properties. *Vision Research*, 43:393–404, 2003.

[19] M. Costa, A. L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.*, 89:068102, Jul 2002.

[20] N. Cowan. The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, 2010.

[21] A. Dasgupta, M. Chen, and R. Kosara. Measuring privacy and utility in privacy-preserving visualization. In *Computer Graphics Forum*, volume 32, pages 35–47. Wiley Online Library, 2013.

[22] A. Dasgupta and R. Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1017–1026, 2010.

[23] B. Ding, S. Huang, S. Chaudhuri, K. Chakrabarti, and C. Wang. Sample+seek: Approximating aggregates with distribution precision guarantee. In *Proceedings of the 2016 International Conference on Management of Data*, pages 679–694. ACM, 2016.

[24] P. Dixon. Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59(4):447 – 456, 2008. Special Issue: Emerging Data Analysis.

[25] S. L. Franconeri. The nature and status of visual resources. *Oxford handbook of cognitive psychology*, 8481:147–162, 2013.

[26] U. Gadiraju, R. Kawase, S. Dietze, and G. Demartini. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1631–1640. ACM, 2015.

[27] E. Gheorghiu, E. Witney, et al. Size and shape-frequency after-effects: same or different mechanism? *Journal of Vision*, 2008.

[28] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. circulation [online]. 101 (23), pp. e215–e220, 2000.

[29] M. R. Greene and A. Oliva. Natural scene categorization from conjunctions of ecological global properties. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Vancouver, July 2006.

[30] M. R. Greene and A. Oliva. Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive psychology*, 58(2):137–176, 2009.

[31] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber's law. *TVCG*, 2014.

[32] C. G. Healey and J. T. Enns. Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE transactions on visualization and computer graphics*, 5(2):145–167, 1999.

[33] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1303–1312. ACM, 2009.

[34] J. M. Hellerstein, P. J. Haas, and H. J. Wang. Online aggregation. In *Acm Sigmod Record*, volume 26, pages 171–182. ACM, 1997.

[35] R. J. Herrnstein and C. Murray. *Bell curve: Intelligence and class structure in American life*. Simon and Schuster, 2010.

[36] K. K. Ho, G. B. Moody, C.-K. Peng, J. E. Mietus, M. G. Larson, D. Levy, and A. L. Goldberger. Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation*, 96(3):842–848, 1997.

[37] H. Karloff and K. E. Shirley. Maximum entropy summary trees. In *Computer Graphics Forum (Proc. EuroVis)*, volume 32, pages 71–80, 2013.

[38] R. Kimchi. Primacy of wholistic processing and global/local paradigm: a critical review. *Psychological bulletin*, 112(1):24, 1992.

[39] S. M. Kosslyn. Understanding charts and graphs. *Applied cognitive psychology*, 3(3):185–225, 1989.

[40] H. R. Lipford, F. Stukes, W. Dou, M. E. Hawkins, and R. Chang. Helping users recall their reasoning process. In *VAST*, 2010.

[41] J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics*, 13(6), 2007.

[42] A. Olivia, M. L. Mack, M. Shrestha, and A. Peeper. Identifying the perceptual dimensions of visual complexity of scenes. In *Proceedings of the Cognitive Science Society*, volume 26, 2004.

[43] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3659–3669. ACM, 2016.

[44] W. Peng, M. O. Ward, and E. A. Rundensteiner. Clutter reduction in multidimensional data visualization using dimension reordering. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 89–96, 2004.

[45] S. Pincus. Approximate entropy (apen) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 1995.

[46] S. M. Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 1991.

[47] M. C. Potter. Meaning in visual search. *Science*, 1975.

[48] M. C. Potter. Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, 1976.

[49] M. Procopio, C. Scheidegger, E. Wu, and R. Chang. Load-n-go: Fast approximate join visualizations that improve over time. *DSIA*, 2017.

[50] K. Rayner, T. J. Smith, G. L. Malcolm, and J. M. Henderson. Eye movements and visual encoding during scene perception. *Psychological science*, 20(1):6–10, 2009.

[51] L. W. Renninger and J. Malik. When is scene identification just texture recognition? *Vision research*, 2004.

[52] R. Rensink. The rapid perception of correlation in scatterplots. In *Journal of Vision*, volume 11, page 1085, 2011.

[53] R. Rensink. An entropy theory of correlation perception. In *Journal of Vision*, volume 16, pages 811–811. The Association for Research in Vision and Ophthalmology, 2016.

[54] R. A. Rensink. The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*, 24(3):776–797, Jun 2017.

[55] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. In G. Melancon, T. Munzner, and D. Weiskopf, editors, *Computer Graphics Forum*, volume 29, pages 1203–1210, 2010.

[56] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology - Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000.

[57] K. Rong and P. Bailis. Asap: prioritizing attention via time series smoothing. *Proceedings of the VLDB Endowment*, 10(11):1358–1369, 2017.

[58] R. Rosenholtz, Y. Li, and L. Nakano. Measuring visual clutter. *Journal of vision*, 7(2):17–17, 2007.

[59] J. Schneidewind, M. Sips, and D. A. Keim. Pixnostics: Towards measuring the value of visualization. In *Visual Analytics Science And Technology, 2006 IEEE Symposium On*, pages 199–206. IEEE, 2006.

[60] P. G. Schyns and A. Oliva. From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, 1994.

[61] J. M. Smith. *Introduction to chemical engineering thermodynamics*. PhD thesis, Rensselaer Polytechnic Institute, 1975.

[62] D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri. Four types of ensemble coding in data visualizations. *Journal of vision*, 16(5):11–11, 2016.

[63] N. Tavakoli. Short communication: Entropy and image compression. *Journal of Visual Communication and Image Representation*, 4(3):271–278, September 1993.

[64] C. Wang and H.-W. Shen. Information theory in scientific visualization. *Entropy*, 13(1):254–273, 2011.

[65] C. Wang, H. Yu, and K.-L. Ma. Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1547–1554, 2008.

[66] Y. Wang, Z. Wang, L. Zhu, J. Zhang, C. W. Fu, Z. Cheng, C. Tu, and B. Chen. Is there a robust technique for selecting aspect ratios in line charts? *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2017.

[67] H. Wickham, D. Cook, H. Hofmann, and A. Buja. Graphical inference for infovis. In *IEEE Transactions on Visualization and Computer Graphics*, volume 16, pages 973–979, November/December 2010.

[68] L. Wilkinson, A. Anand, and R. L. Grossman. Graph-theoretic scagnostics. In *INFOVIS*, volume 5, page 21, 2005.

[69] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou. The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of Biomedical Engineering*, 41(2):349–365, Feb 2013.