INFO 290T

Human-Centered Data Management

Paper Author Role: Qetch

Announcements

- Still a few accessory roles for next week available!
- Remember: you need 5+ credits, including one presenter role
- OH today to talk about projects!

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

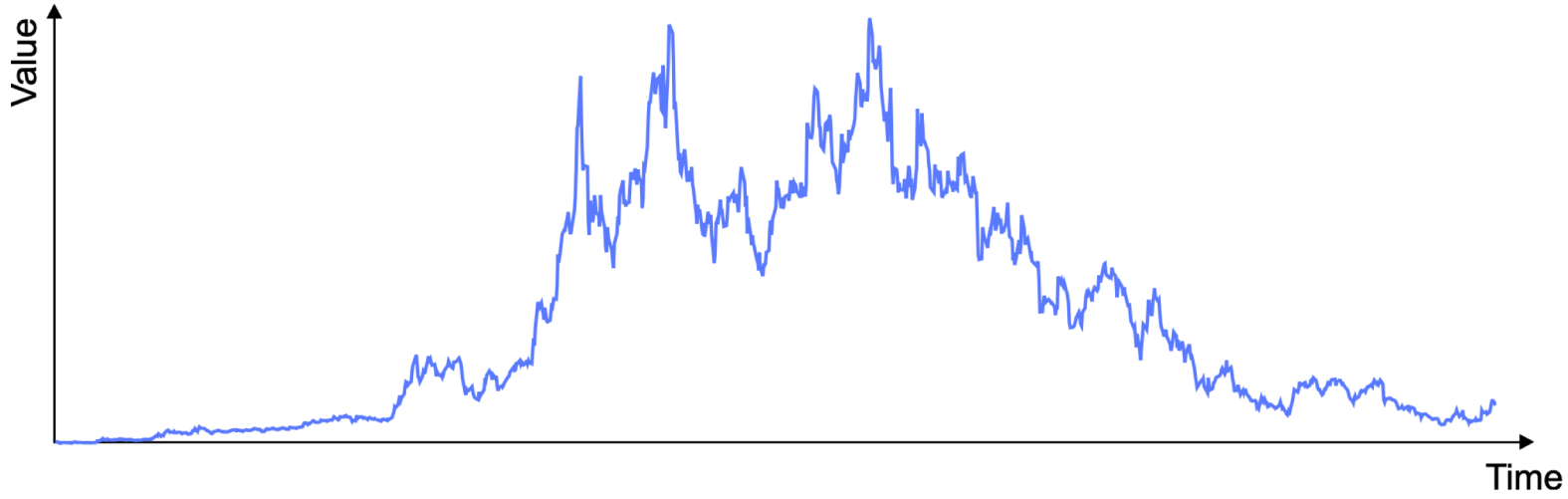
Miro Mannino, Azza Abouzeid

CHI 2018

Best Paper Award

Time Series Visualizations are Important

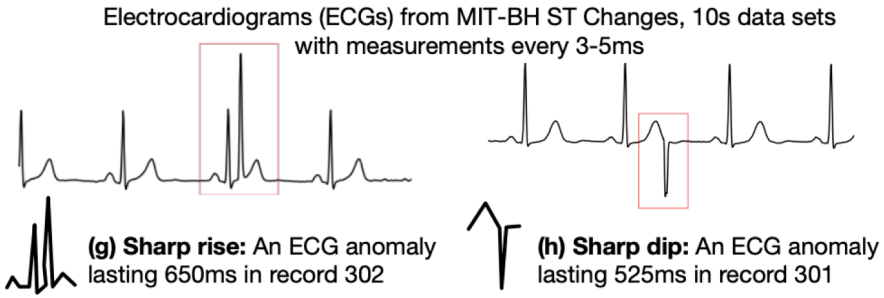
- Occur everywhere
  - Health data, sensors, financial data
- A special case of line charts



Querying = Pattern Matching is a Key Activity

Often want to find patterns in one or more time series

Significant financial or health implications, among others …



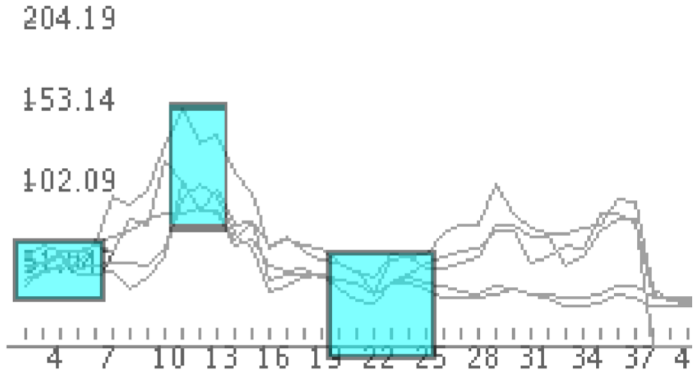
But… How do I Query for Patterns?

- Existing visualization tools (Tableau, …) provide no ways for users to search for specific patterns
  - Onus is on the user to manually sift through the time series for potential matches

- The most natural approach is for the user to simply *sketch* what they are looking for and the system to find matches…
  - Prior work doesn't support sketch enough to make sketch-based querying practical

Prior Work

- Sketch-less querying
  - E.g., timeboxes (timesearcher)
    - User draws constrain boxes
    - Value-based constraints, not shape
- Constrained sketching
  - Using the same sketching canvas as the data
    - User must match the scale and shape exactly
  - Sketches or constraints thereof must conform to a specific shape
    - e.g., lines



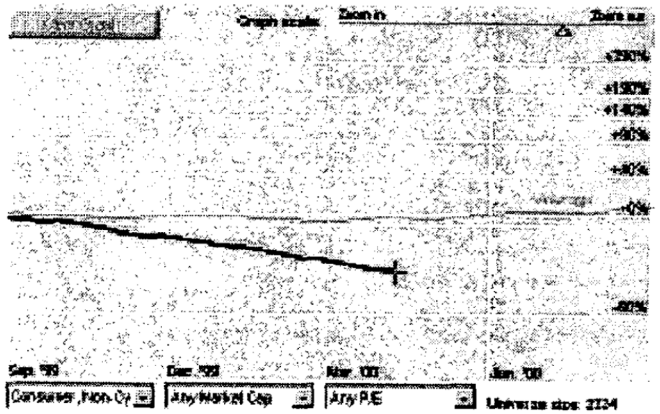
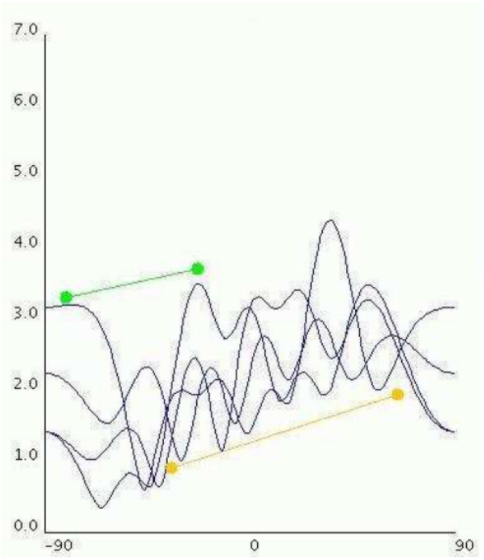


Figure 1: Query in progress. Heavy line is the user's query, light line displays aggregate data for reference.



OK, so let's support "free-range sketching"!

- But we don't actually know how people sketch time series!
  - What features they pay attention to
  - How faithfully they reproduce these features
  - How often they make mistakes

- *"... most humans are not faithful artists ..."* [Eitz et al., 2012]

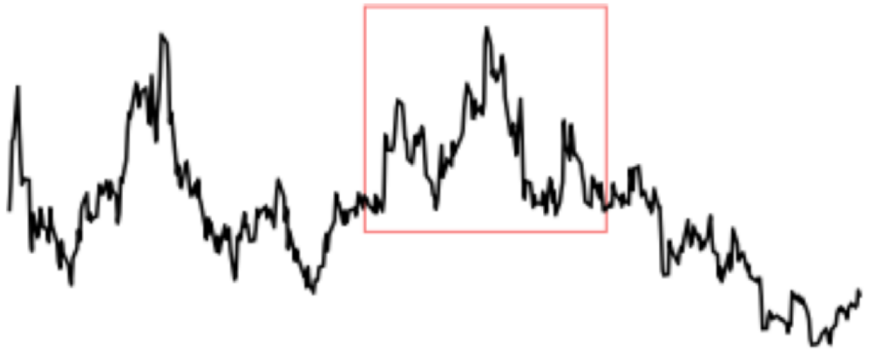
"... Instead people use shared, iconic representations of objects or they make dramatic simplifications or exaggerations…"

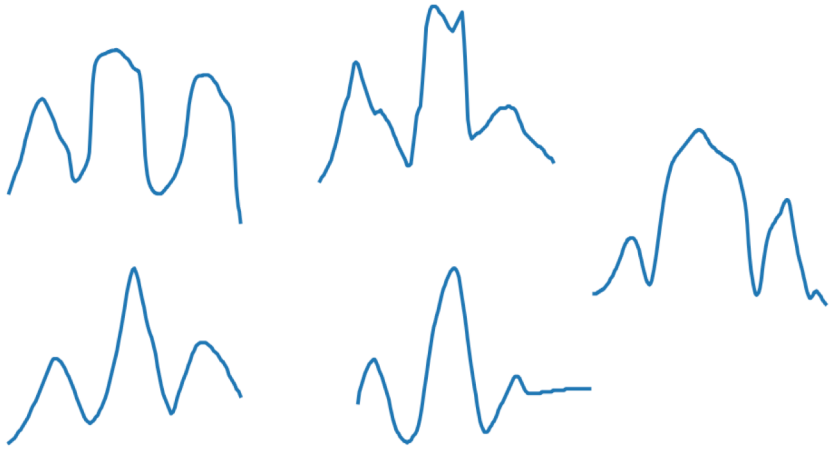
- So let's do a study to understand how people sketch time series!

How do people sketch?

Crowdsourced study: 150 crowd workers

Asked to reproduce a fragment of a time series (without tracing)





Key Findings:

- Visual features are preserved
- Non-uniform global scaling [not all components stretched/squished the same amount]
- Local distortions [humans often exaggerate some features]

Time Series Distance Measures

- Given our study findings, especially the non-global scaling, and local distortions, how do we identify matches to a query?
- Prior work on time series matching
  - Euclidean distance
    - Point-by-point matching
  - Dynamic Time Warping
    - Stretches/Squishes ranges of values to allow for more flexible matches
- However, both are for matching time series against each other, not for a hand-drawn sketch to a time series

So How Do We Match?

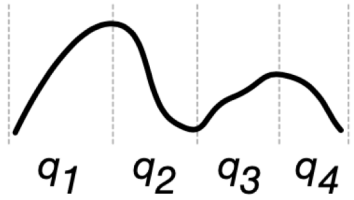
Three steps:

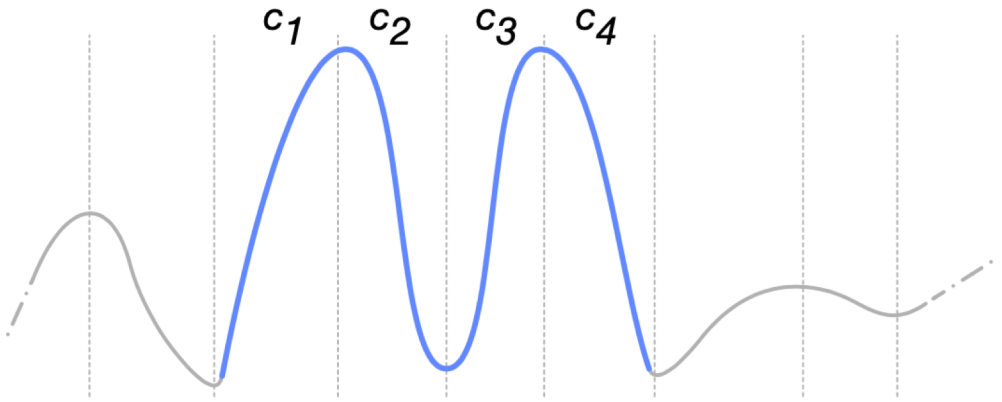
- *Support a variety of "smoothings"*
  - Raw → Smooth
  - Allows for emphasis of local vs. global trends
- Per smoothing, *select a match candidate*
- Per smoothing, per match candidate, *compute the distance*

So How Do We Match? II

Selecting a match candidate

- Segment query & TS into portions of +/- slope (segments)
  - Ignore small local variations [merge with neighboring segments]
- If there are *k* segments in query, consider each contiguous *k* in TS





So How Do We Match? III

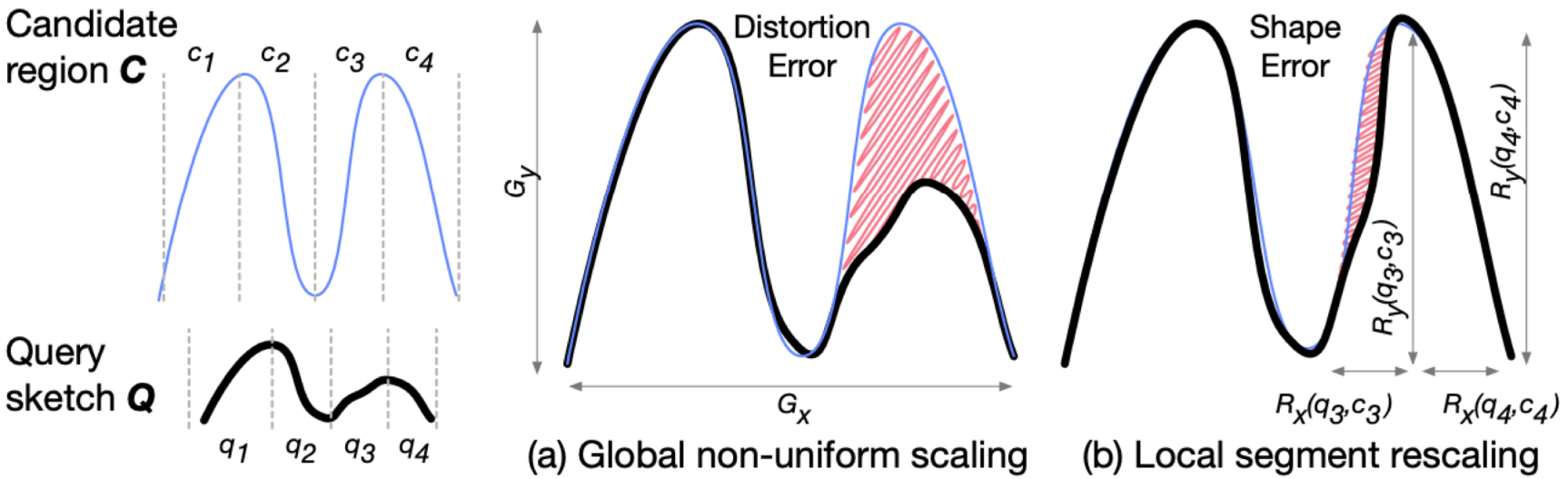
Both these factor into the score

Computing the distance

Compute global rescaling of height and width

Once rescaled with global values, find local issues in scaling (new local scale factors)

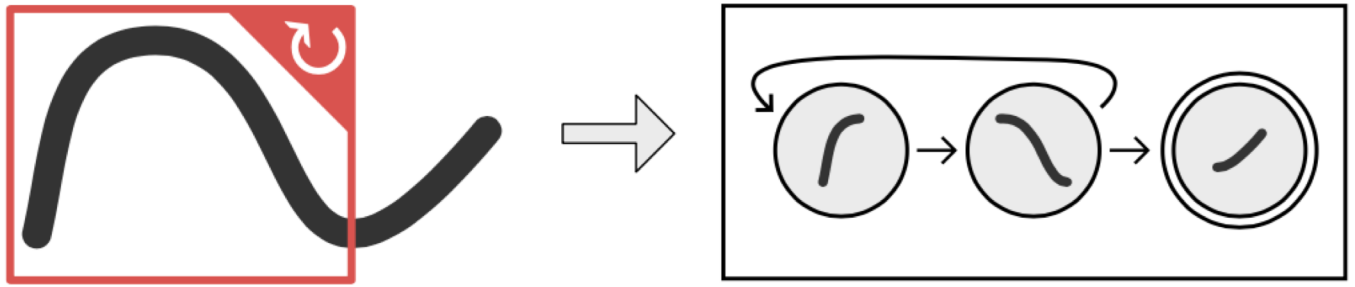
Given global and local rescaling, find point-by-point differences



Demo Video!

https://www.dropbox.com/s/h3njwcb0gsn0jti/qetch%205min.mp4?dl=0

Other Features

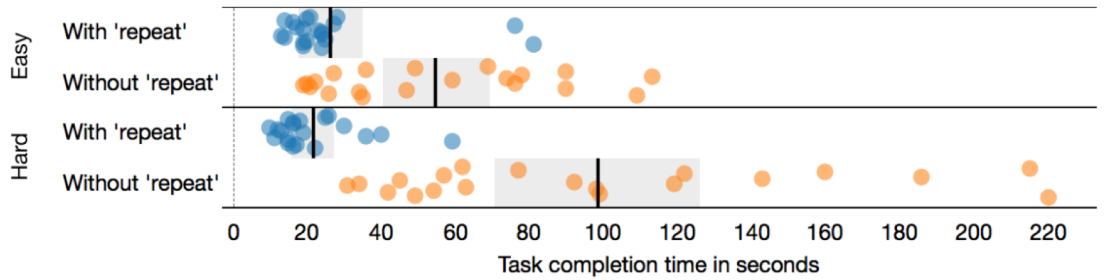


- Regex
  - Operators:
    - Repeat (avoid having to specify unknown repetitions, or draw many times),
    - Not (find anomalies)
    - Concat
  - Each segment in the sketch represents a state
  - All possible transitions are explored as long as the matches can be extended

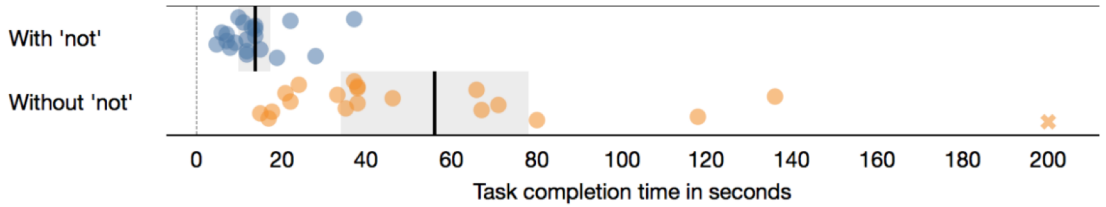
- Simultaneous match on multiple timeseries
  - Qetch uses relative positioning to define temporal constraints between sketches across time aligned data sets

Evaluation: Interaction Features

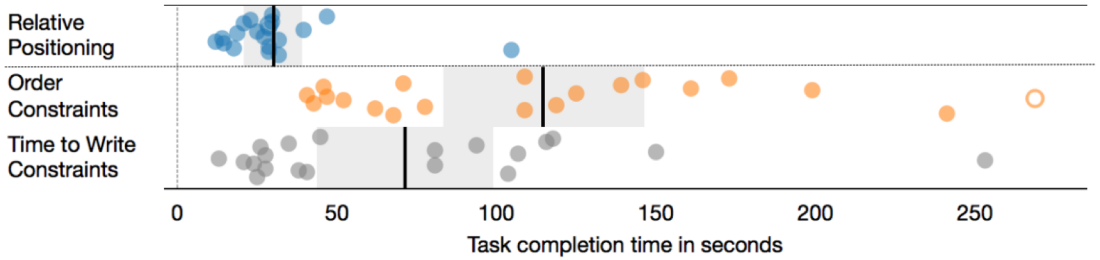
- User study on multiple tasks
- Tasks that require repetition, not operator, or multi-time series querying are easier with those features
- (Aditya note: duh!)



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



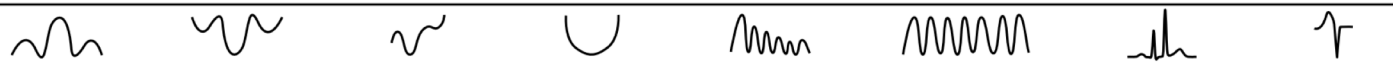
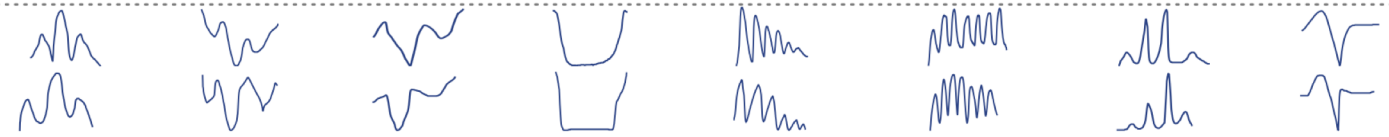
(b) Task completion times when searching for an anomaly with and without the not regex operator



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

Evaluation: Targeted Search

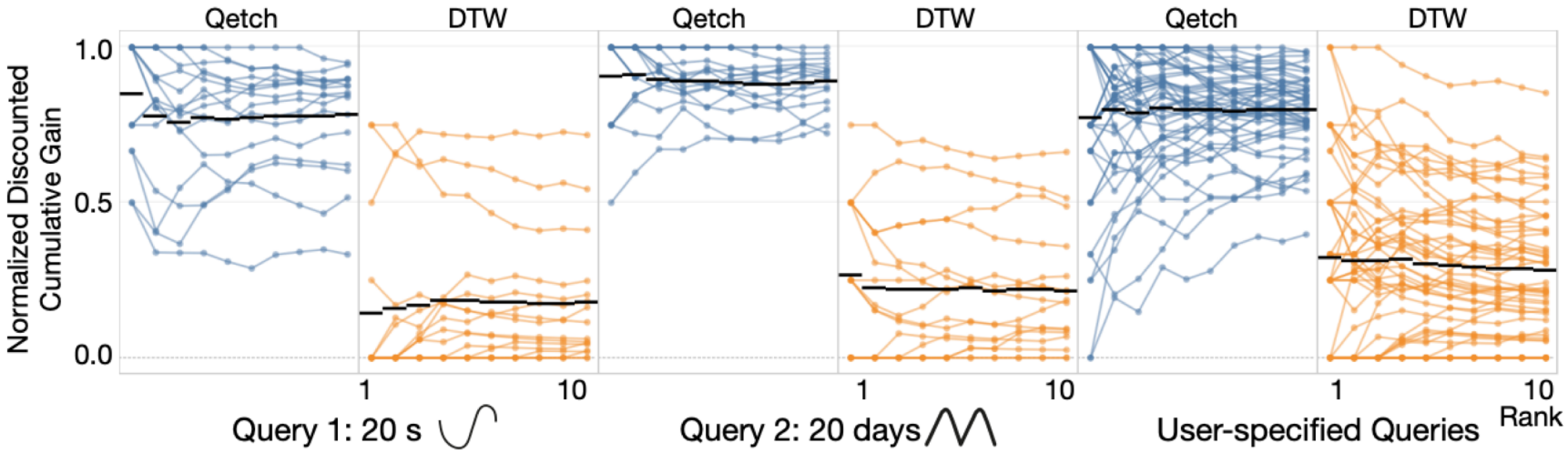
- Accuracy evaluation of Qetch's algorithm
  - Uses the crowdsourced dataset, where workers draw a ref region

Queries	
Sketch Samples Typical sketches preserve key perceptual features but have local distortions.	
DTW ranks the reference region at 16+ and Qetch ranks it within 1-15	
DTW ranks the reference region within 1-15 and Qetch ranks it at 16+	

- Counted how often each algorithm placed this reference region in its top 5 results
  - Qetch outperforms DTW in 5 data sets out of 8
  - Qetch outperforms ED in 6 data sets out of 8
- But, DTW and ED require query length

Evaluation: Exploratory Search

- User study on exploring Qetch
  - Asked users to draw 2 queries and three queries of their choice
  - Asked to rank the top 10 results of both DTW and Qetch
- Qetch outperforms DTW across all queries



Conclusions

- Prior matching algorithms not designed for sketch matching for humans
- Querying time series with human sketches is challenging

- Qetch = a tool and algorithm free-range sketch-based querying
  - Distance measure outperforms DTW and ED in many scenarios
  - No need to specify query length, no constraints in sketching
  - Regular expression and multiple matches help typical cases

Discussion

- What did you all think of the paper?
  - Key contributions
  - Evaluation
  - Writing
  - Interface
  - Adoption

Discussion: Query Completeness

- Operators: Not, Repeat
- Is the range of expressiveness sufficient?
- What else might be useful to add to the mix?

Discussion: Evaluation

- Multiple experiments = great!
- How could the evaluation be improved?
- Thoughts:
  - Experiments done on synthetic datasets --> ecological validity
  - Crowd worker study had no "skin in the game"
  - Real world use cases?
  - Within/Between subjects issues

Discussion: Adoption and Usability

- Do you think the system is easy to adopt?
- How would it fit into current workflows?

Discussion: Informal Archaeology

- Paper that this builds on:
  - Martin Wattenberg: Sketching a graph to query a time series database, 2001
  - First paper that uses "sketching" for querying
- Paper that builds on this paper:
  - ShapeSearch: A Flexible and Efficient System for Shape-based Exploration of Trendlines, 2020 [SIGMOD Best paper]
  - Expanded regex
  - Natural language -> Regex