CS 8803-MDS
Human-in-the-loop Data Analytics

Lecture 24

11/16/22

Today's class

Vega-lite: A grammar of interactive graphics

Reviewer: Qiandong

Archaeologist: Haotian

Expressive Time Series Querying with Hand-Drawn Scale-Free Sketches

Authors: Harshal, Cangdi

Reviewer: Haotian

Archaeologist: Akshay

Practioner: Siddhi

Expressive Time Series Querying with Hand-Drawn Scale-Free Sketches

Authors

Harshal Gajjar, Cangdi Li

8803-MDS Fall 22

Prof. Kexin Rong

Background

- **Our ability to describe complex objects with hand-drawn sketches and easily recognize them** predates our ability to do so with language
- **Many search interfaces have capitalized on this ability**, providing users with intuitive sketching interfaces for search queries
- **Generally, such “querying by sketching systems” assume** that **sketched objects resemble their real-world counterparts**

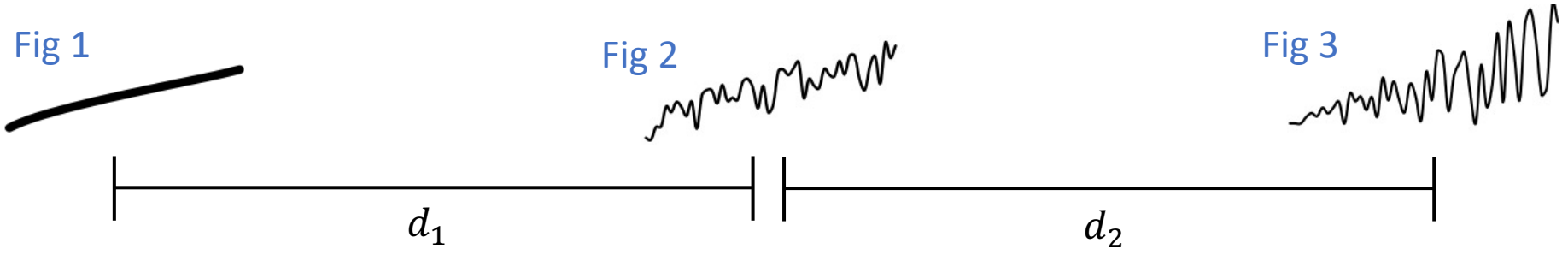
Problem

https://github.com/googlecreativelab/quickdraw-dataset

- Most humans are **not** faithful artists
- Most people **don’t have access to fancy drawing equipment** like stylus

Existing systems & Perceived Distance

- The accuracy of many existing time series matching algorithms is often evaluated by **how well they cluster similar time series patterns** extracted from the data — and **not how well they cluster data patterns with hand-drawn sketches**
- Consequently, "good" matching algorithms may fail to produce good similarity rankings when "goodness" is assessed by humans



Euclidean distance metric: $d_1 > d_2$

Goodness/Perceived distance metric: $d_1 < d_2$

Concrete Motivating Example

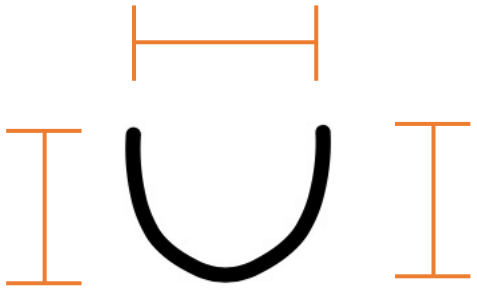
- **Economist looking for transient historical periods of recession** marked by a sharp decrease in gross domestic product (GDP) and then a rise, as well as an increase in unemployment and then a fall.
- Immediate thought about the curves give the following mental representations:

GDP

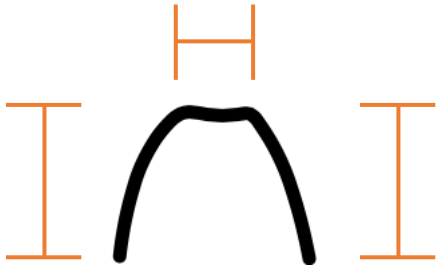
Unemployment

Concrete Motivating Example

- With existing systems, however, the economist must consider **questions which are not straightforward to answer**: how long was the recession? How big was the fall in gross domestic product? How high was unemployment? etc.



GDP

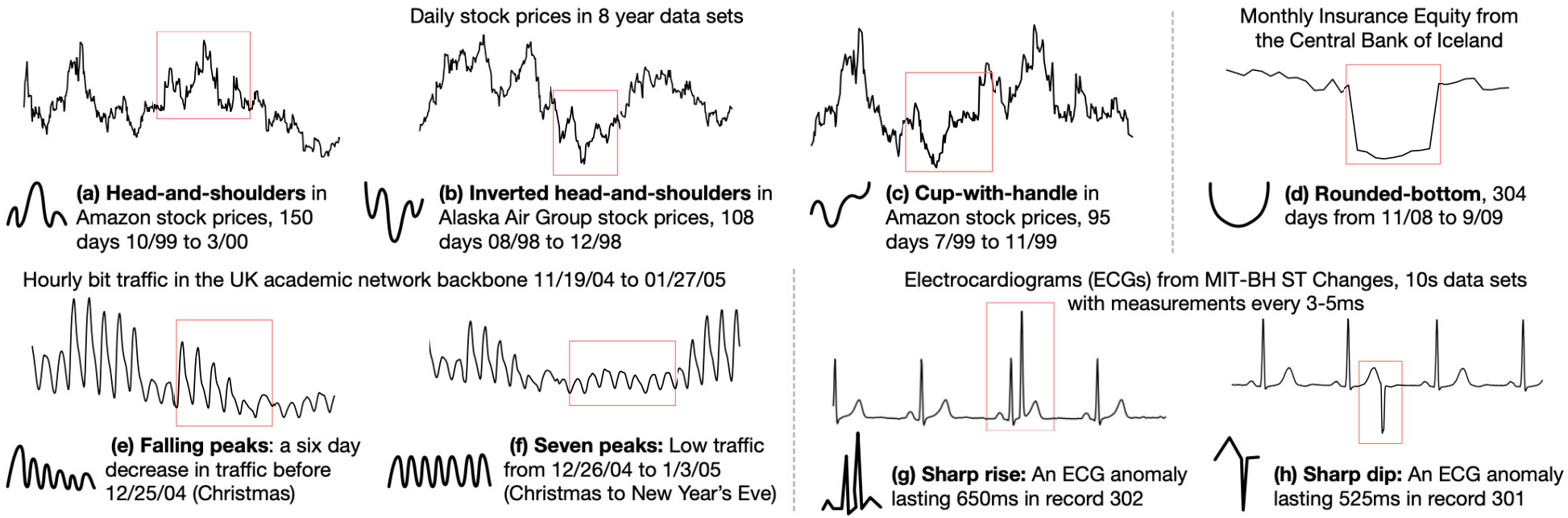


Unemployment

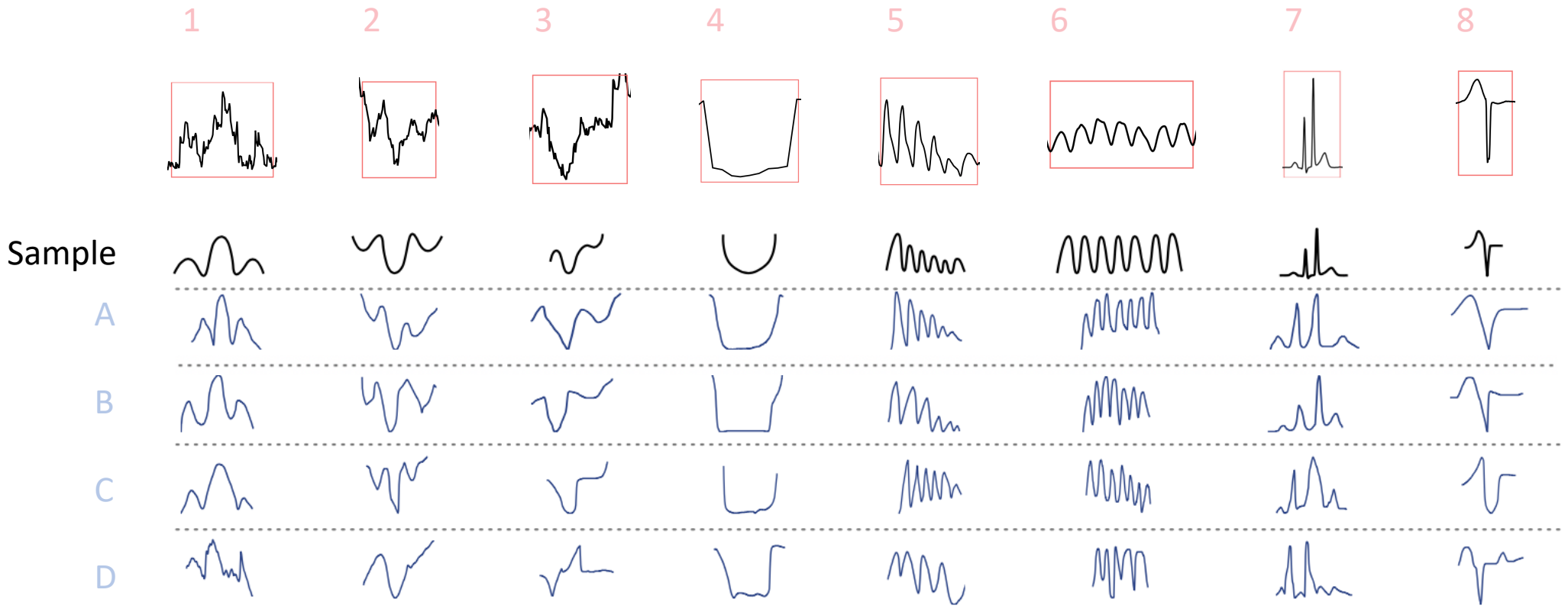
Introducing Qetch

Wait. How do people sketch?

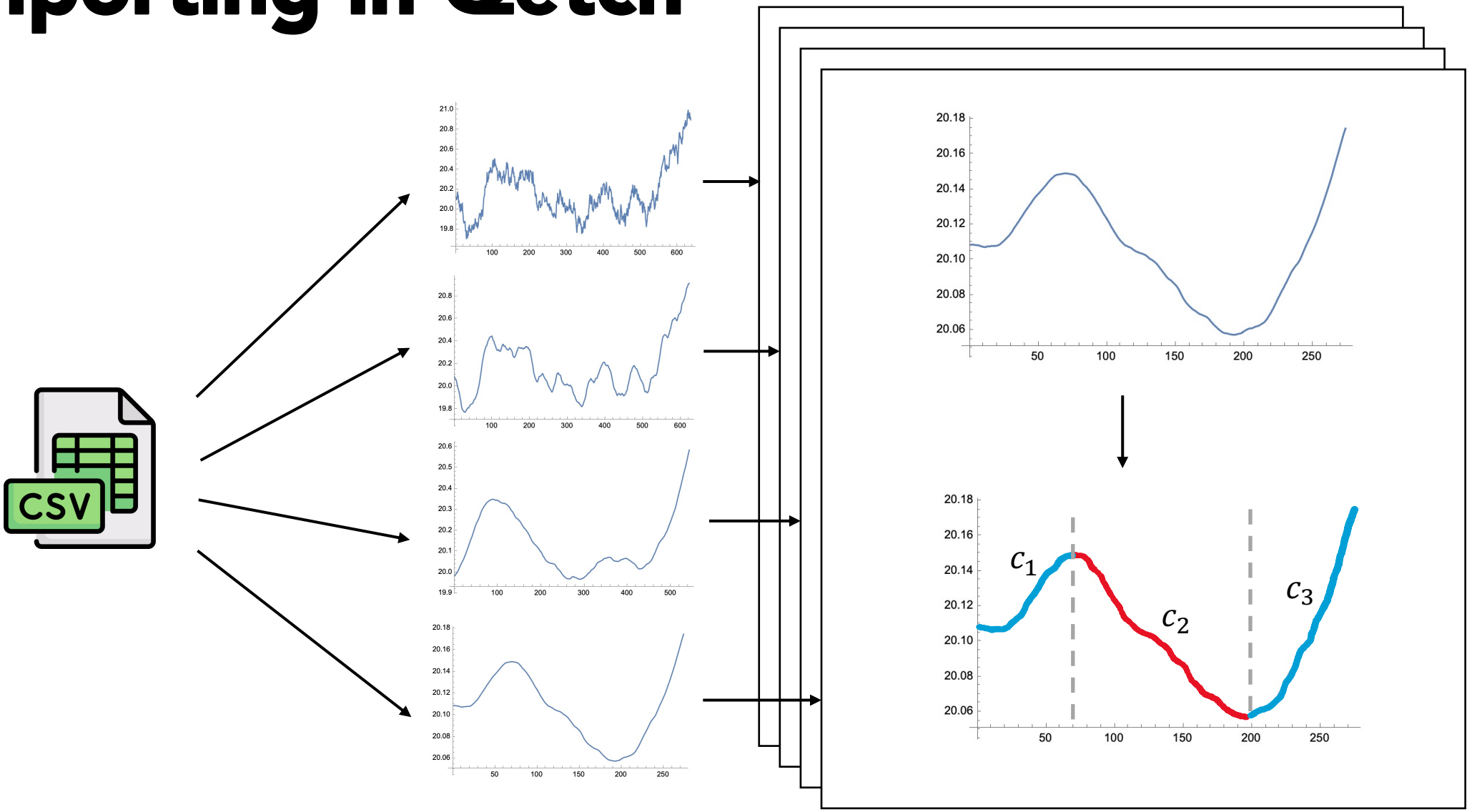
- To study how people represent various curve trends as sketches we **used AMT** to get a total of **1200 sketches from 150 people** which was later **cleaned by 41 more people** to get a total of **930 sketches**



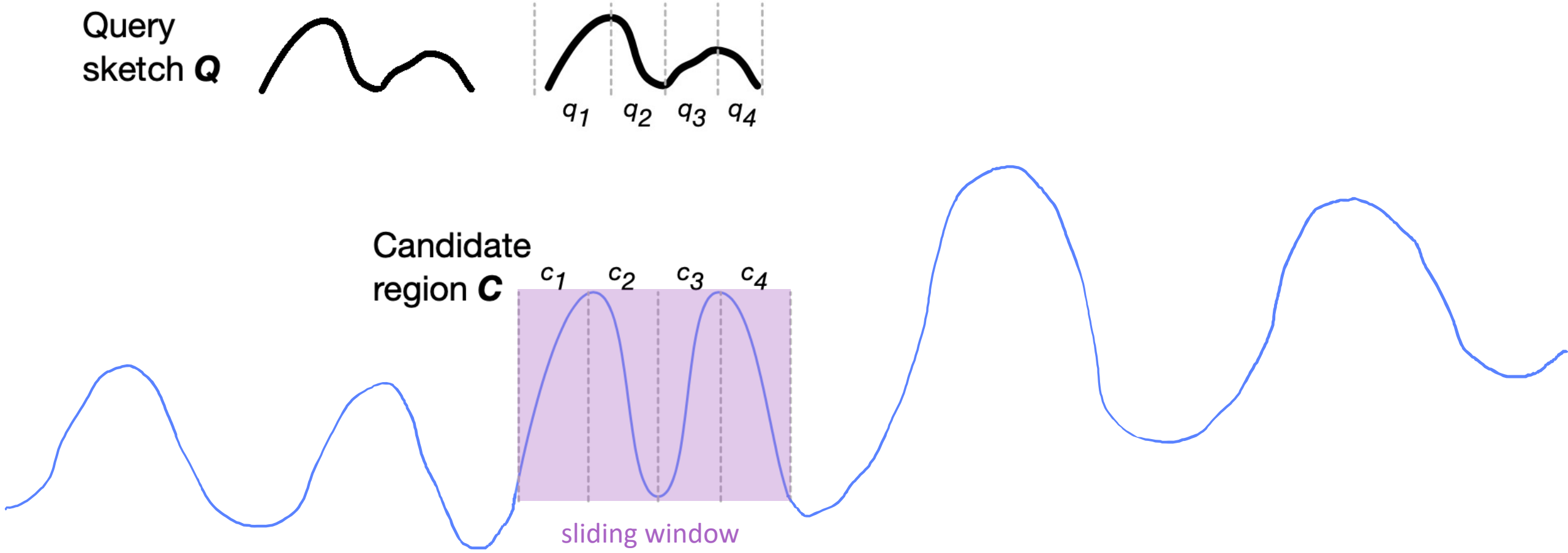
Observations from crowd study



Importing in Qetch

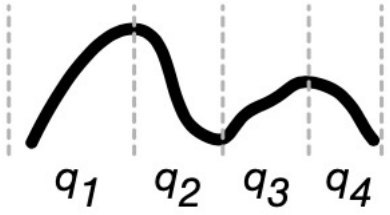


Qetch's Matching Algorithm

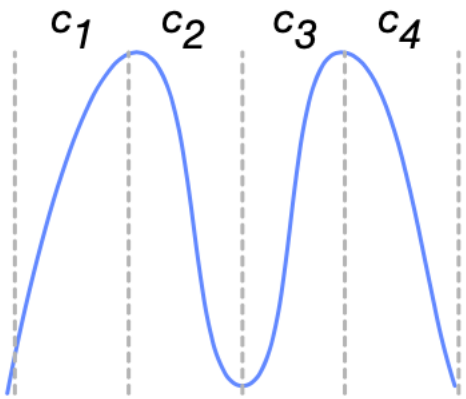


Qetch's Matching Algorithm

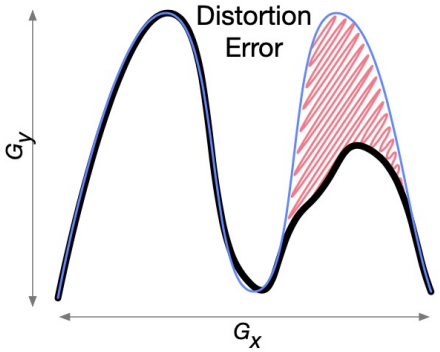
Query sketch **Q**



Candidate region **C**



Error 1: Distortion Error



(a) Global non-uniform scaling

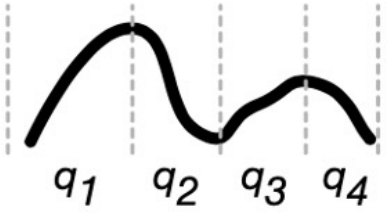
$$R_y(q_i, c_i) = \frac{\text{height}(c_i)}{G_y * \text{height}(q_i)}$$

$$= \frac{\text{height}(c_i)}{\text{height of } q_i \text{ after global scaling}}$$

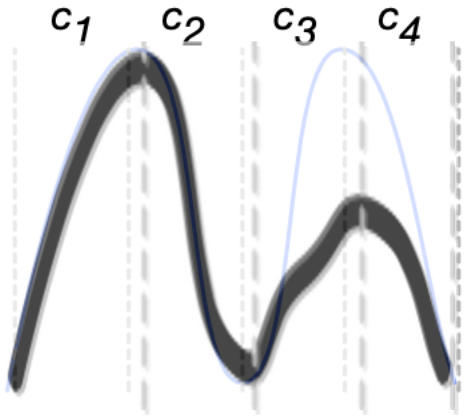
$$\text{LDE}(q_i, c_i) = \log(R_x(q_i, c_i))^2 + \log(R_y(q_i, c_i))^2$$

Qetch's Matching Algorithm

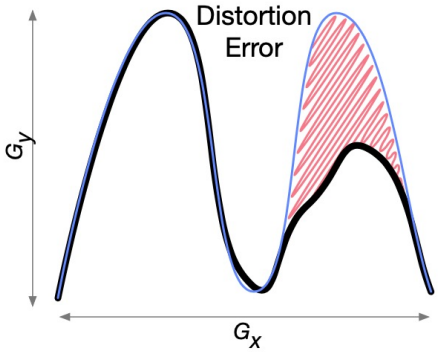
Query sketch **Q**



Candidate region **C**



Error 1: Distortion Error



(a) Global non-uniform scaling

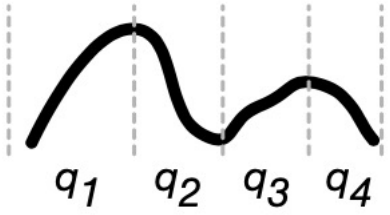
$$R_y(q_i, c_i) = \frac{\text{height}(c_i)}{G_y * \text{height}(q_i)}$$

$$= \frac{\text{height}(c_i)}{\text{height of } q_i \text{ after global scaling}}$$

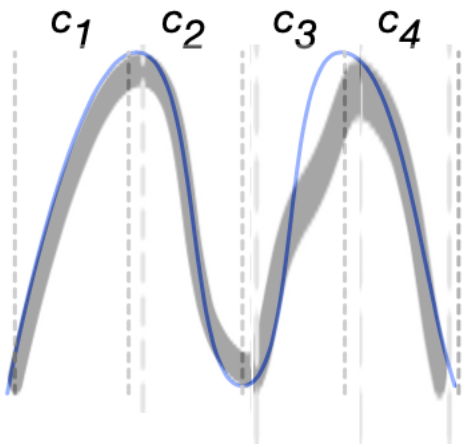
$$\text{LDE}(q_i, c_i) = \log(R_x(q_i, c_i))^2 + \log(R_y(q_i, c_i))^2$$

Qetch's Matching Algorithm

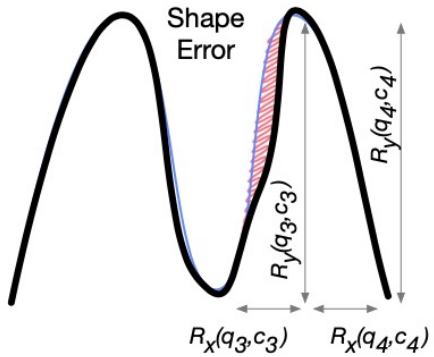
Query sketch **Q**



Candidate region **C**



Error 2: Shape error

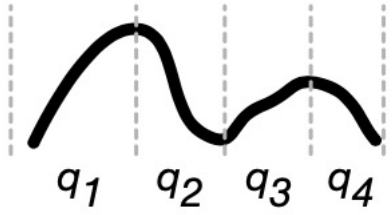


(b) Local segment rescaling

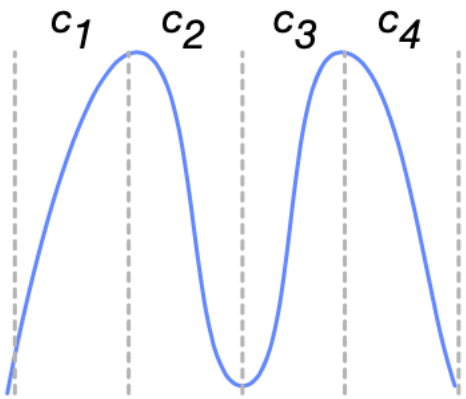
$$\mathrm{SE}(q_i, c_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \left| \frac{G_y * R_y(q_i, c_i) * q_i[j].y - c_i[j].y}{\mathrm{height}(C)} \right|$$

Qetch's Matching Algorithm

Query sketch **Q**



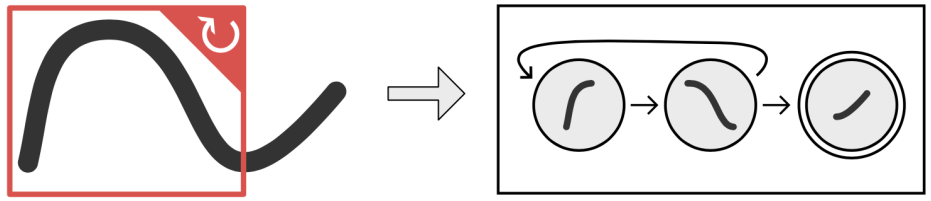
Candidate region **C**



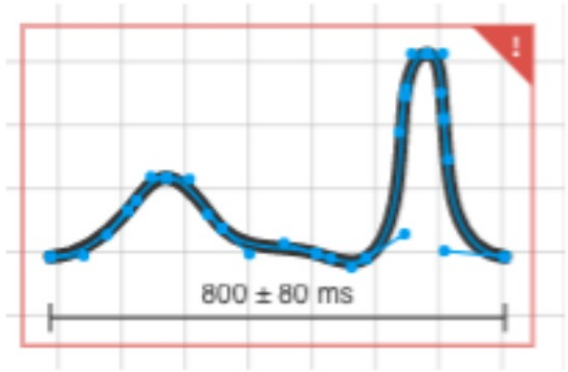
Total error = Error 1 + Error 2

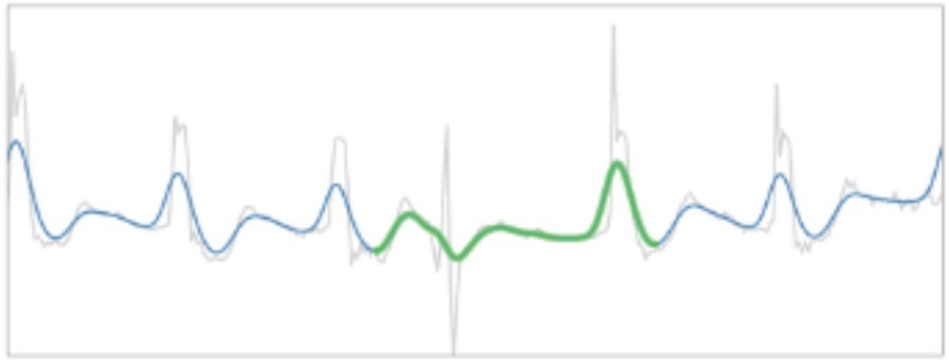
Cherry on the cake (waiting for reviewers)

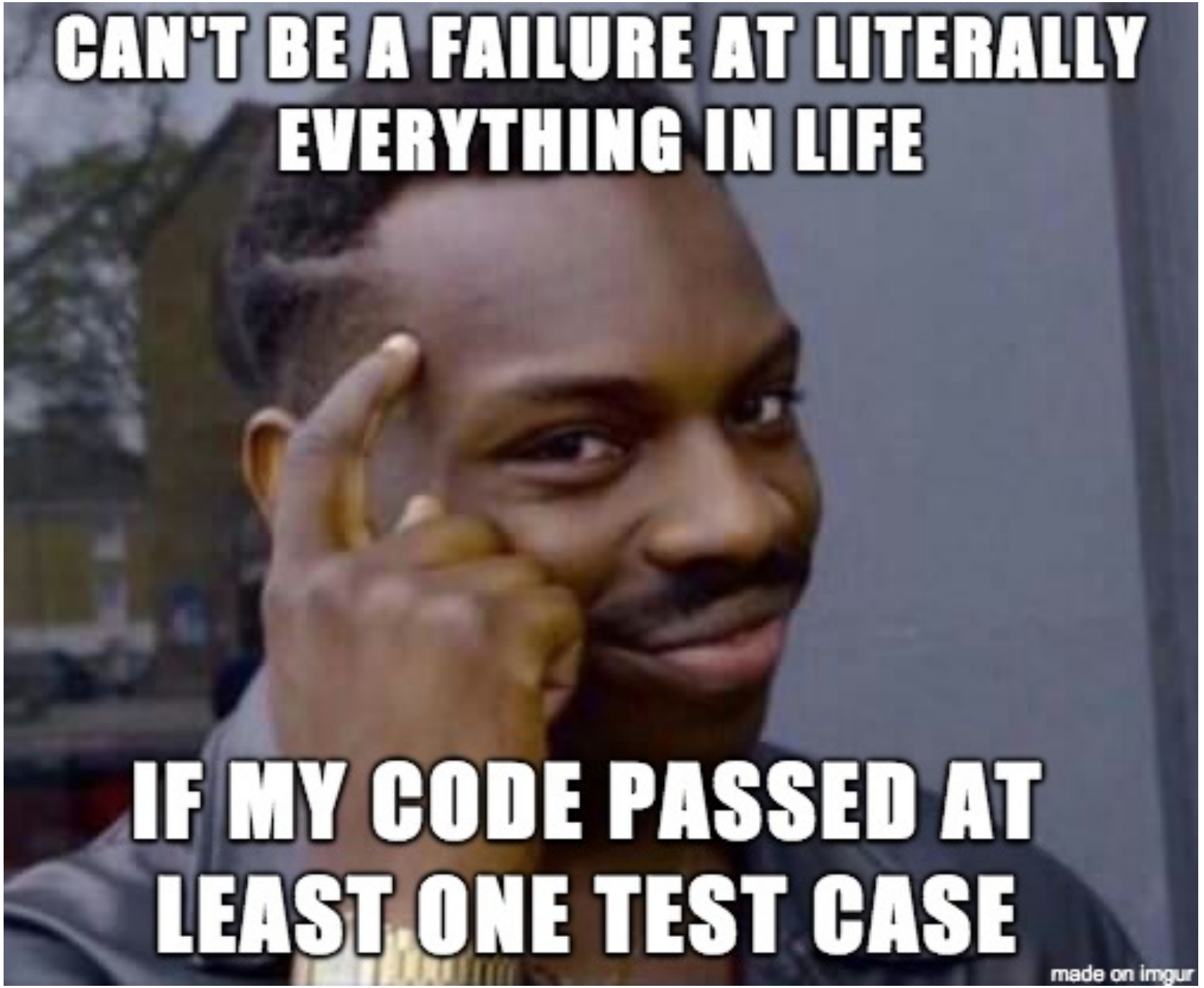
- Support for Regex in sketches



Example






CAN'T BE A FAILURE AT LITERALLY EVERYTHING IN LIFE
IF MY CODE PASSED AT LEAST ONE TEST CASE
made on imgur

Evaluation

- We conducted a within-subjects comparative user study of Qetch's novel time series querying features:
  - (i) regular expressions for querying repeated patterns and anomaly detection VS no regular expressions
  - (ii) relative positioning of sketches for querying across multiple data sets VS specifying order constraints over sketches.
  - Then use query completion times to determine whether Qetch's features improved querying.
- We also evaluate smoothing preference and Qetch interface.

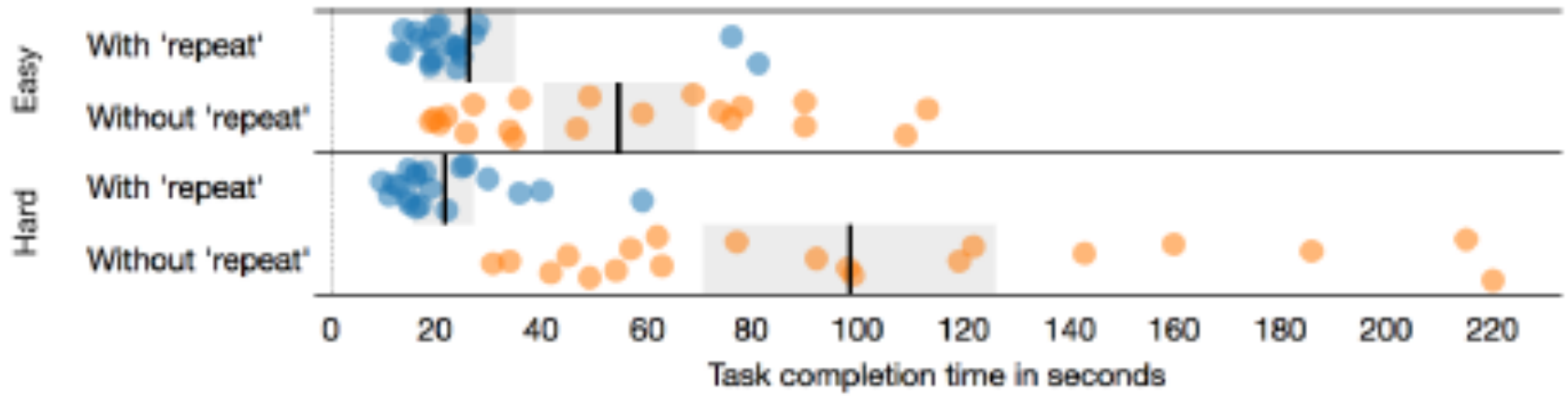
Evaluation Setup

- 20 university students with some training on how to use the tool.
- A series of querying task within 600 seconds on synthesized datasets with/without a particular Qetch feature.
- Synthesized datasets are designed to be easily verified and new to all participants.
- Only mouse inputs are allowed to ensure results are generalized.

- The comparative user study are split into 3 tasks
  - Task set A: Search for repeated patterns.
  - Task set B: Detect an anomaly
  - Task set C: Query across multiple data sets.

Evaluation: The repeat operator

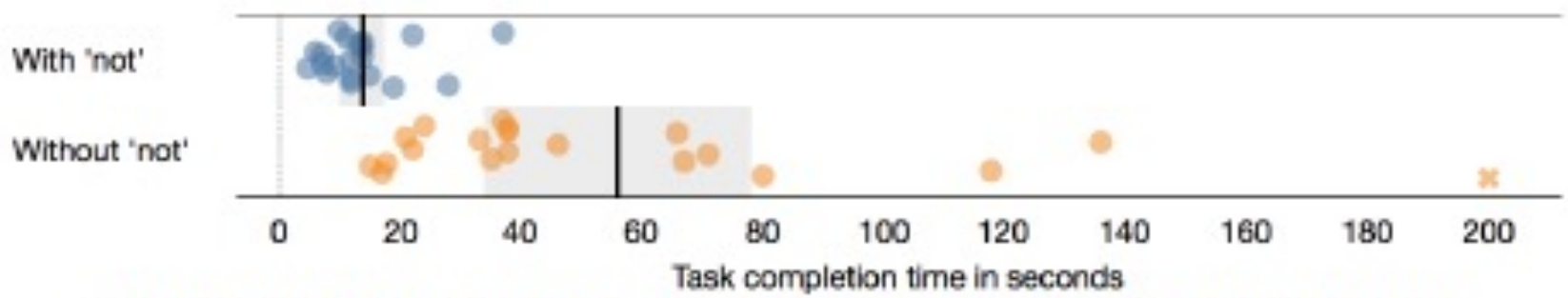
- A two-way repeated measures ANOVA with task difficulty and operator-use as independent factors.
  - For both easy and hard tasks, there was a significant difference in completion times between using and not using the repeat operator
  - Without the repeat operator, there was a significant difference in task completion times between easy and hard tasks



(a) Task completion times when searching for repeated patterns with and without the repeat regex operator

Evaluation: The not operator

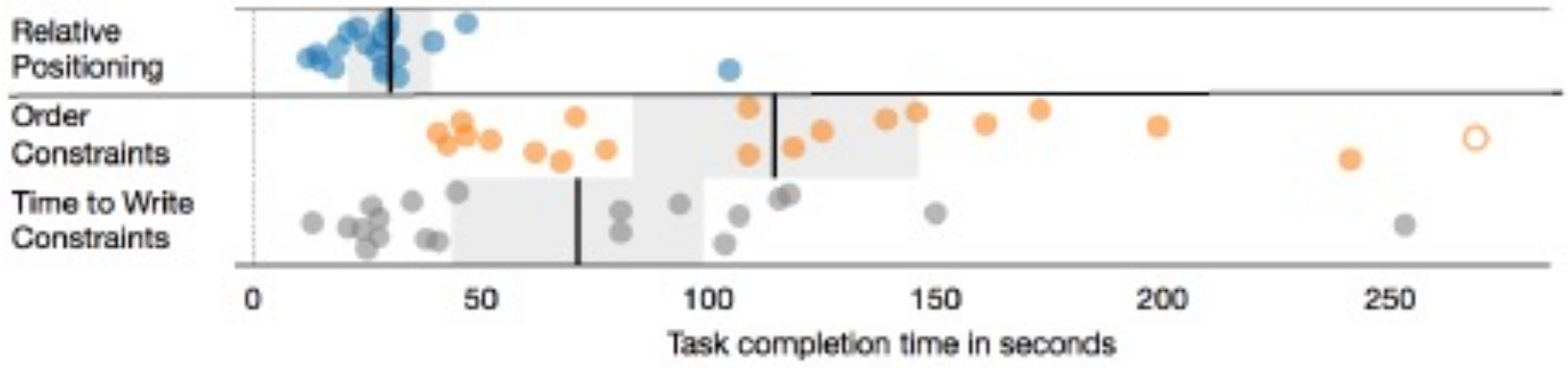
- One-way repeated measures ANOVA with operator-use as an independent factor.
  - We observe a statistically significant effect of using the not operator on task completion times



(b) Task completion times when searching for an anomaly with and without the not regex operator

Evaluation: Relative positioning

- The effectiveness of relative positioning of sketches as a technique for querying multiple data sets **VS** the explicit specification of order constraints with keywords on sketches:
  - One-way repeated measures ANOVA with technique used as an independent factor.
- We found a significant effect of technique used (relative positioning).
- In the post-study questionnaire, 18 users of the 20 users preferred relative positioning over the specification of order constraints, for specifying queries over multiple datasets.



(c) Task completion times when searching for patterns across multiple data sets with relative positioning of sketches vs. specifying order constraints with keywords

Evaluation: Smoothing Preferences

- For each of the 8 queries, we presented a query sketch, and asked users to slide three sliders to indicate the minimum, preferred and maximum degree of smoothing of the dataset. Observation:
  - For most queries, users want a minimum non-zero degree of smoothing to exist.
  - Users set a maximum degree of smoothing well below 50% of the highest possible smoothing
- Conclusion: Qetch's smoothing choices are within a 95% confidence interval of the mean preferred smoothing degree for six of the eight queries.

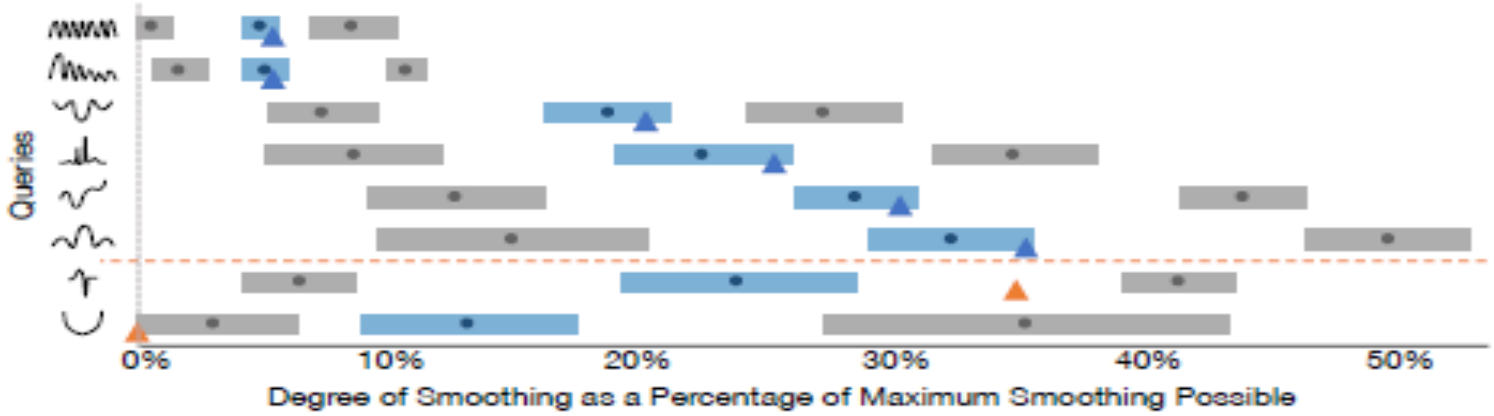


Figure 7. For each query, we present 95% CI bars for the minimum, preferred, and maximum mean smoothing degrees as selected by users. The triangle glyph marks Qetch's degree of smoothing.

Evaluation: Qualitative Evaluation of the Qetch Interface from questionnaire

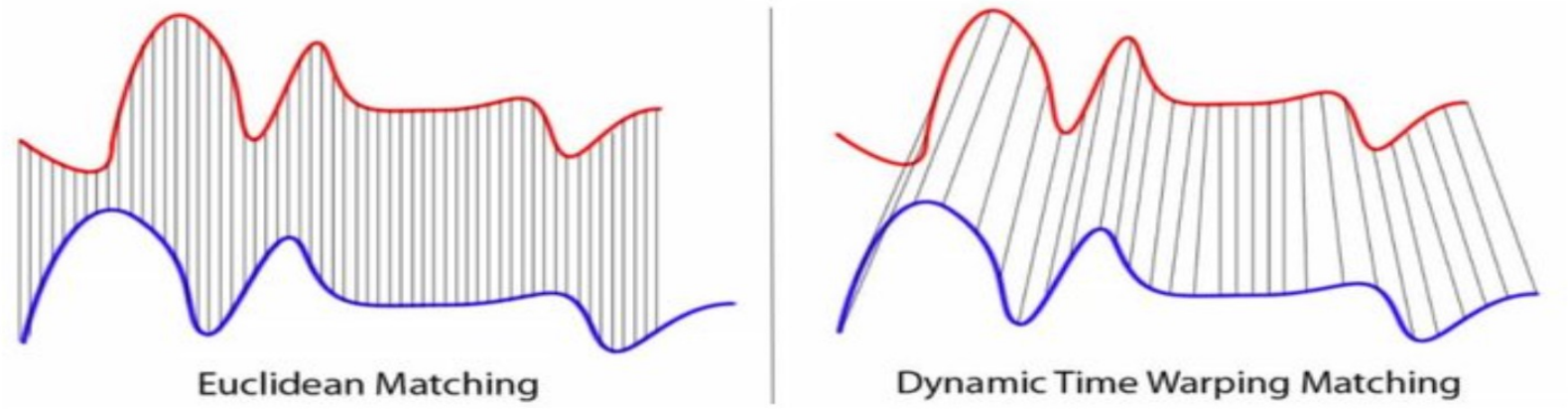
- Overall, participants found it easy to use.
- Query-by-sketching is an intuitive and effective paradigm.
- It's interactive and responsive enough to freely correct sketching errors.
- In general, users found the table visualization and the ordering of results by different attributes effective.

Evaluate Qetch's Matching Performance on Search Tasks

- We evaluate Qetch on two types of search tasks:
- Targeted search: users search the data set for a specific region
- Exploratory search: users search for several regions in the data set that match their sketch.
- We compare the results with ED and DTW

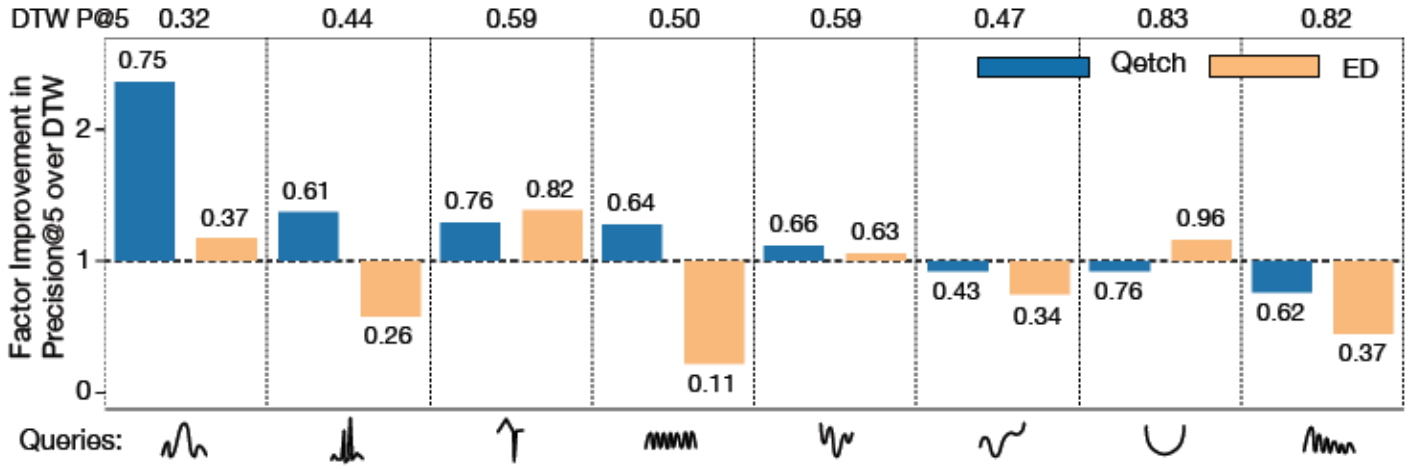
Measuring the similarity between two time-series sequences data

- **ED: Euclidean Distance Matching**
  The square root of the sum of the squared length of the vertical hatch lines
- **DTW: Dynamic Time Warping Matching**
- Rakthanmanon et al. state that “after an exhaustive literature search of more than 800 papers, we are not aware of any distance measure that has been shown to outperform DTW by a statistically significant amount on reproducible experiments”



Compare targeted search

- Compare the precision of Qetch to both DTW and ED.
- Use the previous crowd study, We evaluated how often each algorithm placed this reference region in its top 5 results when the sketch was used to query the time series.
- We find that Qetch outperforms DTW and ED with a few exceptions.

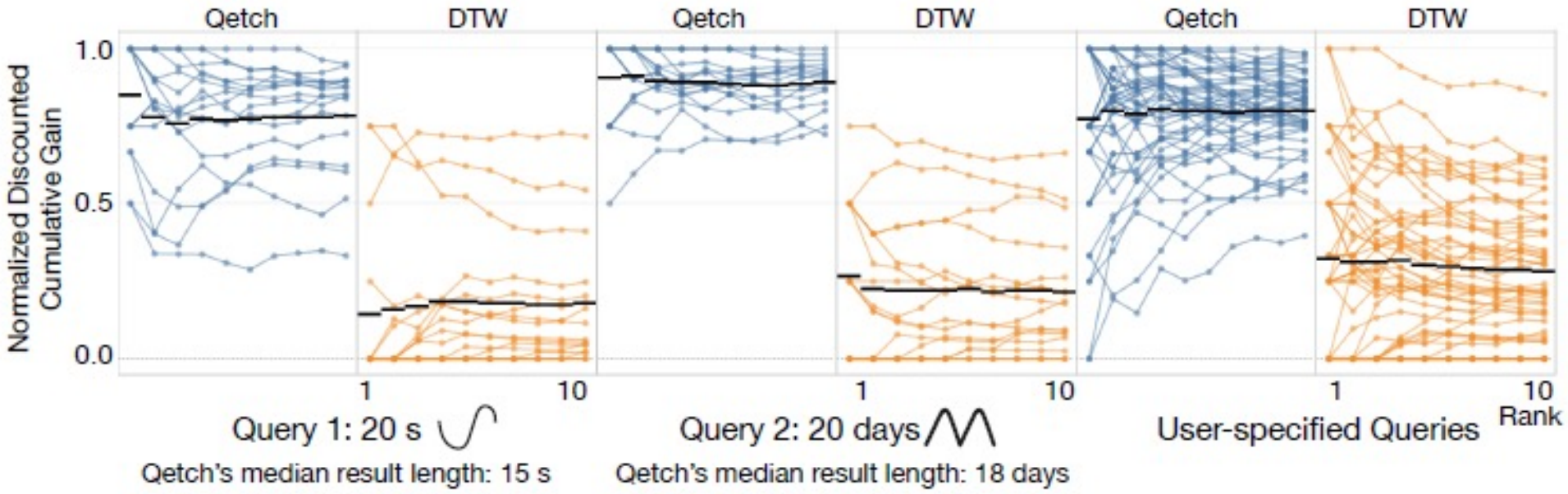


Compare exploratory search

- We recruited another 16 university students, applied the same training procedure.
- We ask the users to perform two query tasks:
  - A pattern lasting roughly 20 seconds on the MIT-BIH heart rate series 1
  - A pattern lasting roughly 20 days on the mean daily temperature in Saugeen river.
  - We then asked users to freely sketch three queries of their choice on 3 other datasets.
- For each user query, we presented the top-10 results of DTW(add description) (no smoothing) and Qetch.
- The users rated the relevance of each result from bad (1) to perfect (5).
- We evaluated the effectiveness of DTW and Qetch with the popular normalized discounted cumulative gain (NDCG) measure (usefulness)

Compare exploratory search

- Qetch noticeably outperforms DTW across all queries.
- Users attributed Qetch's superior performance to two aspects:
  - Qetch's smoothing choices and its presentation of smoothed results.
  - Qetch gives equal importance to each perceptual feature within a sketch, where Qetch is better at finding smaller features and DTW tends to look at the bigger picture and ignore minor peaks.



Limitation Discussion

- **Cons**
  - Qetch does not outperform DTW in standard time-series benchmarking tests
  - Poor choice of distance measure: measuring the distance between one time-series region and another.
- **Pros**
  - Qetch is suitable for matching a rough, hand-drawn, scale-less sketch to a time series.
  - For the small/medium scale time series, Qetch provides interactive performance with high precision. (For large-scale, it performs similar to DTW and ED)

Conclusion

- Proposed Qetch, a query-by-sketch tool for time series data.
- Designed a noval matching algorithm to tolerate the human sketch errors.
- Evaluation demonstrate that Qetch outperforms standard algorithms on targeted and exploratory search tasks.
- Publicly release our crowd-sourced data set of sketches and source code.

Expressive Time Series Querying with Hand-Drawn Scale-Free Sketches: Archaeologist

Akshay Iyer

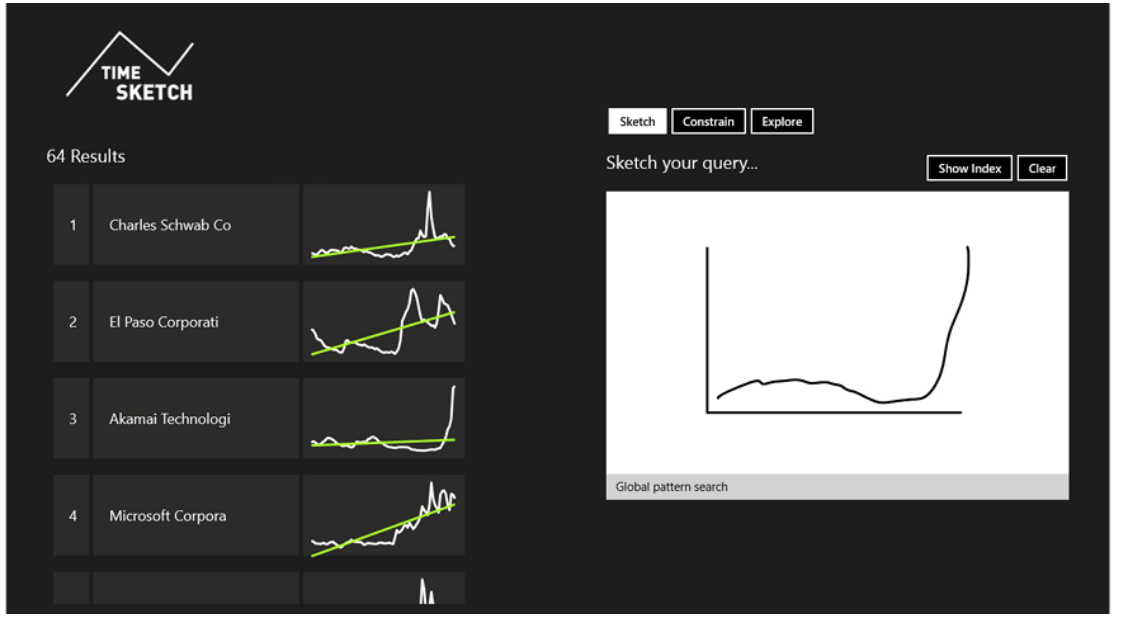
Georgia Tech

Paper Summary

- Qetch: Lets users freely sketch patterns which translate into queries on the dataset
- Most existing interfaces are restrictive from a user's point of view. These restrictions usually fall into overlays, shape restrictions, and pre-sketching constraints
- Key contribution is the novel matching algorithm.
  - Segments and smooths time series data
  - Abstracts scale, applies global rescaling factor to determine closest segment to sketch
  - Custom distance metric accounts for human errors, summing local distortion error and shape error
- Not suitable for a number of time-series tasks such as determining distances between regions and finding recurring patterns
- Generally good user feedback for small/medium time series, but challenges with scaling

Previous Paper: Evaluating Subjective Accuracy in Time Series Pattern-Matching Using Human-Annotated Rankings

- A similarity search algorithm that a human would output is inherently subjective. Authors argue that current datasets do not support this subjectivity
- Gist of the paper is to create human-annotated rankings, determine which similarity-based ranking best align with human intuition.
- Relatively brute-force approach:
  - Find all the possible subsequences that match the length of some query pattern
  - Look at every pair and ask users to determine which pair resembles query pattern more
  - Identify the level to which participants agree and sort by similarity



Georgia Tech

Previous Paper: Evaluating Subjective Accuracy in Time Series Pattern-Matching Using Human-Annotated Rankings

- Comparing two human-annotated and machine-based rankings:
- Experimental setup involved 8 hand drawn query patterns. 4 datasets used and accuracy of 3 different similarity measures w/4 sampling strategies was gauged
- Between DTW, Euclidean Distance and SpADe, DTW had the most “human-like” results. (Rankings were closest to human-annotated ones)
- Also found that explored sampling methods(Geometric Linear, PLS, PIP) did not significantly affect accuracy of algorithms
- Acknowledge that other sampling methods might achieve better results. Authors explore idea of defining a sketching-language that supports common types of queries

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

where d_i is the difference of the ith ranking in R_S and R_H respectively and n is the total number of ranks. A Spearman coefficient of 1 or -1 indicates that the two ranking correlation perfectly (positively or negatively) which means the produced ranking is identical to the human created one. We use this correlation coefficient as a measure of how well two

Georgia Tech

Future Paper: Is this Real? Generating Synthetic Data that Looks Real

- Authors developed a tool called Synner which helps users generate realistic synthetic data
- Motivation for this problem includes better education initiatives as well as stronger software presentations
- 4 main design principles:
  - 1.) Visual Lifting & Declarative Specification—Generating familiar spreadsheet visualizations and telling users what features they can continually explore
  - 2.) Example-driven interaction—Recommendation for generating further data and suggests dependencies between different domains
  - 3.) Communicating randomness—Let users know about misconceptions
  - 4.) Separation of concerns—Distinguish between errors that do not affect data generation
    - Give users the option to add missing/erroneous values in the dataset after generation
- User feedback was used to determine how “real” a dataset is

Future Paper: Is this Real? Generating Synthetic Data that Looks Real

- Baseline was Mockaroo, another data generation tool. Based on user study with college students, users spent less time specifying data generation tasks in Synner than with Mockaroo.
- Synner satisfied more checks of realism based on users's responses.
- Limitation is that its interactivity and visualization does not scale to modern big data workload
- Same authors who wrote this paper wrote the Expressive Time Series Querying paper. Cite from their own work how users tend to misjudge time-series patterns, overemphasizing certain features like tapering

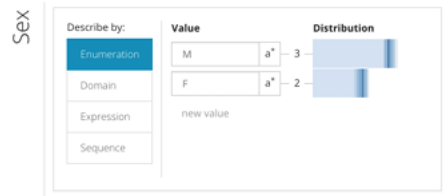


Figure 3. Imani specifies the ratio of males to females in her dataset with the help of an enumeration.

Figure 4. Synner suggests different distributions from a few examples entered by Imani in the Years of Education column

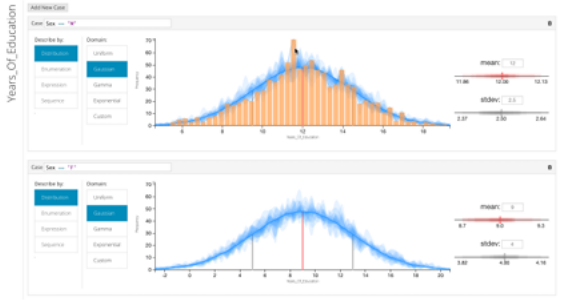


Figure 5. Imani specifies the relationship between years of education and sex using dependencies by cases. Synner illustrates the many plausible data generation outcomes using the Layers visualization. Hovering over a layer reveals the underlying histogram.

the domain database such as 'city', or 'continent'. There is no which allows 'country' a. To enable these rich

5 / 13

Georgia Tech

Thank you!

Georgia Tech

Expressive Time Series Querying with Hand-Drawn Scale-Free Sketches

•••

Reviewer: Haotian Sun

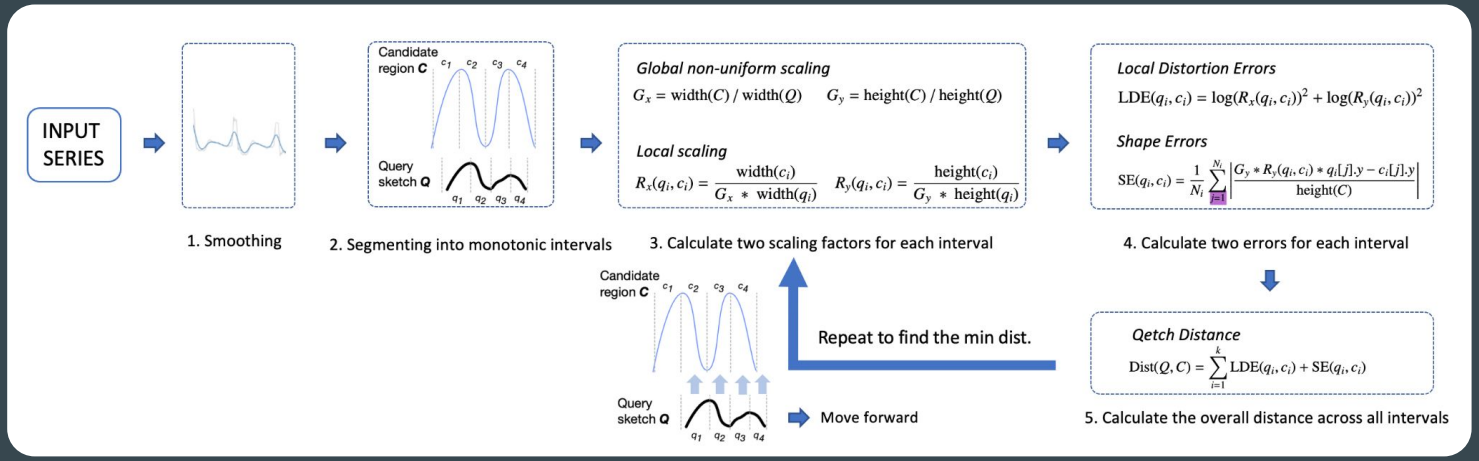
In Summary

- Qetch, a query-by-sketching framework that allows users to sketch and search patterns on a scale-less canvas without specifying query length or amplitude;

- Outperforms the conventional matching algorithms and supports regular expressions for querying.

In Summary

- The sketches are first smoothed using Bézier curves
- Segmented into several monotonic intervals
- The distances for all matchings are calculated by combining local distortion and shape errors



Strong Points

- User-friendliness: interactive frontend UI to adjust smoothing level, perform query-by-sketching, and regular expression query
- Sketching freedom: allow users to freely sketch on canvas without shape restrictions
- Fewer requirement: get rid of explicitly defining query features like time length or amplitude
- More functionalities: support three regular expression operators

Weak Points

- Limited #regex supported: Qetch only supports three regular expression operators (repeat, repeat for n times, and not.)
- Limited performance in standard time-series benchmarking (compared w/ DTW)
    - Maybe promising to propose a holistic method combining Qetch & DTW for different tasks
- Sensitivity to smoothing levels: query result is very sensitive to the smoothing level of the time series data[1], since varying smoothing level may affect the monotonic interval segmentations.

[1] C. Fan, K. Matković and H. Hauser, "Sketch-Based Fast and Accurate Querying of Time Series Using Parameter-Sharing LSTM Networks," in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 12, pp. 4495-4506, 1 Dec. 2021,

Overall evaluation

- Novelty in designing the holistic querying system
- Proposed a matching algorithm with enhanced performance
- Good awareness of what people working with time series really need

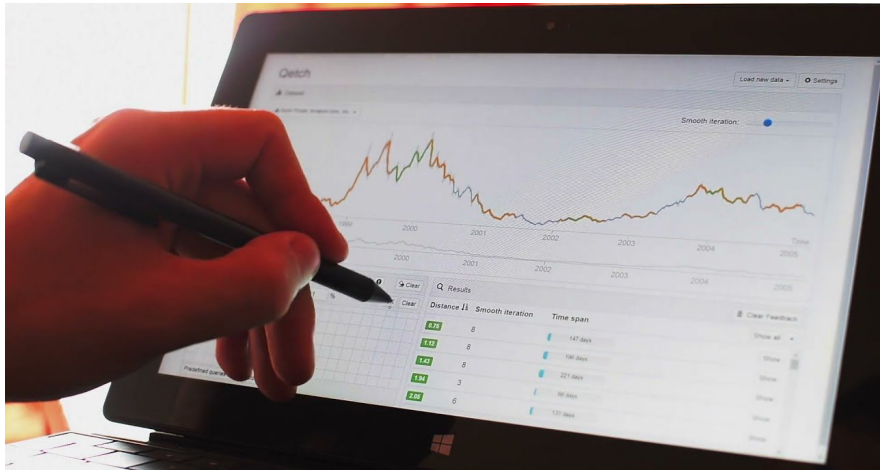
Qetch: Expressive Time Series Querying with Hand-Drawn Scale-Free Sketches

Practitioner Pitch

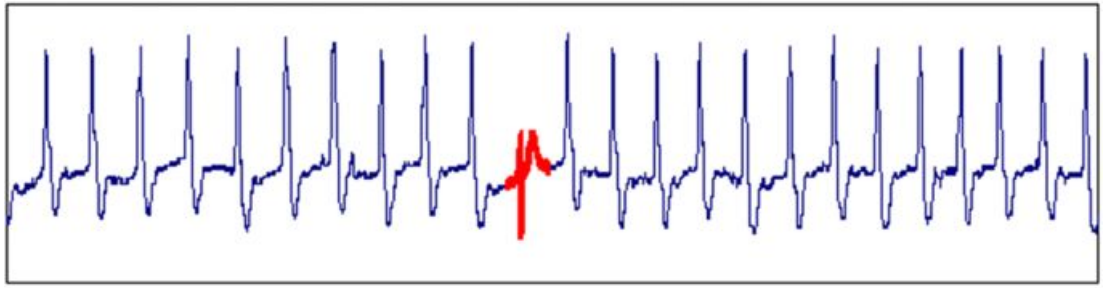
Siddhi Pandare

Features

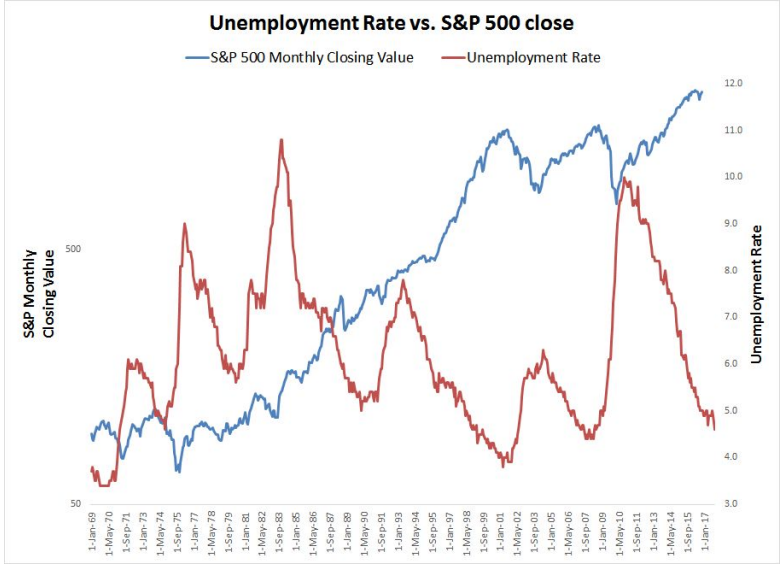
- Qetch allows users to query time series data using sketch patterns on a scale free canvas.
- Analysts can annotate their sketch with regular expressions.
- Query multiple time aligned series.



Why use Qetch for our company?



- Regex operators
  - Not, Repeat
- Suggestions of functions to find
- Search within multiple datasets
- Multiple Time Series Queries



Negatives

- Qetch could be influential in the decision making

  → precision matters

- The query result is sensitive to the smoothing level of the time series data.
- Not operator needs accurate sketching

Questions?

Discussion

What queries are difficult to specify using Qetch's interface?

- Patterns over a specific window: e.g., cities with steepest rise or fall in temperature over 3 months, peaks with a width of 4 months.
- Summarizing common trends: e.g., find cities with typical weather patterns
- Sequence of patterns: e.g., cities with the following temperature trends over time: rise, flat, and fall, stocks with decreasing and then rising trends.

Discussion

What are other ways to write time series queries?

- Query by example
- Natural language description
- Better regex

Additional reading

ShapeSearch: A Flexible and Efficient System for Shape-based Exploration of Trendlines

Table 1: Comparison between specification mechanisms

Mechanism	Intuitiveness	Control	Expressiveness
Natural language	high	low	high
Sketch	high	high	low
Regex	low	high	high

Next class

Falx: Synthesis-Powered Visualization Authoring

Authors: Vishnu

Reviewer: Sankalp

Archaeologist: Yiheng