

# Interactive Pattern Search in Time Series

Paolo Buono      Aleks Aris      Catherine Plaisant  
Amir Khella      Ben Shneiderman

By: Hany M. SalahEldeen  
CS 796/896 - Spring 2011, Old Dominion University

---

# About the Paper

**Authors:** Paolo Buono, Aleks Aris, Catherine Plaisant, Amir Khella, Ben Shneiderman

- **Affiliation:** University of Bari & University of Maryland College Park.
- **Venue:** Visualization And Data analysis (VDA 2005).
- **Year:** 2005.
- **Software:** TimeSearcher2.

# **Interactive Pattern Search in Time Series**

# **Interactive**

Real time negotiation and display of results

# **Interactive Pattern Search in Time Series**

# Pattern

Detecting repeating cycles across time or the way the data increases or decreases.

**Interactive Pattern Search in  
Time Series**

# Search

Not only answering the question of “what” but also “when”





Travel

Travel

Flights

Hotels

Deals

Destinations

Resources

Flight search

Flight results

Flexible search

▲ TIMES

[View in grid](#)

Leave: Depart | Arrive

05:45a 7:35p

Return: Depart | Arrive

12:25a 11:45p

Price\* ▲

Airline

Airports

Leave – Arrive

Stops

Duration

**\$694**

[Select](#)

Continental

ORF>SFO 5:35p – 11:06p 1 | 8h 31m | econ

SFO>ORF 11:05a – 10:19p 1 | 8h 14m | econ

[⊕ Flight details](#)

Book with: [\\$694 continental.com](#) · [\\$694 orbitz.com](#)

**\$694**

[Select](#)

Continental

ORF>SFO 5:35p – 11:06p 1 | 8h 31m | econ

SFO>ORF 7:25a – 7:45p 1 | 9h 20m | econ

[⊕ Flight details](#)

Book with: [\\$694 continental.com](#) · [\\$694 orbitz.com](#)



Travel

Travel

Flights

Hotels

Deals

Destinations

Resources

Flight search

Flight results

Flexible search

▲ TIMES

[View in grid](#)

Leave: Depart | Arrive

05:45a 08:19a

Return: Depart | Arrive

2:25p 8:49p

Price\* ▲

Airline

Airports

Leave – Arrive Stops | Duration

**\$1055**

[Select](#)

American

ORF>SFO

6:05a – 2:25p

1 | 11h 20m | econ

SFO>ORF

3:00p – 7:25p

1 | 25h 25m | econ



[⊕ Flight details](#)

Book with: [\\$1055 aa.com](#)

**\$1195**

[Select](#)

American

ORF>SFO

6:05a – 11:55a

1 | 8h 50m | econ

SFO>ORF

3:00p – 7:25p

1 | 25h 25m | econ



[⊕ Flight details](#)

Book with: [\\$1195 aa.com](#)



Travel

Travel

Flights

Hotels

Deals

Destinations

Resources

Flight search

Flexible search



Browse Bing Travel

1-10 of 210 results

Reset

DEPARTURE CITY

Norfolk - ORF ›

DESTINATIONS

☒ San Diego - SAN ›

Select a city ›

30-DAY DATE RANGE

Depart between

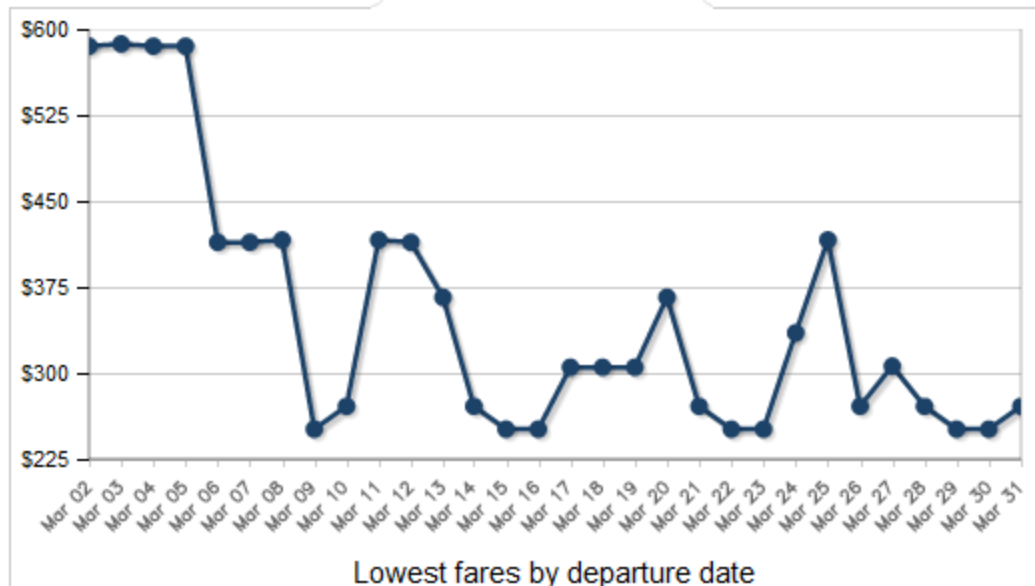
03/02/2011

and

3/31/2011

Map – where to travel

Graph – when to travel



**Interactive Pattern Search in  
Time Series**

# **Time Series**

# Time Series

- Simply data spread across time.
- It is the occurrences of events or observations to a certain variable at equal time interval



# Time Series

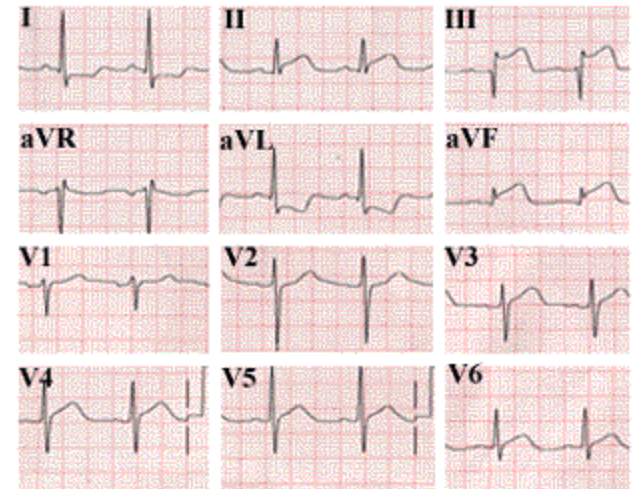
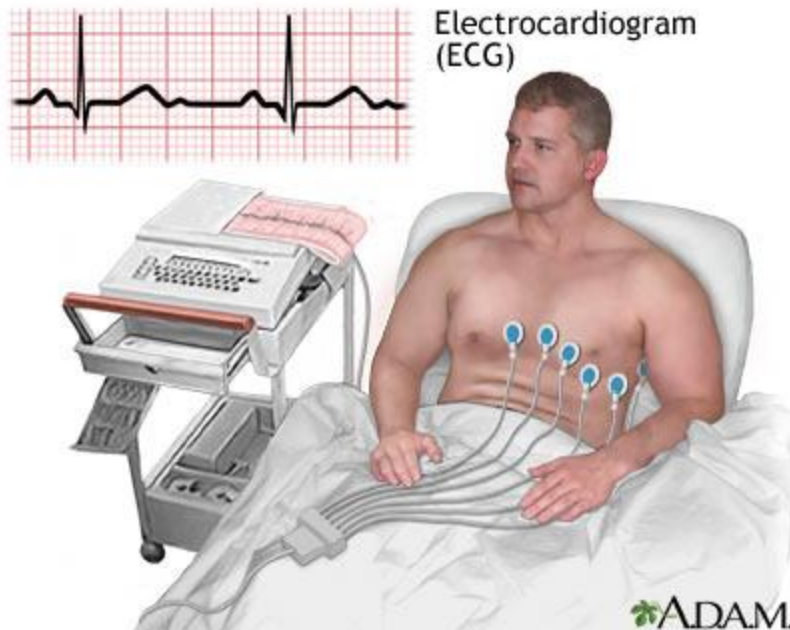
Applications:

- Sound Recordings
- Meteorology

# Time Series

Applications:

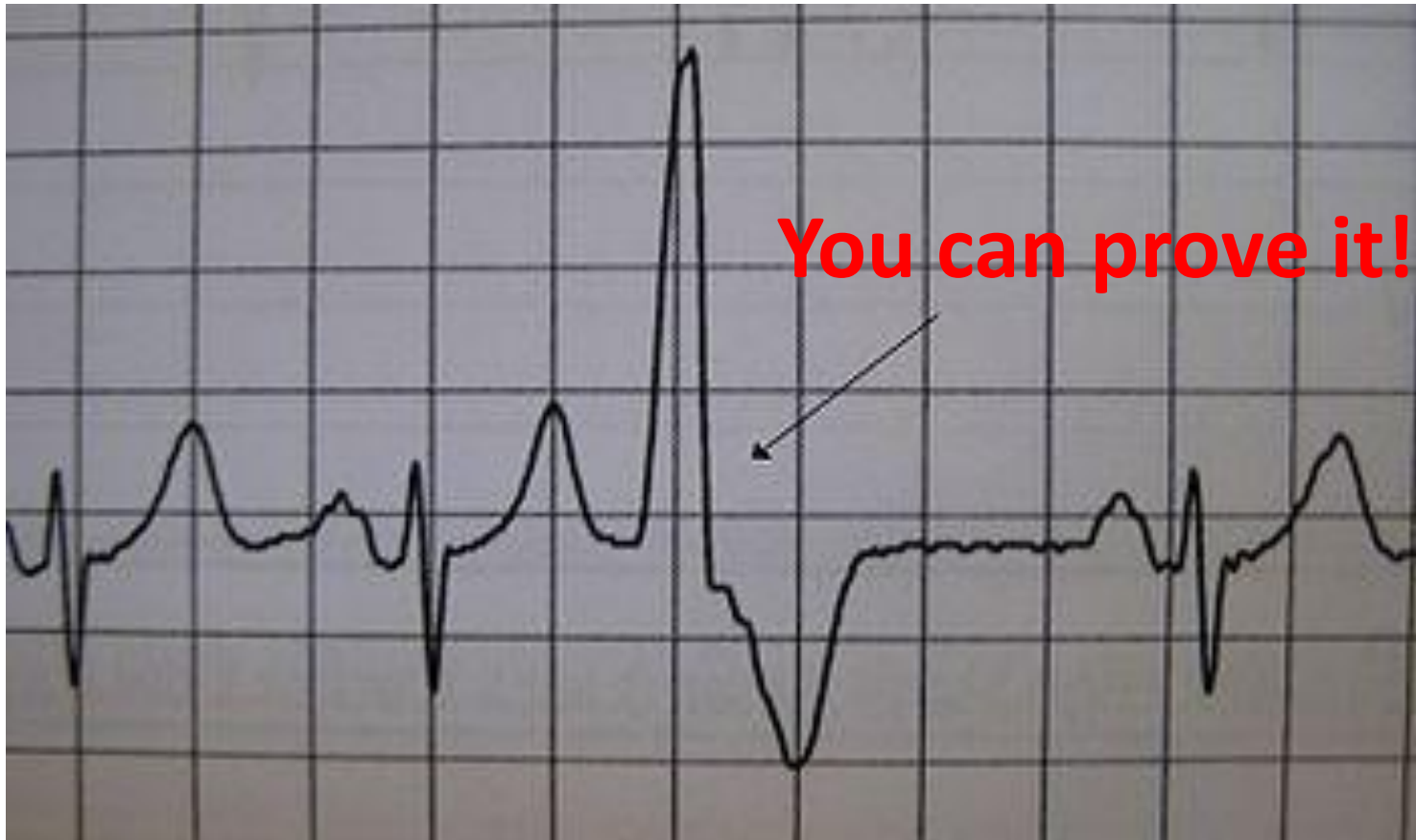
- Sound Recordings
- Meteorology
- Electrocardiograms ECG





You give me *premature ventricular contractions!*

A.K.A: you make my heart skip a beat!



# Time Series

There is a need to visualize long periods of data to:

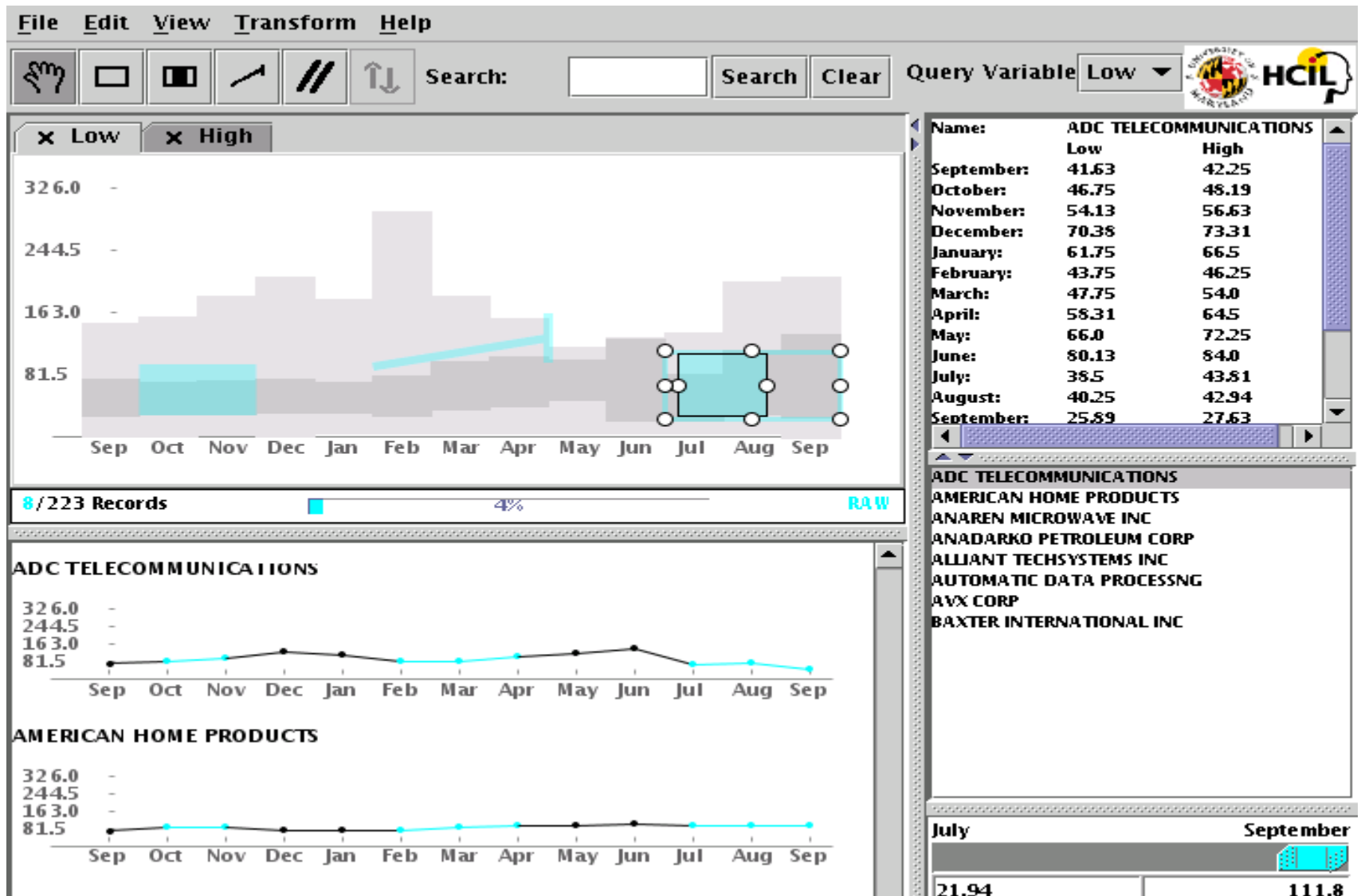
- Apply perceptual abilities in analysis.
- Identify trends.
- Spot Anomalies.

# TimeSearcher 1

Is the original model that this paper built its newer version on.

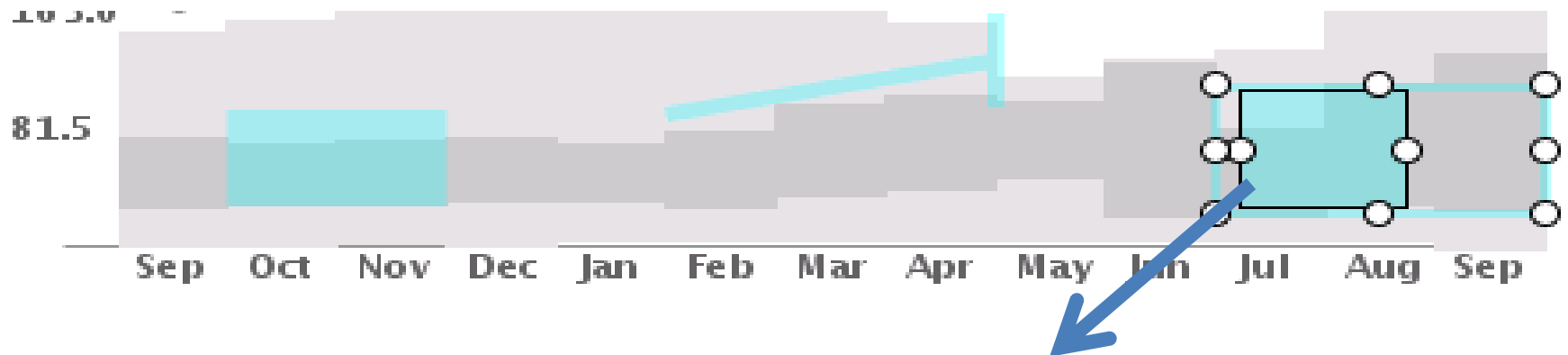
- Easy to use.
- No statistical analysis skills needed
- Introduce concept of TimeBoxes

# TimeSearcher 1



# TimeBox

- It is a rectangular region selected on timeline overview of data.
- Reflect the query on data through the boundaries.



**= Select from Data where time>July and  
time<Aug and income>60% and income<86%**

# TimeBox

Version 1: Filter data + Reduce search scope

Version 2: Perform pattern search across data

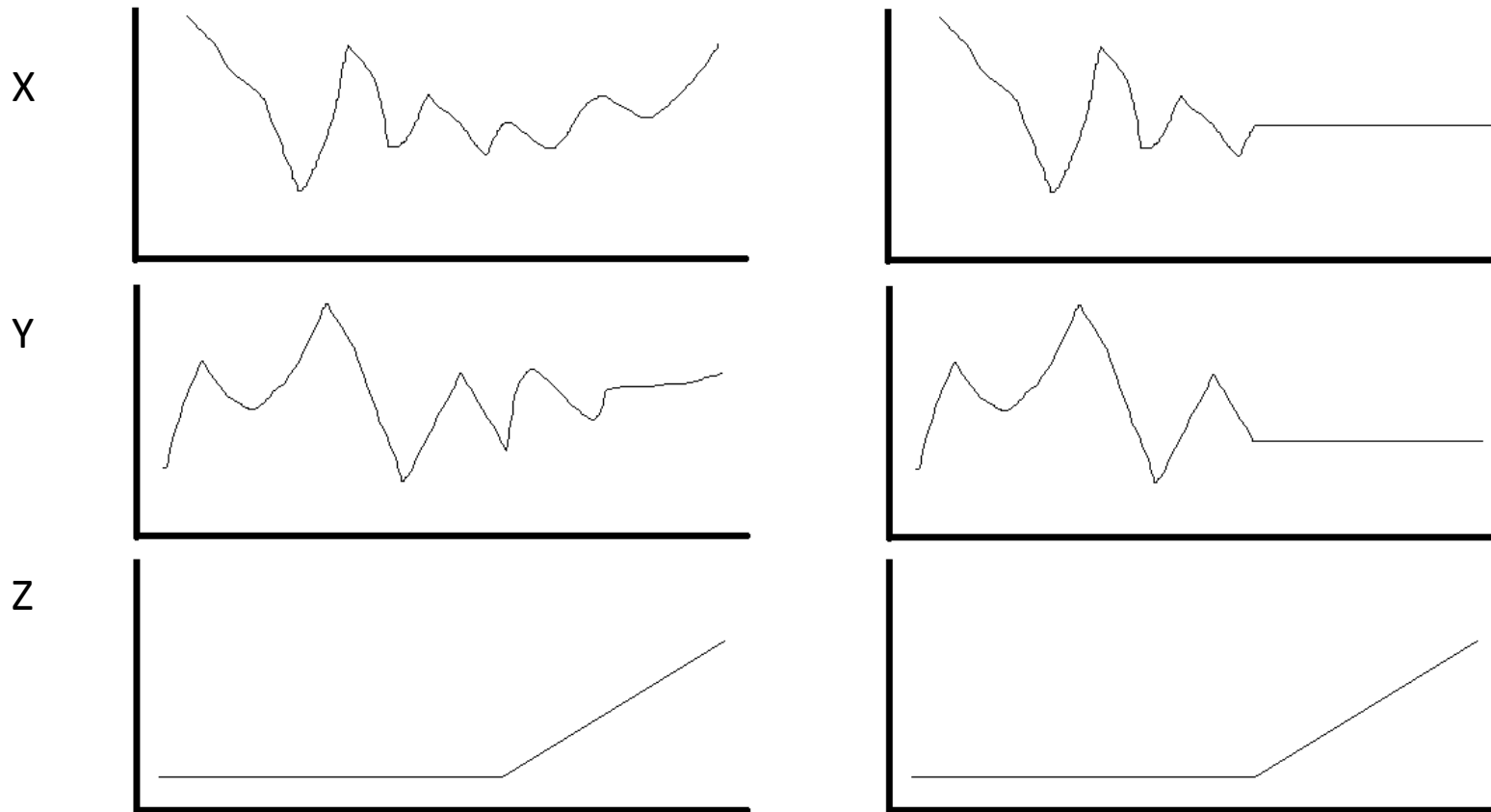
# TimeSearcher2

## Contribution:

- Extend TimeBoxes to incorporate pattern search + filter search.
- Three-step approach
- Enhance user-interface for long time series.
- Multiple heterogeneous variables up to 8!

# TimeSearcher2

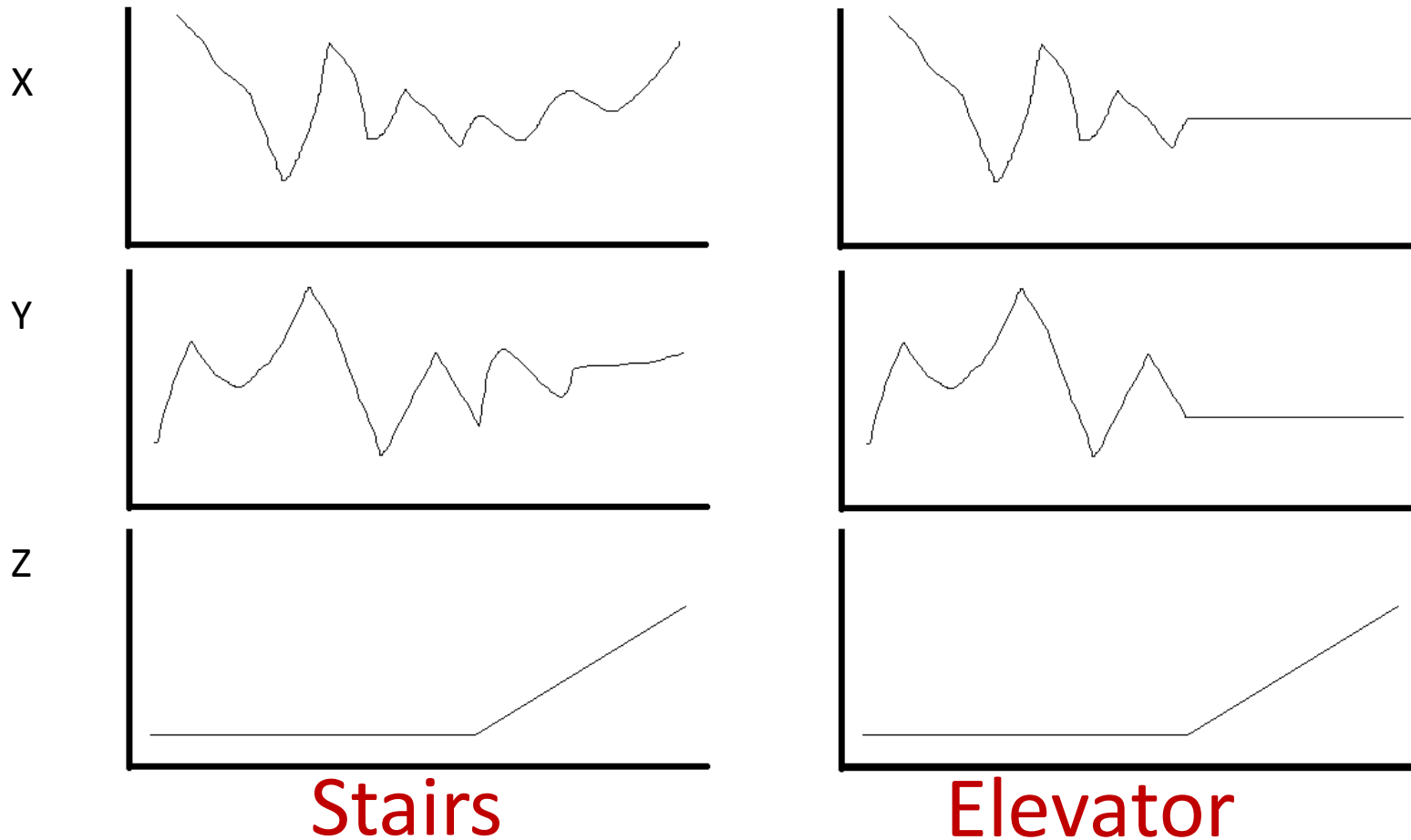
Multiple heterogeneous variables





# TimeSearcher2

Multiple heterogeneous variables



# Related Work

## Diamond Fast:

### Pros:

- One of the first.
- Visualize, move, resize time series.
- Compare more than 2 series.
- Management for missing values.

### Cons:

- Only manage very short times.

# Related Work

## ILOG & Personal Stock Monitor:

### Pros:

- High level interaction
- Enhanced zoom features.

### Cons:

- No search capabilities.
- Only one single time series.

# Related Work

**Semantic Zoom:** Very long time series.

**Spiral View:** Tightening and relaxing view

**Choratas:** Define patterns by numerical parameters.

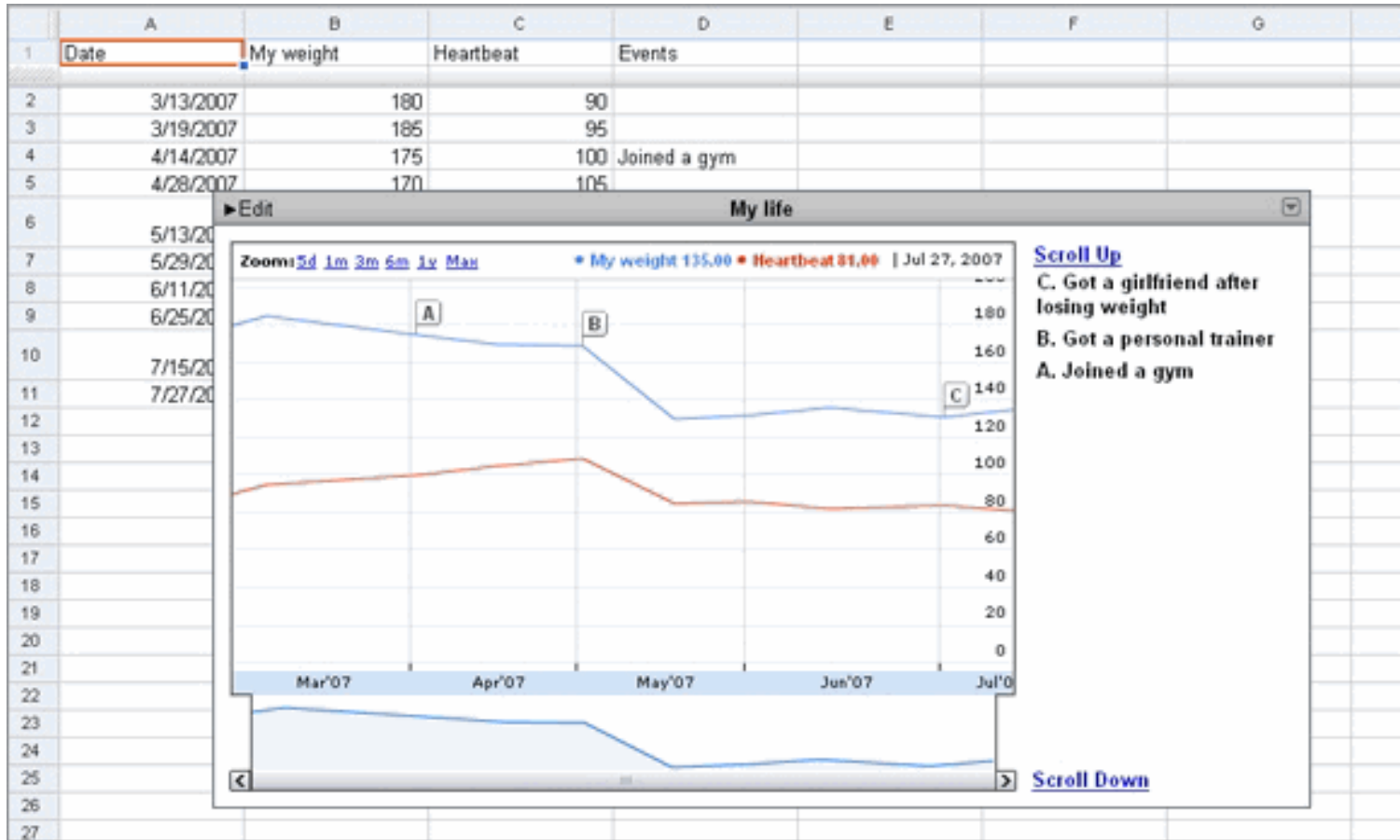
**VizTree:** Define patterns by shape

**Query Sketch:** Direct sketching the shape of the required pattern

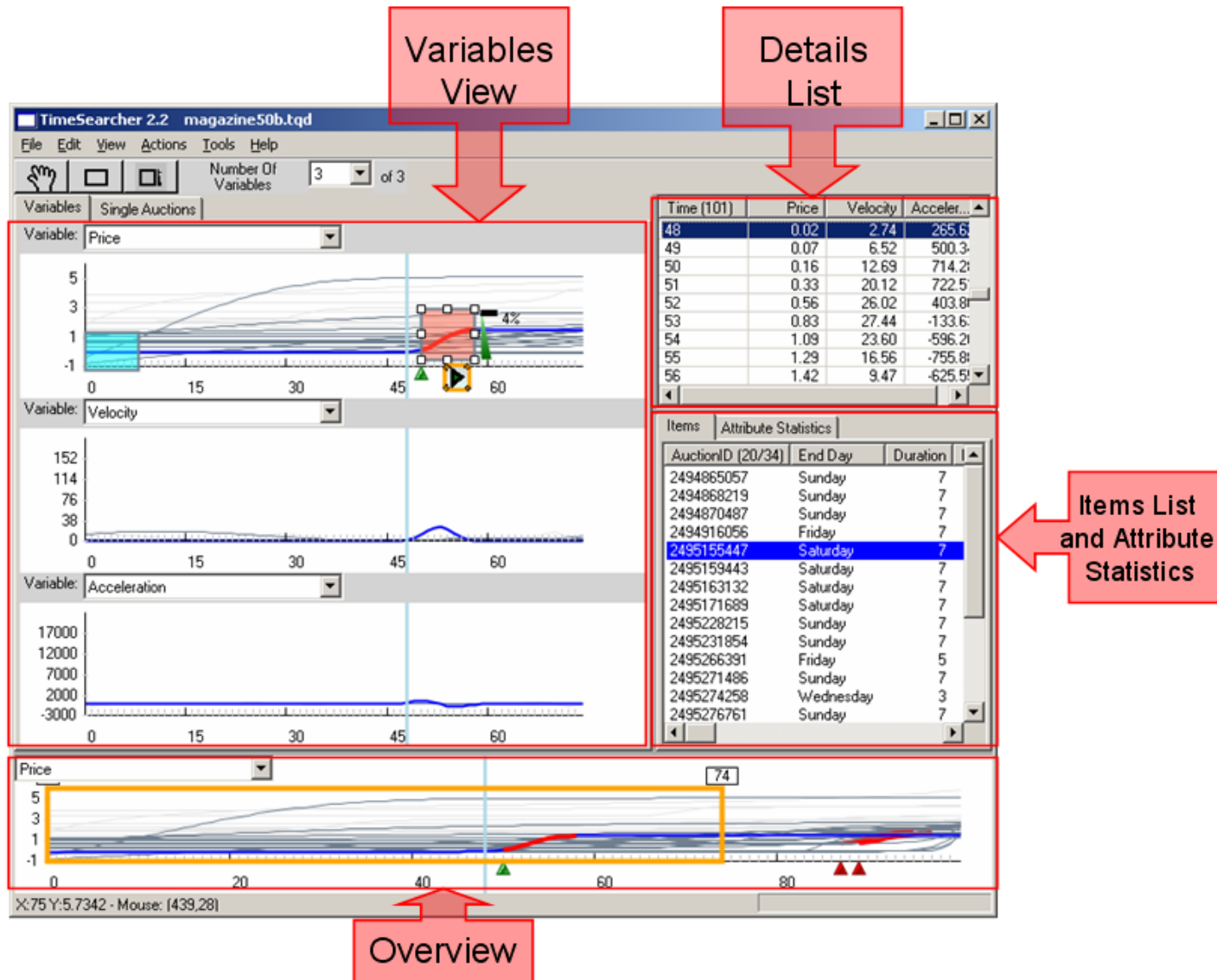
**IPBC:** 3D tool

# Extra Related Work

## Google Docs & Gadgets:



# Interface



# Demo

# Interactive Pattern Search

**Problem:** Direct search for pattern will be very slow. → Risk Interactivity.

**Solution:** The 3-Step Framework.

**Aim:**

- 1- Reduce the scope of the query by slicing and dicing.
- 2- Let user specify a pattern.
- 3- Let the user refine the results.



# Three step search framework

## Step 1:

- Create many Timeboxes.
- Reduce big long view by ANDing the boxes

→ Same like looking in a phone book or a dictionary, proximity search.

# Three step search framework

## Step 2:

- Select a specific pattern by selecting an element.
- Search within scope.
- Adjust tolerance slider to reach desired results.

# Three step search framework

## Step 3:

- Select search parameters to tune.
- Adjust the parameters till finding the desired result.

# Search Algorithm

**Euclidean Distance:** Simple algorithm utilizes getting difference between all corresponding points on time series.

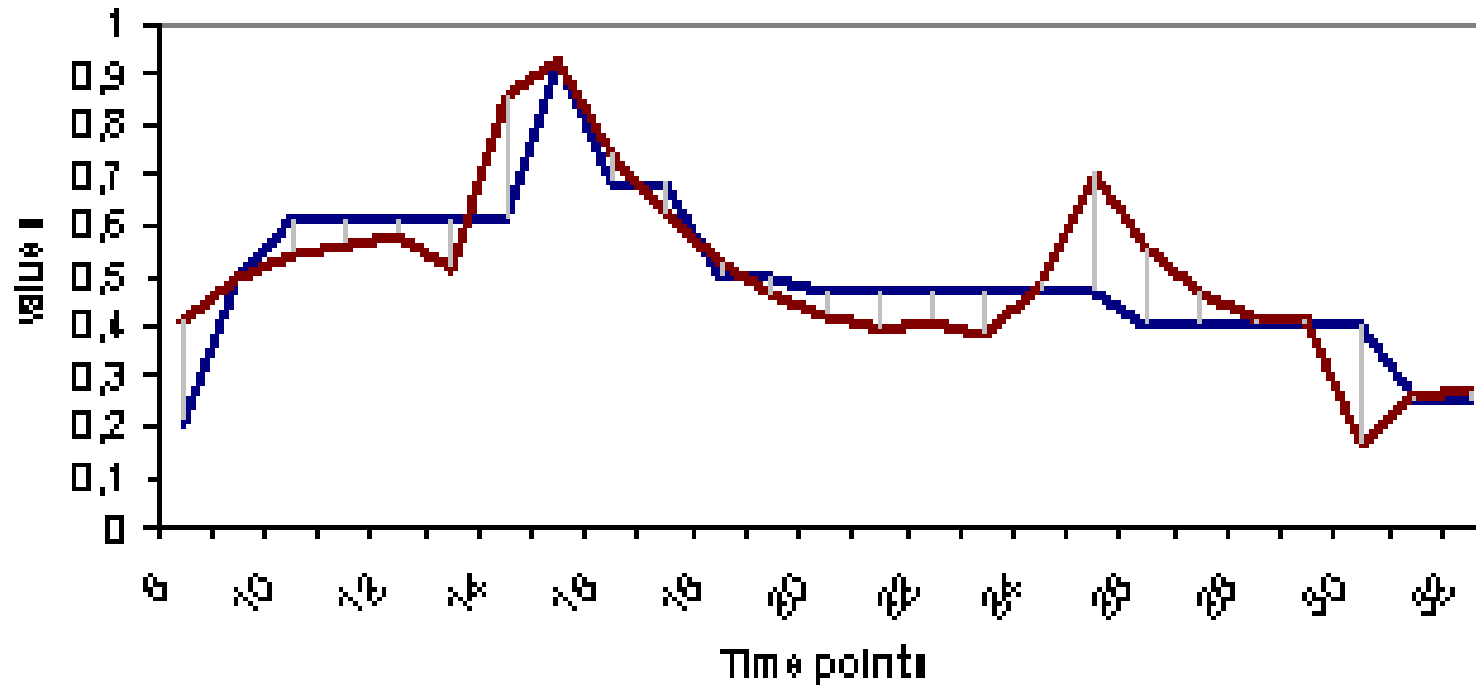
$$D(Q, C) \equiv \sqrt{\sum_{i=1}^n (Q_i - C_i)^2}$$

Since it is monotonic:

$$D(Q, C) \equiv \sum_{i=1}^n (Q_i - C_i)^2$$

# Search Algorithm

Euclidean Distance:



# Search Algorithm

## Euclidean Distance:

### Pros:

- Simple to calculate and very easy to implement.
- Effective when the data is non periodic in date

### Cons:

- Very slow.
- Non-scalable.
- Only for small sets.
- Freezing and stopped interaction with big sets.

# Search Algorithm

## Euclidean Distance:

### Solution:

- Offset translation.
- Magnitude scaling.
- Linear trend removal.
- Noise Reduction.

→ Enable/Disable by checking the boxes.

# Problems

## Dealing with Large Datasets:

- With big datasets dynamic view fails.
- Introducing steps is not always a good idea (what if I didn't know what am looking for!?).
- Reducing sometimes not enough.



# Problems

## Missing Data:

- Can detect Missing data but no way to recover.

## Evaluation:

- There is no kind of evaluation of the system, user analysis is needed.

## Enhance Interaction:

- You can start from any step.
- Enhance drawing patterns not just select.

# My insight: Ways to improve

## Dealing with Larger Datasets:

- Since data is predefined and not on the fly then appropriate preparation is needed
  - Multilayer indexing.
  - Across parameter range indexing (B-Trees).
- Use intelligent Heuristics from info about the data.
- Clustering portions of similar data to speed up pattern search.

# My insight: Ways to improve

## Dealing with Larger Datasets:

- Apply feature detection techniques and then classification of patterns → good for not close matches and scaled matches.

# My insight: Ways to improve

## Dealing with Larger Datasets:

- Bottleneck is Euclidean Distance:
  - Upon performing the search calculation it could be done by transforming to approximate other dimension (good for pattern with multiscale)
  - Use dynamic programming in what you calculated once never calculate again. (index differences)

# My insight: Ways to improve

## Missing Data:

- Since some of the data is cyclic or periodic.
  - Try to make intelligent estimate.
  - Make a step in the data preparation initially to retrieve missing data.
  - Ignore estimated portions in pattern search.

## Evaluation:

- A user experience analyst and designer needs to be hired in the process.

# My insight: Big Problems

- Not maintained at all, still in .Net 1.1 (Now 3.0).
- Last Contribution was in 2005.
- Version 3.0 is frozen.
- Bad transition from extremely portable Java implementation to C#.
- Would be wonderful if they created an API to integration with other programs.
- Data needs to be prepared first before entering (CSV,...etc.)

# My insight: Questions

- What do we need in time series analysis?
- How to make input data on the fly?
- What possible algorithms could be utilized to replace the Euclidean distance?

Questions?

---