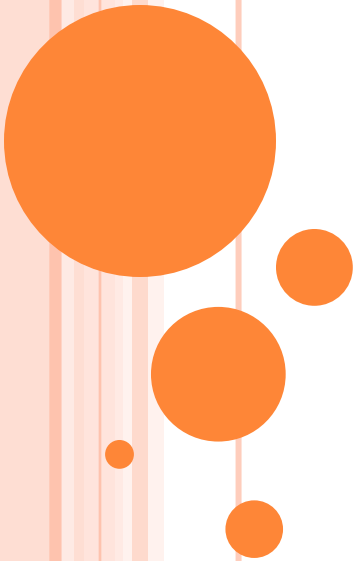


Time Series Data

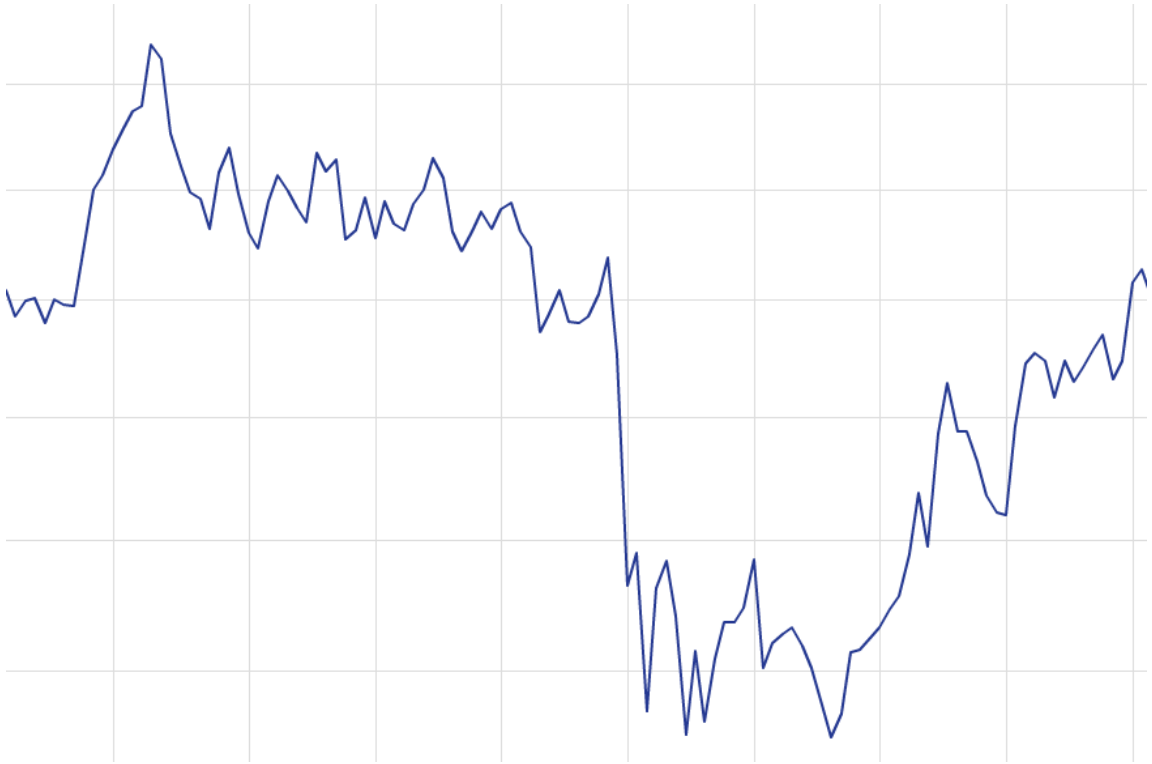
Papers Covered

- Interactive Visualization of Serial Periodic Data
  - John V. Carlis and Joseph A. Konstan
- Visualizing and Discovering Non-Trivial Patterns in Large Time Series Databases
  - Jessica Lin, Eamonn Keogh, Stefano Lonardi
- Time-series Bitmaps: A Practical Visualization Tool for working with Large Time Series
  - Nitin Kumar, Nishanth Lolla, Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana

What is time series Data?

What is time series Data?

- A value over time



What is time series Data?

- ~~A value over time~~
  - not too useful
- A sequence of time point + value pairs
  - $< t_0, v_0 >$
  - $< t_1, v_1 >$
  - $< t_2, v_2 >$
  - …
  - $< t_n, v_n >$

What is time series data?

- $t_i \leq t_{i+1}$
  - not $t_i < t_{i+1}$
- Low resolution of time
- Errors
- Discontinuities
- Multiple sources of measurement

What is time series data?

- common examples:
  - financial data
  - electrocardiograms
  - meteorological data
  - production rates
  - …

What is time series data?

- Doesn't need to be a numerical value over time
  - routes
    - position over time
  - schedules
    - Activity over time (resource focused)
    - resource over time (activity focused)

Tasks with time series data

- Finding patterns
  - periodic vs non-periodic
  - finding known patterns
    - searching
    - sequence matching
    - classification
  - finding common unknown patterns
    - motif discovery
    - clustering
  - finding rare patterns
    - anomaly detection

Tasks with time series data

- Finding trends
  - general increasing/decreasing
  - abrupt changes
    - anomaly detection
  - correlation between variables

Paper 1

- Interactive Visualization of Serial Periodic Data
  - John V. Carlis and Joseph A. Konstan

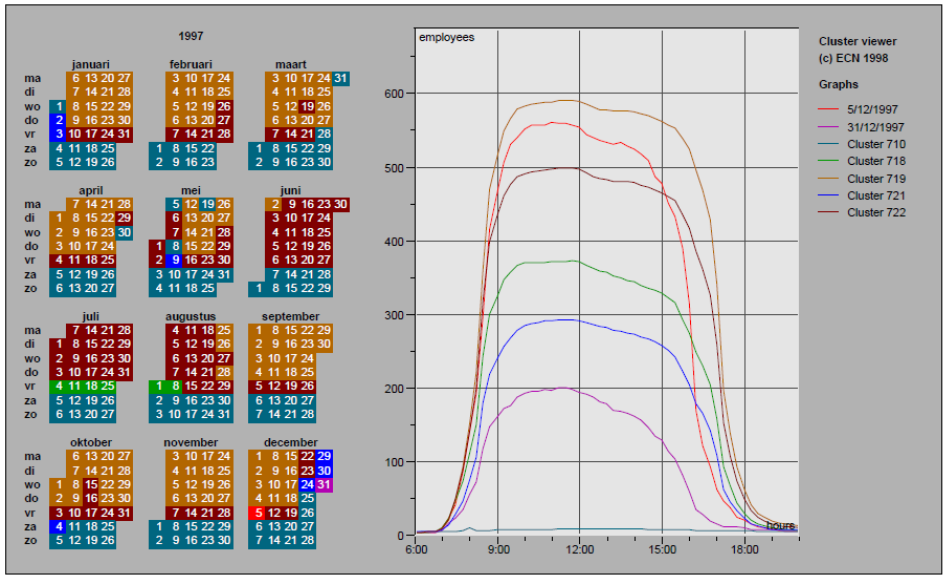
Periodic Data

- “Pure” periodic data
  - each period has identical duration
- vs event anchored periodic data
  - periods start following some event
  - time between events may be inconsistent

- Focus is on pure periodic data

Periodic Data

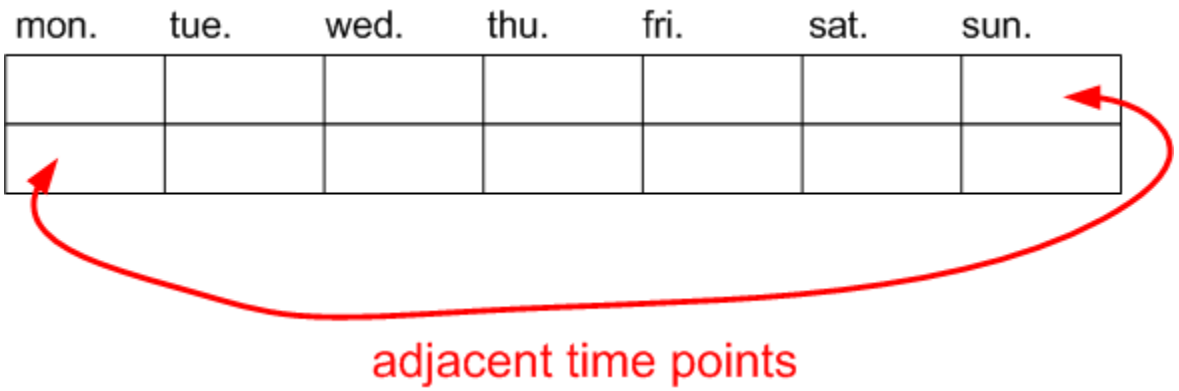
- Initial Approach: Calendars (tabular layouts)



Cluster and Calendar based Visualization of Time Series Data. Jarke J. van Wijk and Edward R. van Selow, Proc InfoVis 99

Periodic Data

- Calendar (tabular) layouts exaggerate distance between adjacent periods



Periodic Data

- Calendar (tabular) layouts exaggerate distance between adjacent periods
- Solution: layout the series in a spiral

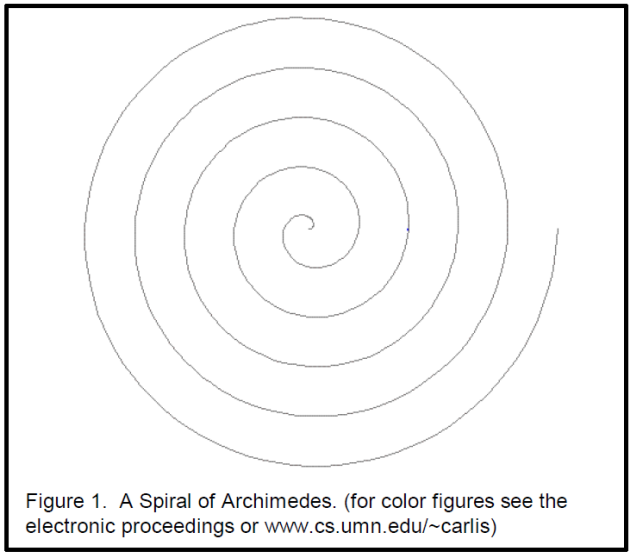


Figure 1. A Spiral of Archimedes. (for color figures see the electronic proceedings or www.cs.umn.edu/~carlis)

Periodic Data

- The end of one period is close to the start of the next.
- Encodes time with two visual attributes
  - distance from center is time
  - angle is time relative to start of period
- Values at time points must be encoded some other way
  - same with tabular layouts

Periodic Data

- dot size

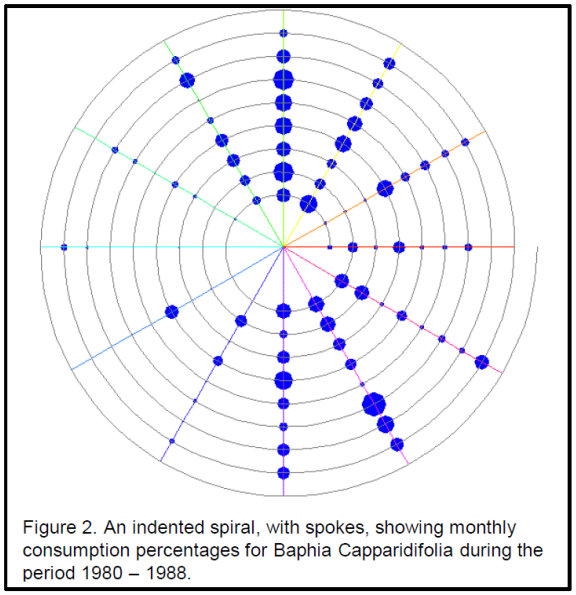
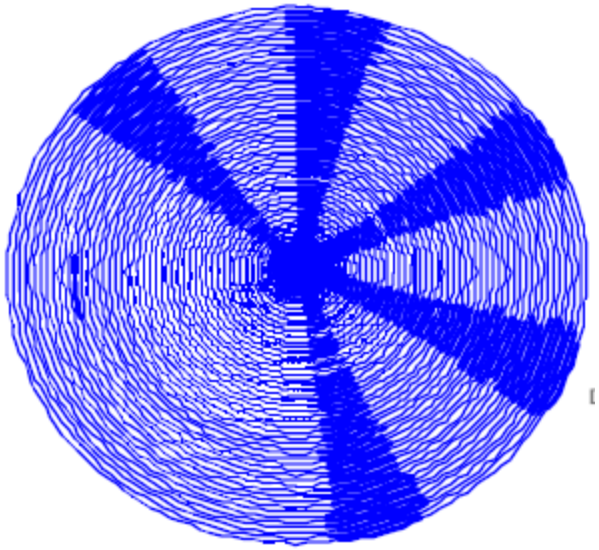


Figure 2. An indented spiral, with spokes, showing monthly consumption percentages for Baphia Capparidifolia during the period 1980 – 1988.

- line width



Periodic Data

- glyph

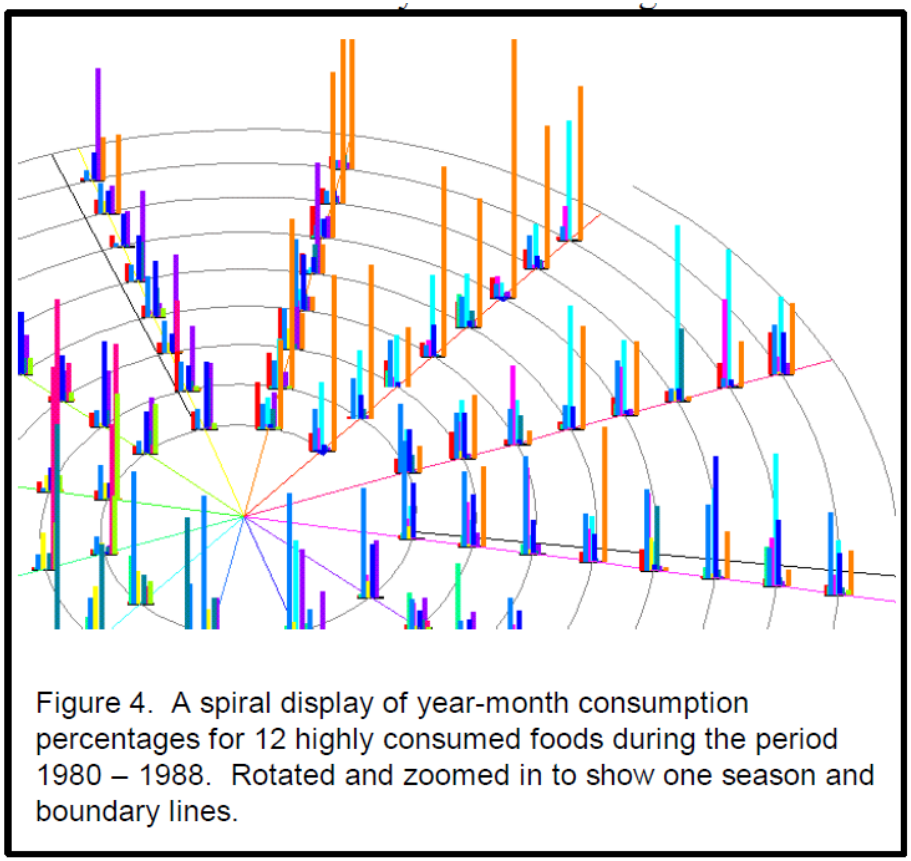


Figure 4. A spiral display of year-month consumption percentages for 12 highly consumed foods during the period 1980 – 1988. Rotated and zoomed in to show one season and boundary lines.

Periodic Data

- Interaction
  - manually adjust period length

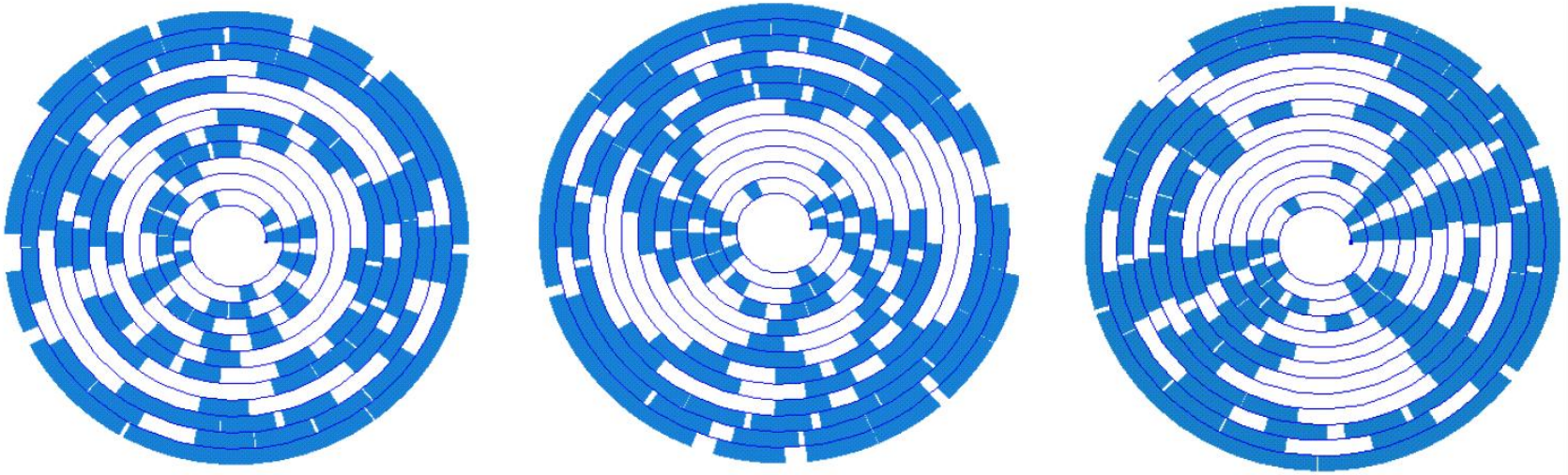


Figure 7. Tightening a spiral view of sound data from five instruments. From left to right the structure of the sound reveals itself.

Periodic Data

- Interaction
  - change point of view (for 3D spirals)

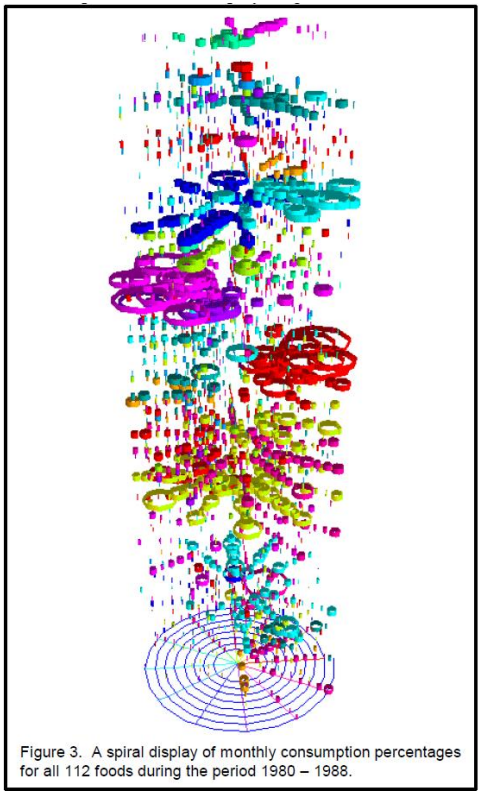


Figure 3. A spiral display of monthly consumption percentages for all 112 foods during the period 1980 – 1988.

Periodic Data

- good:
  - space efficient
  - neighbouring points are always near each other
  - easy to tell where a point is within a period
- bad:
  - points within the same period may be very far apart
  - inconsistent density
  - can‘t display many variables
    - glyph occlusion
    - bewildering 3D views

Paper 2 & 3

- Visualizing and Discovering Non-Trivial Patterns in Large Time Series Databases
  - Jessica Lin, Eamonn Keogh, Stefano Lonardi

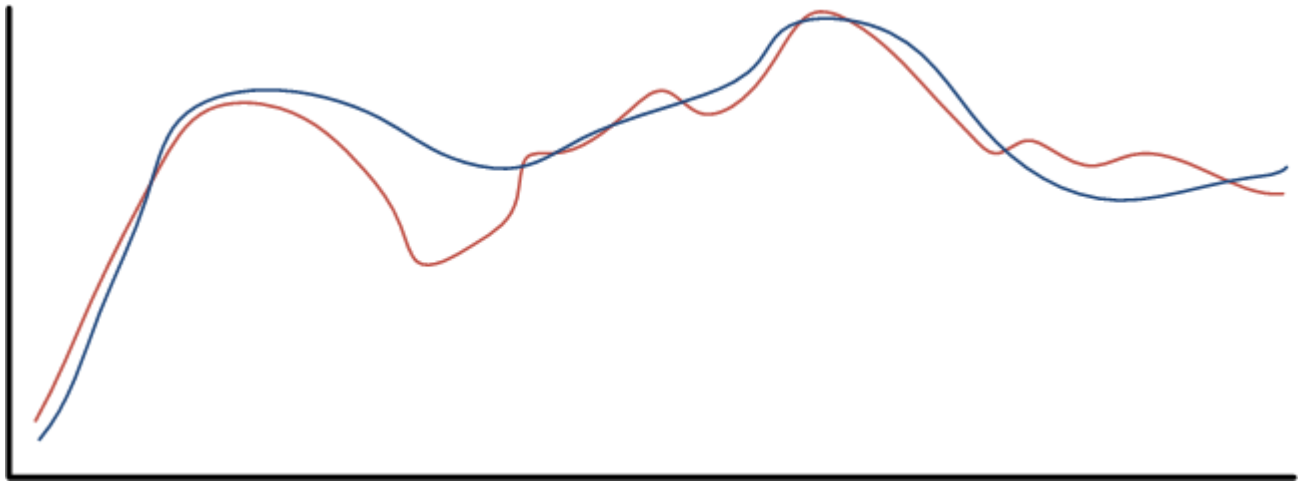
- Time-series Bitmaps: A Practical Visualization Tool for working with Large Time Series
  - Nitin Kumar, Nishanth Lolla, Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana

Pattern Detection

- Observation:
  - sequence matching and pattern detection is a lot easier for strings
- Symbolic Aggregate approXimation (SAX)
  - dimensionality reduction

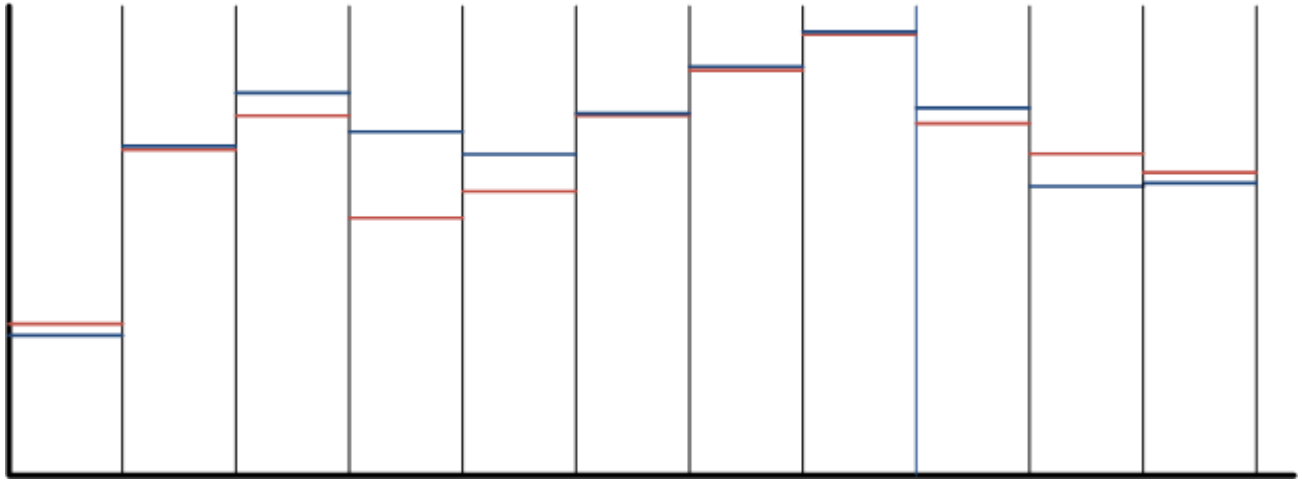
Pattern Detection - SAX

- From initial time series…



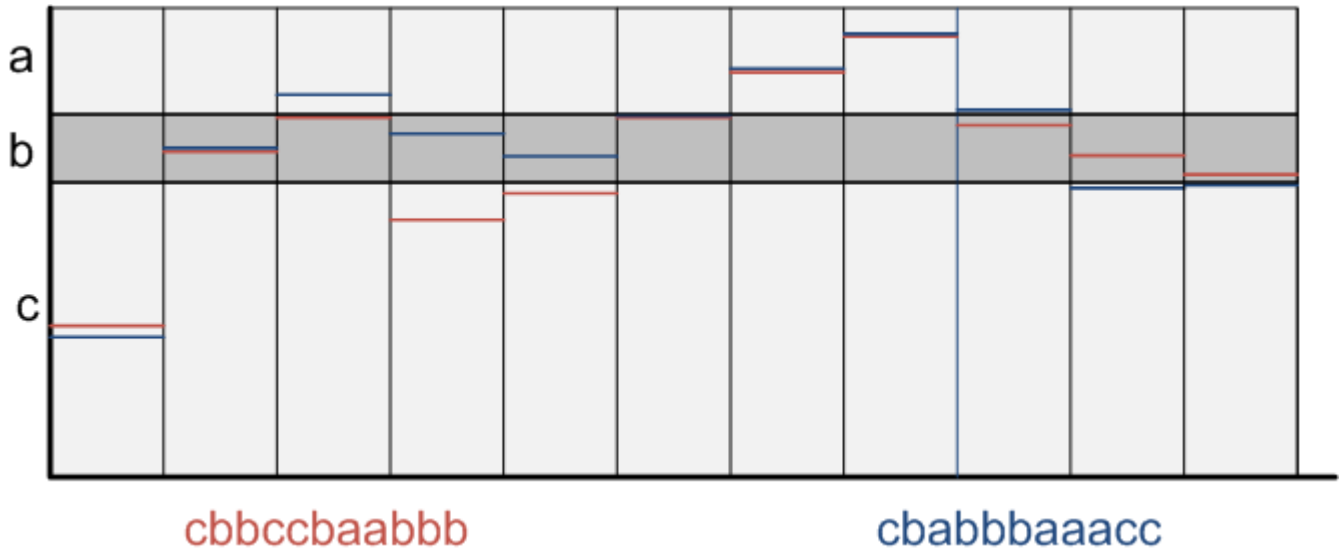
Pattern Detection - SAX

- First step, discretize time into ***w*** equal sized intervals
  - aggregate the points within each interval (ie, average)



Pattern Detection - SAX

- Second step, discretize the value for each interval into an alphabet of size ***α***
  - should result in equiprobable symbols

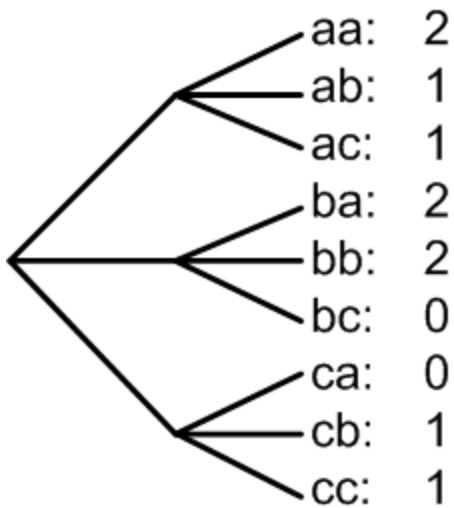


Pattern Detection - SAX

- Linear trends could make patterns meaningless
  - Could get patterns like aaaaabbbbbbccccc.
- Use a short sliding time window
  - symbols are equiprobable within the time window
  - produces a set of strings instead of just one

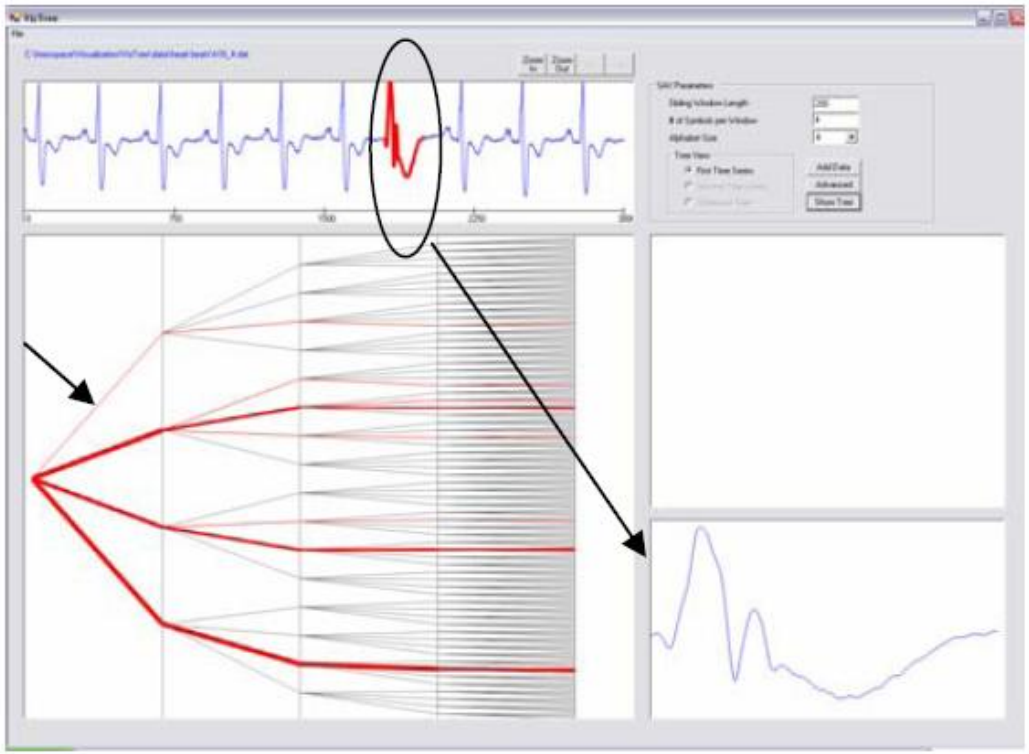
Pattern Detection – VizTree

- VizTree Idea:
  - The set of strings produced by SAX can be encoded as a suffix tree
  - Using a time window of length, 2 cbabbbaaacc becomes {cb, ba, bb, bb, ba, aa, ac, cc}



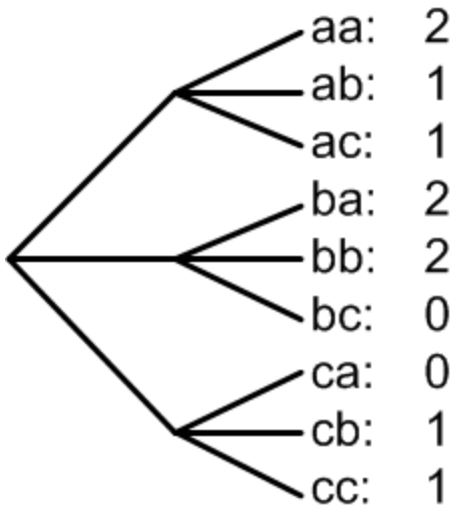
Pattern Detection – VizTree

- Increase edge width paths containing large # of matching sequences
  - Frequent patterns and anomalies are easily recognizable



Pattern Detection – Time Series Bitmaps

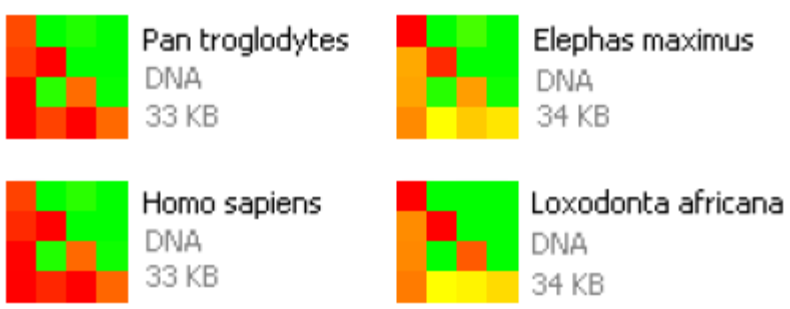
- Instead of using node-link diagrams to represent a suffix tree we can create a treemap
  - encode # of matches as colour of each cell
- Restrict # of cells to a small value (~16)

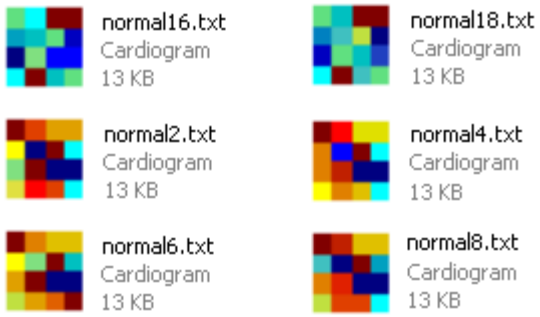


aa	ab	ba	bb
ac		bc	
ca	cb		
cc			

Pattern Detection – Time Series Bitmaps

- Very difficult to interpret what a sequence looks like from the map
  - No good for analyzing an individual time series
- Easy/quick to compare different time series, useful for
  - overviews of many time series
  - spotting clusters & anomalies





Pattern Detection

- Good:
  - Fast method for approximating time series as symbolic strings
  - Easy to see common/uncommon subsequences with suffix trees
  - Easy to compare multiple time series with bitmaps
- Bad:
  - unclear how to determine key parameters; (1) length of sliding window, (2) # of intervals to use, (3) alphabet size