

Interaction  
Visualization  
Usability + UX  
laboratory

UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

## Visualizzazione di dati orientati al tempo

Paolo Buono  
paolo.buono@uniba.it  
ivu.di.uniba.it/people/buono.htm

## USAGE AGREEMENT

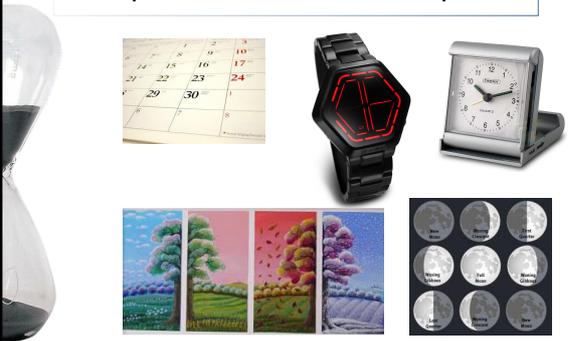
BY USING THESE SLIDES OR PARTS OF THEM YOU INDICATE YOUR ACCEPTANCE OF THE FOLLOWING AGREEMENT, UNDERSTAND THE CONDITIONS, AND AGREE TO BE LEGALLY BOUND BY THEM.

I understand that the slides contain copyright protected material and therefore the following conditions of use apply:

- The slides may only be shared as PDF files instead of the original PowerPoint files
- The access to the slides must be securely restricted to the **attending students** (e.g., using password authentication)
- The students cannot share or redistribute the slides

Visualizzazione di serie temporali

## Il tempo ha una struttura complessa



Visualizzazione di serie temporali

## Alcune considerazioni

- Non si può (ancora) tornare indietro nel tempo
- Si può attendere il futuro per renderlo presente
- Non percepiamo direttamente il tempo
- Con le visualizzazioni si rende visibile l'invisibile
- Si deve tener conto delle granularità
- Ci sono caratteristiche particolari da tener conto (illustrate in seguito)

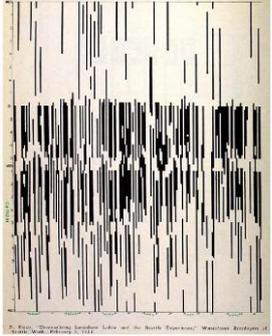
Visualizzazione di serie temporali

## Alcune visualizzazioni classiche

- In aggiunta a quelle viste per l'infovis

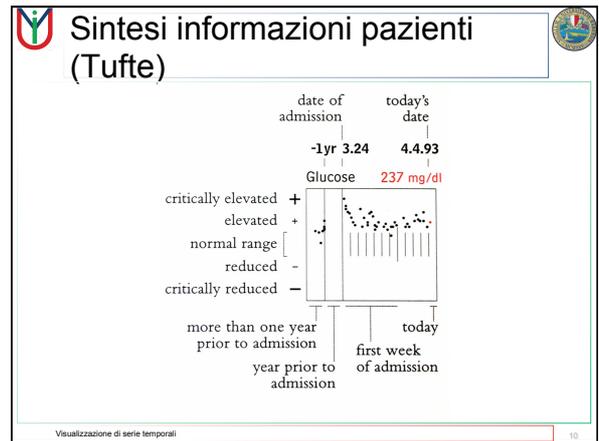
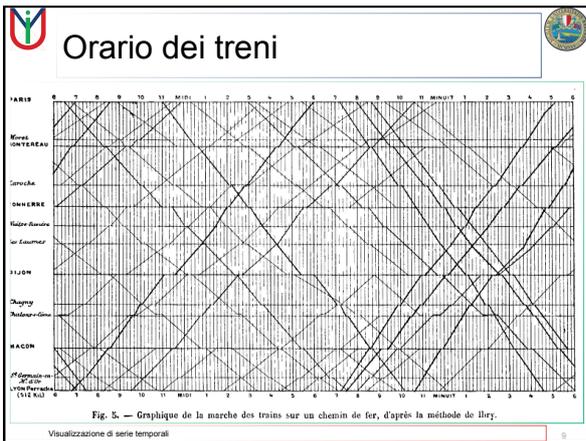
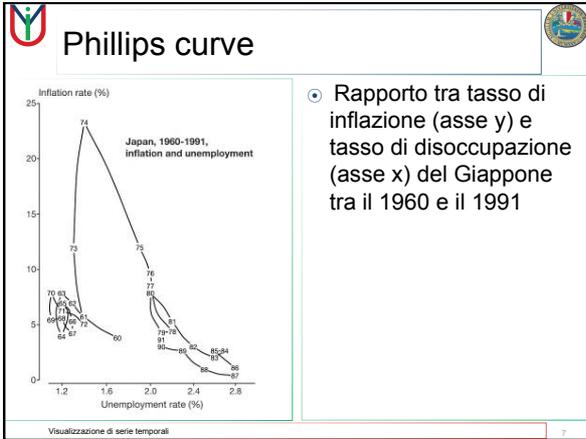
Visualizzazione di serie temporali

## Ore lavorate



**Fig. 2.14** Exact hours and days worked in 1929 by an employee at the Oregon ports. Days are mapped on the horizontal axis and hours per day worked are represented as bars on the vertical axis. The representation shows extreme irregularities in working hours.  
*Source: Brinton (1939, p. 250).*

Visualizzazione di serie temporali



### Analisi dei dati orientati al tempo

Visualizzazione di serie temporali 13

### Modellazione del tempo

scale	ordinal	discrete	continuous
scope	point-based	interval-based	
arrangement	linear	cyclic	
viewpoint	ordered	branching	multiple perspectives

Visualizzazione di serie temporali 14

### Modellazione del tempo

Abstractions

granularity & calendars	none	single	multiple
time primitives	instant	interval	span
determinacy	determinate	indeterminate	

Visualizzazione di serie temporali 15

### Esempio di paradosso della granularità

Relationship of A and B:

Years	2008	2009	
Weeks	Week 1		
Days	1/25	1/26	1/27

A not equals B  
A equals B  
A not equals B

Visualizzazione di serie temporali 16

### Modellazione del tempo

scale	quantitative	qualitative
frame of reference	abstract	spatial
kind of data	events	states
number of variables	univariate	multivariate

Visualizzazione di serie temporali 17

### Analisi dei dati orientati al tempo

Visualizzazione di serie temporali 18

## Visualizzazione del tempo

- Tempo -> Tempo (Animazione)
  -
- Tempo -> Spazio
  -

Visualizzazione di serie temporali 19

## TimeRider

Visualizzazione di serie temporali 20

## TimeRider

- Utenti
  - Medici specialistici
  - Medici di ospedale
- Dati
  - Pazienti
  - Dati multivariati, astratti
  - Tempo lineare, istanti
- Task
  - Trova un pattern in dati multivariati nel tempo (gruppi di pazienti)

Visualizzazione di serie temporali 21

## Ricerca tra pazienti: problemi per scatter plot animati

- Campionamento irregolare
  -
- Durata del dato
  -
- Porzioni di tempo considerati diversi su diversi dataset
  -

Visualizzazione di serie temporali 22

## Visualizzazione Tempo->Spazio

Visualizzazione di serie temporali 23

## Variabili visuali

- Posizione, lunghezza, angolo, inclinazione, connessione, spessore, ...

Visualizzazione di serie temporali 24

### Analisi dei dati orientati al tempo

Caratterizzazione di tempo e di dati orientati al tempo

Visualizzazione dei dati orientati al tempo

Interazione con il tempo

Analisi di dati orientati al tempo

Visualizzazione di serie temporali

### Interazione con il tempo

- Strumenti di visualizzazione di serie temporali, si vedranno dopo

Visualizzazione di serie temporali

### Analisi dei dati orientati al tempo

Caratterizzazione di tempo e di dati orientati al tempo

Visualizzazione dei dati orientati al tempo

Interazione con il tempo

Analisi di dati orientati al tempo

Visualizzazione di serie temporali

### Analisi orientata al tempo

- Astrazione dati temporali
- Statistica
- Data Mining temporale

Temporal data mining tasks

prediction clustering classification

search & retrieval pattern discovery

Visualizzazione di serie temporali

### Serie temporali

Una serie temporale è una sequenza di numeri (tipicamente reali) rappresentanti le misurazioni di una variabile nel tempo

- Temperatura giornaliera
- EEG data
- Andamento delle scorte di magazzino
- Andamento del prezzo di un prodotto

Teoricamente possiamo trovare le serie temporali in qualunque ambito: medico, scientifico, commerciale, ecc.

Visualizzazione di serie temporali

### Visualizzare per problem solving

Visualizzazione di serie temporali

### Le serie temporali sono ovunque! 1

Si misurano molte cose...

- Il grado di popolarità di George Bush
- La pressione del sangue
- La piovosità annuale di Bari
- Il valore delle azioni Parmalat...
- Il numero di click al secondo in un sito web

... e queste cose cambiano nel tempo.

Quindi le serie temporali sono presenti virtualmente in ogni dominio, medico, economico, scientifico.

Visualizzazione di serie temporali

### Le serie temporali sono ovunque! 2

In un campione di 4,000 grafici presi da 15 giornali mondiali, pubblicati dal 1974 al 1989 si è trovato che il 75% di tutti i grafici rappresentavano serie temporali (Tuftte, 1983)

Visualizzazione di serie temporali

### Analisi immagini

Visualizzazione di serie temporali

### Analisi video

Visualizzazione di serie temporali

### Analisi dati biometrici (manoscritti)

George Washington manuscript

Visualizzazione di serie temporali

### Alcuni settori di ricerca

- Bioinformatics:** Aach, J. and Church, G. (2001). Aligning gene expression time series with time warping algorithms. *Bioinformatics*. Volume 17, pp 495-508.
- Robotics:** Schmill, M., Oates, T. & Cohen, P. (1999). Learned models for continuous planning. In *7th International Workshop on Artificial Intelligence and Statistics*.
- Medicine:** Caiani, E.G., et al. (1998) Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume. *IEEE Computers in Cardiology*.
- Chemistry:** Gollmer, K., & Posten, C. (1995) Detection of distorted pattern using dynamic time warping algorithm and application for supervision of bioprocesses. *IFAC CHEMPAS-4*
- Gesture Recognition:** Gavrilu, D. M. & Davis, L. S. (1995) Towards 3-d model-based tracking and recognition of human movement: a multi-view approach. In *IEEE IWAFGR*
- Meteorology/ Tracking/ Biometrics / Astronomy / Finance / Manufacturing ...**

Visualizzazione di serie temporali

### Perchè sono difficili da gestire? I Parte

Le serie temporali risiedono su DB molto grandi

- ▶ 1 Hour of EKG data: 1 Gigabyte
- ▶ Typical Weblog: 5 Gigabytes per week
- ▶ Space Shuttle Database: 200+ Gigabytes
- ▶ Macho Database: 3 Terabytes, updated with 3 gigabytes a day

La maggior parte dei dati sono sul disco (o su nastro), pertanto abbiamo bisogno di una rappresentazione dei dati per poterli manipolare efficacemente

Visualizzazione di serie temporali 37

### Perchè sono difficili da gestire? II Parte

Ci sono problemi di soggettività

La similitudine dipende dall'utente, dal dominio e dallo scopo per cui si vuole calcolare questa misura. È necessario poter trattare questa soggettività

Visualizzazione di serie temporali 38

### Perchè sono difficili da gestire? III Parte

Ci sono problemi di gestione di dati differenti

- Differenze nel formato
- Differenze nelle frequenze di campionamento
- Disturbo, valori nulli, ecc

Visualizzazione di serie temporali 39

### Cosa vogliamo fare con le serie temporali?

Visualizzazione di serie temporali 40

### Hanno in comune il problema della similarità

Visualizzazione di serie temporali 41

### Motivazione

Un paziente va dal medico a causa di un dolore al petto. L'elettrocardiogramma sembra insolito. Il medico vorrebbe cercare un comportamento **simile** nel database nella speranza di avere indizi circa la condizione del paziente

Due domande:

- Cos'è la similarità?
- Come cerchiamo velocemente?

Visualizzazione di serie temporali 42

## Cos'è la similarità?

The quality or state of being similar; likeness; resemblance; as, a similarity of features.



**Webster's Dictionary**  
 È difficile definire la similarità ma "la riconosciamo quando la vediamo"  
 Il significato reale di similarità è un fatto filosofico  
 La scienza usa un approccio pragmatico

Visualizzazione di serie temporali 43

## Tipi di analisi

- Ricerca di similarità
  - whole matching
  - subsequence matching
- Analisi di trend
- Periodicità
- Predizione
- Associazione
- Ricerca di pattern
- Clustering

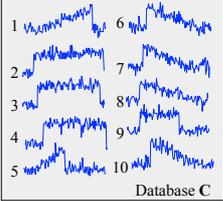
Visualizzazione di serie temporali 44

## Whole Matching

Query Q



Dato un database C, una query Q e una funzione per la misura della distanza, trovare la serie C più simile a Q.



Database C



C<sub>6</sub> è la serie più simile.

Visualizzazione di serie temporali 45

## Subsequence Matching

Query Q



Sequenza C



Sottosequenza più simile

Data una sequenza C, una query Q e una funzione per la misura della distanza, trovare le posizioni in cui si ha il migliore match.

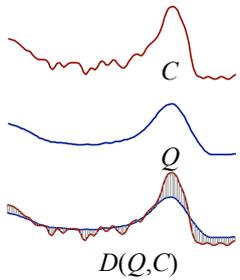
Visualizzazione di serie temporali 46

## Misura della distanza

**Definizione:** Siano O<sub>1</sub> e O<sub>2</sub> due oggetti dell'insieme di tutti gli oggetti. La distanza (dissimilarità) è denotata come D(O<sub>1</sub>, O<sub>2</sub>)

**Distanza Euclidea**

Date due serie temporali  
 $C = c_1 \dots c_n$   
 e  
 $Q = q_1 \dots q_n$   
 la loro distanza Euclidea è definita come:

$$D(Q, C) = \sqrt{\sum_{i=1}^n (q_i - c_i)^2}$$


Visualizzazione di serie temporali 47

## Distanza Euclidea - proprietà

- ✓ Semplice da calcolare
- ✓ Ottimale quando gli errori nella serie sono additivi (indipendenti, uniformemente distribuiti)
- ✓ Stima bene la dissimilarità reale
- ✓ Consente soluzioni scalabili ad altri problemi (es: clustering)

⚠ Non supporta differenti baseline:  
 es: il prodotto X oscilla intorno a 100 €, il prodotto Y intorno a 30 €

⚠ Non supporta scale differenti:  
 es: il prodotto X oscilla fra 60 € e 120 €, il prodotto Y fra 20 € e 40 €

Visualizzazione di serie temporali 48

### Distanza pesata

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

$$D(Q, C, W) \equiv \sqrt{\sum_{i=1}^n w_i (q_i - c_i)^2}$$

L'altezza dell'istogramma indica l'importanza relativa di ogni punto della query

Visualizzazione di serie temporali

### Preprocessing dei dati

Se tentiamo di misurare la distanza fra due serie temporali "grezze", rischiamo di ottenere dei risultati ben lontani dalla realtà.

Ciò dipende dal fatto che la distanza Euclidea è molto sensibile ad alcune distorsioni nei dati. Per la maggior parte dei problemi queste distorsioni non sono significative per cui è possibile eliminarle.

Le quattro distorsioni più comuni:

- Offset Translation
- Amplitude Scaling
- Linear Trend
- Noise reduction

Visualizzazione di serie temporali

### Spostamento origine (Offset translation)

$$Q = Q - \text{mean}(Q)$$

$$C = C - \text{mean}(C)$$

$$D(Q, C)$$

Visualizzazione di serie temporali

### Normalizzazione ampiezza (Amplitude scaling)

$$Q = (Q - \text{mean}(Q)) / \text{std}(Q)$$

$$C = (C - \text{mean}(C)) / \text{std}(C)$$

$$D(Q, C)$$

Visualizzazione di serie temporali

### Andamento lineare (linear trend)

Trovare la retta che meglio interseca le serie temporali e sottrarla dalle serie

Rimozione di offset translation  
Rimozione di amplitudine scaling

Rimozione di linear trend  
Rimozione di offset translation  
Rimozione di amplitudine scaling

Visualizzazione di serie temporali

### Riduzione del disturbo (noise reduction)

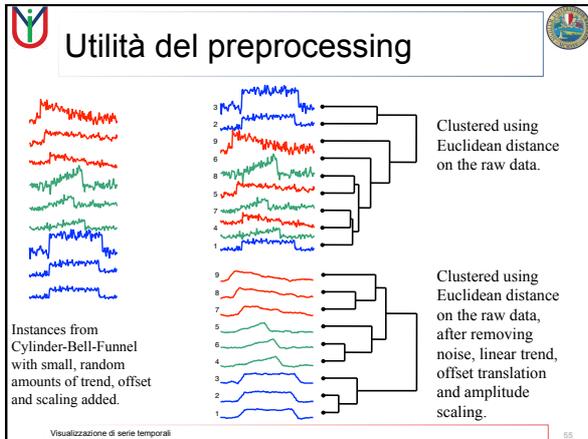
Eliminazione semplice: media di ogni punto con i punti più vicini (media mobile)

$$Q = \text{smooth}(Q)$$

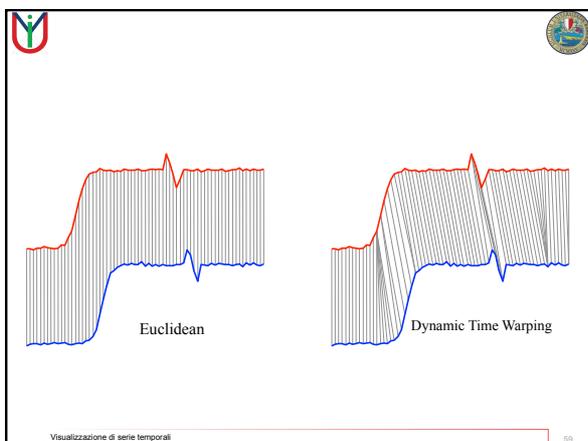
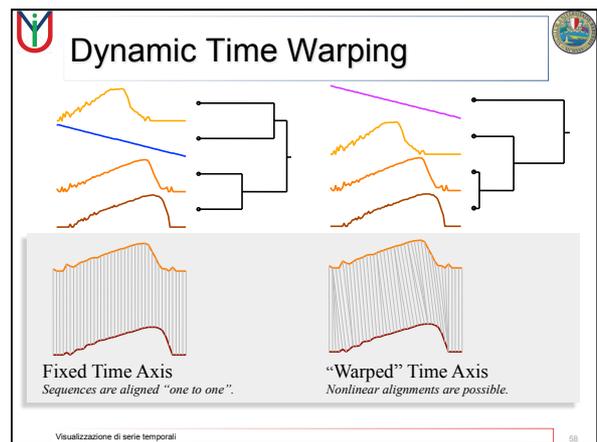
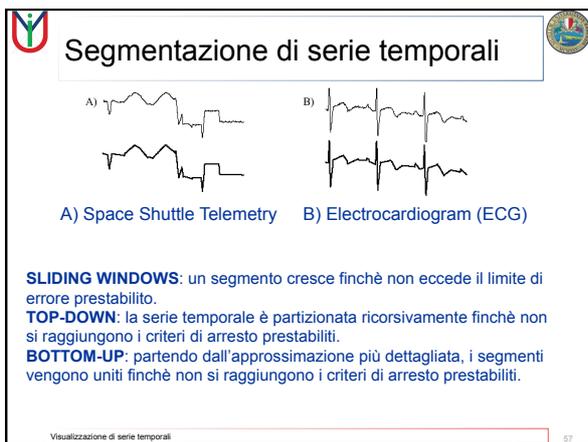
$$C = \text{smooth}(C)$$

$$D(Q, C)$$

Visualizzazione di serie temporali



- ### In breve...
- I dati originali potrebbero avere distorsioni che si possono rimuovere prima di applicare il clustering, classificazioni, ecc..
  - A volte però, le distorsioni sono le cose più interessanti
  - Alcune rappresentazioni di alto livello dei dati possono tenere conto anche delle trasformazioni
- Visualizzazione di serie temporali



### Results: Error Rate

DataSet	Euclidean	DTW
Word Spotting	4.78	1.10
Sign language	28.70	25.93
GUN	5.50	1.00
Nuclear Trace	11.00	0.00

We use 1-nearest-neighbor, leaving one out evaluation

Visualizzazione di serie temporali

### Results: Time (msec)

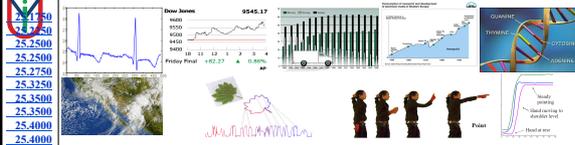
DataSet	Euclidean	DTW	
Word Spotting	40	8,600	215
Sign language	10	1,110	110
GUN	60	11,820	197
Nuclear Trace	210	144,470	687

DTW is hundred to thousands of times slower than Euclidean distance.

Visualizzazione di serie temporali 61



## Time series tools



- Time series: a **sequence** of real numbers, representing the measurements (observations) of a **real variable** at **equal time intervals**
- TS may be found in almost all the real life domains:
  - Weather forecast, stock trend, ECG, DNA sequences
  - Image analysis, video analysis
  - Math, chemistry, physics, economy, ...

Visualizzazione di serie temporali 63

### Pattern search in time series

- Classic example: of a doctor that wants to search a database to find similar ECGs, in order to have clues about a patient condition...
- Similarity implies: non exact match, efficiency, indexing subsequence similarity problem, clustering problem, rule discovery problem
- Properties are: symmetry, constancy of self-similarity, positivity, triangular inequality

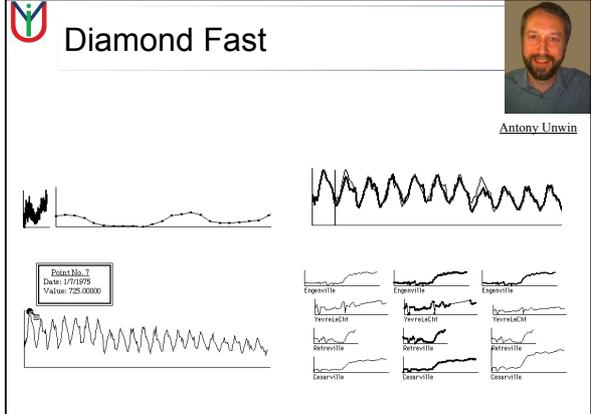
Visualizzazione di serie temporali 64

### Some tasks

- Find trends with a given behavior
- Find companies with similar stock prices over a **time interval**
- Use patterns to **classify** a given time series
- Find similar **subsequence** of an ECG or in DNA seq.
- Find products with similar sell **cycles**

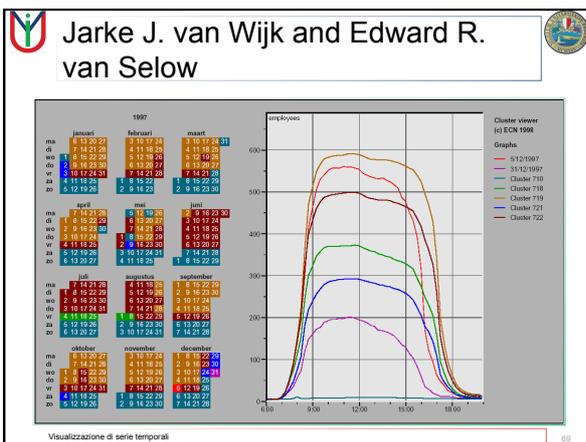
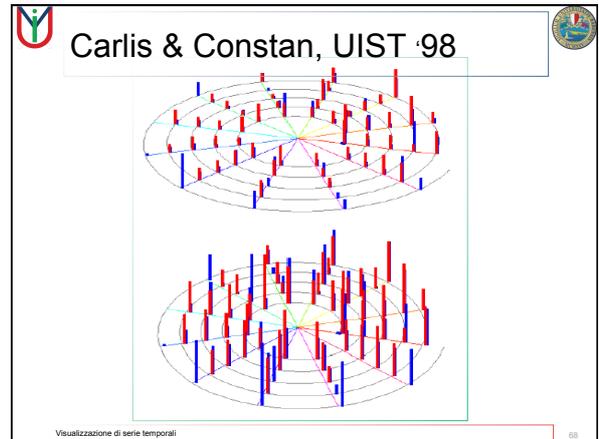
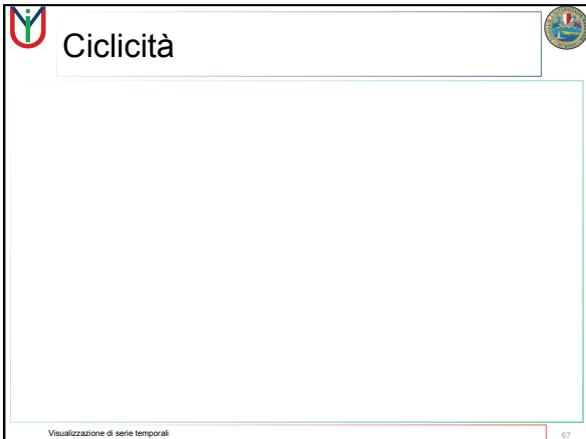
Visualizzazione di serie temporali 65

### Diamond Fast



Antony Unwin

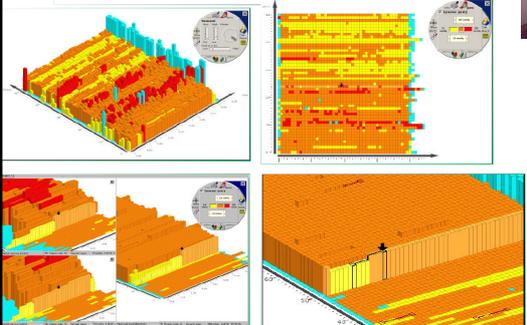
Visualizzazione di serie temporali 66



University of Udine, Italy



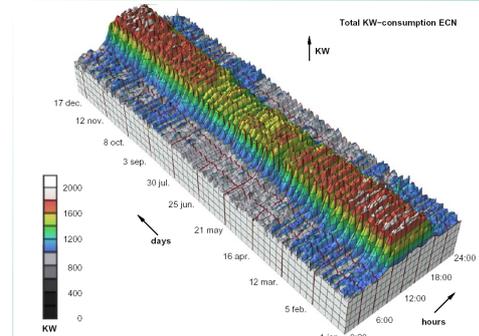
Luca Chittaro



Visualizzazione di serie temporali

73

Keller, P.R. and Keller, M.M.



Total KW-consumption ECN

↑ KW

days

hours

2000  
1600  
1200  
800  
400  
0

17 dec.  
12 nov.  
8 oct.  
3 sep.  
30 jul.  
25 jun.  
21 may  
16 apr.  
12 mar.  
5 feb.  
1 jan. 0:00

24:00  
18:00  
12:00  
6:00

Visualizzazione di serie temporali

74

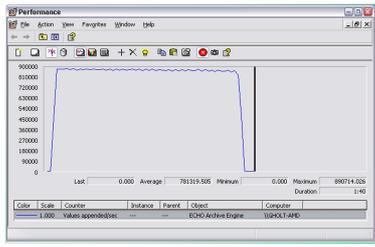
Prestazioni



Visualizzazione di serie temporali

75

ECHO



Running on a dual 1.8 MHz AMD-processor machine, an ECHO historian clocked top speeds in excess of 890,000 values per second writing to a single datastream

Visualizzazione di serie temporali

76

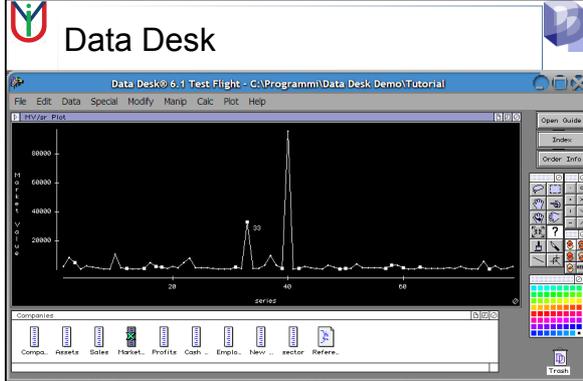
Prediction



Visualizzazione di serie temporali

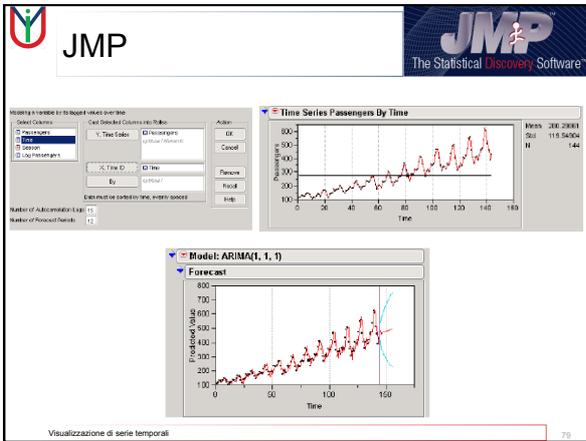
77

Data Desk



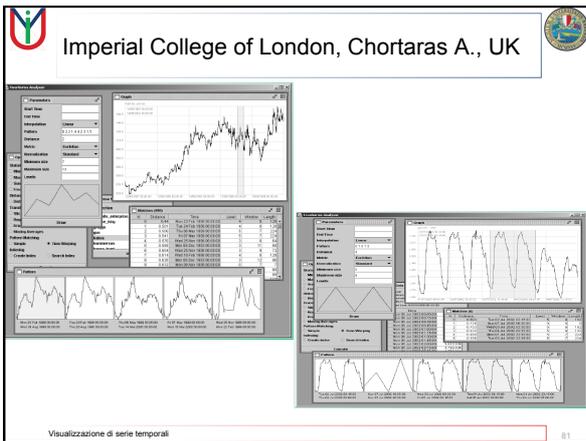
Visualizzazione di serie temporali

78



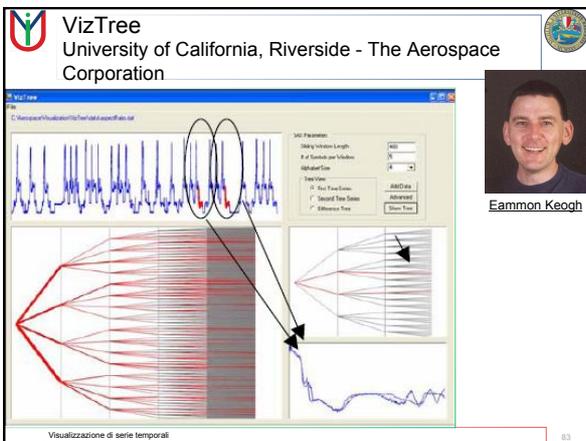
### Pattern specification

Visualizzazione di serie temporali



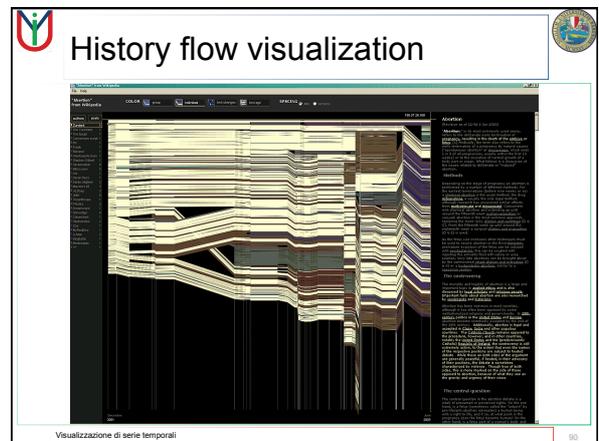
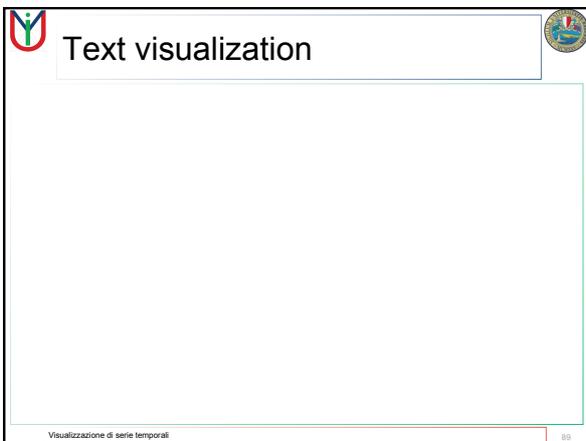
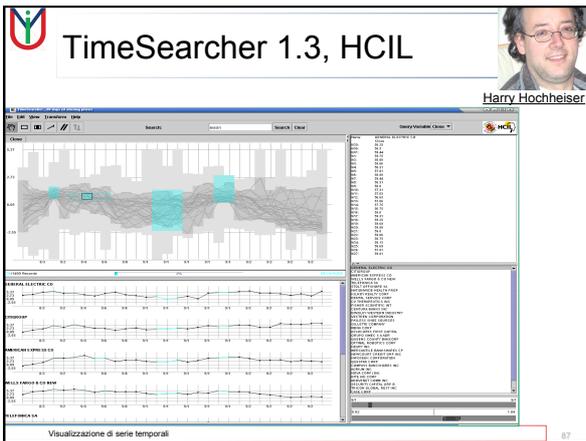
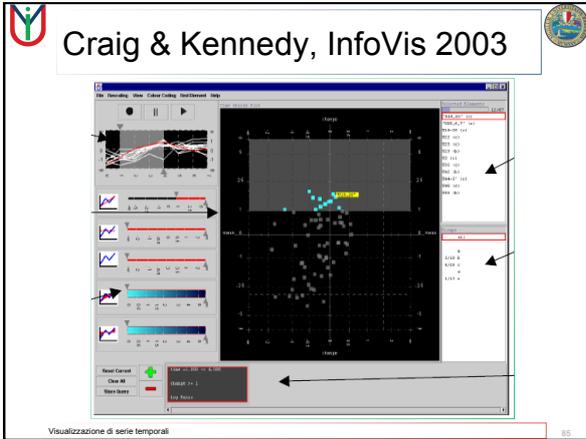
### Anomaly detection

Visualizzazione di serie temporali



### Advanced interaction / direct manipulation

Visualizzazione di serie temporali



### Readability analysis

Visualizzazione di serie temporali

### Visualization of time-oriented data

Visualizzazione di serie temporali

### Interactive Pattern Search

- Select pattern
- Present results
- Adjust tolerance
- Interactive process
  - 1st step: explicit initiation of search  
Find matches
  - 2nd step: interactive implicit initiation of search  
Change the result list by adjusting tolerance

Visualizzazione di serie temporali

### Dataset example

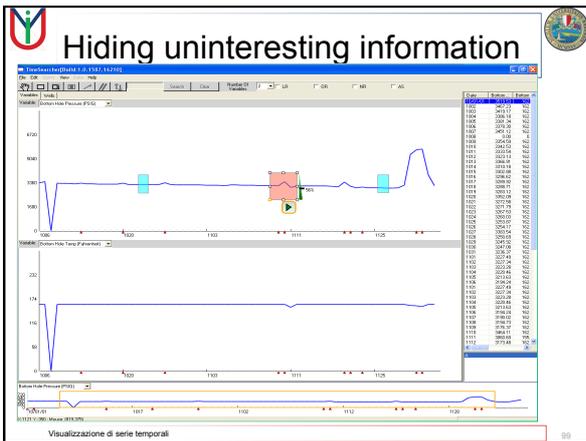
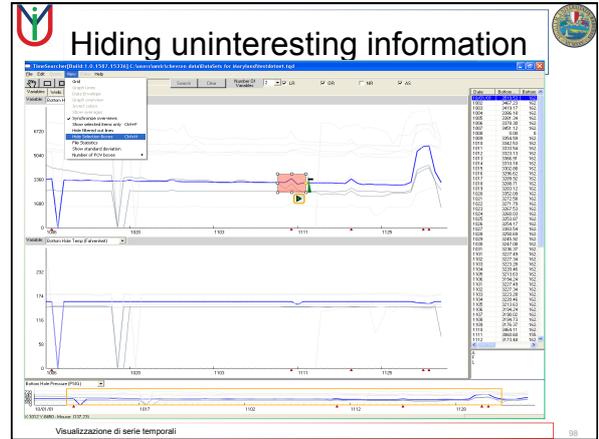
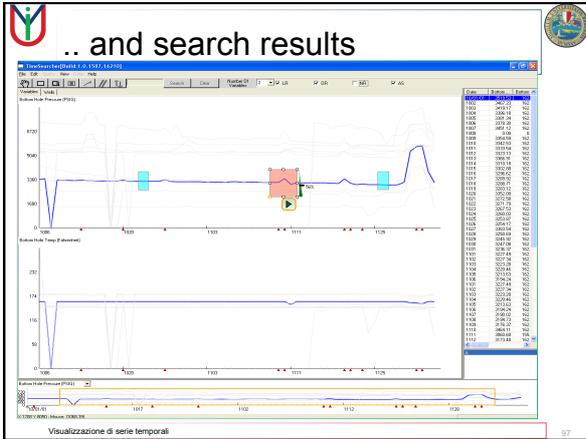
Visualizzazione di serie temporali

### Filter using timebox

Visualizzazione di serie temporali

### ... filter more...

Visualizzazione di serie temporali



### Sample datasets

- TimeSearcher 2.0 has been tested on several dataset:
  - Chevron Texaco data oil production
  - 5 years meteo data in a region of the South of Italy

Visualizzazione di serie temporali

### TimeSearcher 2.0

Visualizzazione di serie temporali

- Oil Production
  - New in 2.0:
    - Multiple variables
    - Overview/details
    - Pattern Search
- Weather
  - 10 variables
  - 15 locations
  - 5 years of data
  - ~286,000 values

### Searchbox tolerance adjustments

Visualizzazione di serie temporali

Direct manipulation adjustments of envelope

Indirect parameter adjustments

- amplitude
- linear trend
- offset
- noise

# Ciclicità

## Serie Temporali

- Sequenze di numeri reali i quali rappresentano le misure o le osservazioni di una variabile reale a intervalli di tempo costanti

- Domini
- scienza
- ingegneria
- affari
- finanza

- Applicazioni
- elettrocardiogrammi
- sismografi
- processi industriali
- rilevazioni meteorologiche

Visualizzazione di serie temporali 104

## Componenti e proprietà

- Componenti elementari
  - Trend
  - Ciclo
  - Variazione stagionale
  - Componente residua
- Proprietà
  - Tipi di dati: interi, reali, immagini
  - Dominio di tempo: rappresentazione ordinale, calendario
  - Estensione di vita: intervallo di punti temporali validi
  - Regolarità: valore per ogni punto temporale

Visualizzazione di serie temporali 105

## Dati Seriali Periodici

*Che sia generato dalla natura, o imposto artificialmente dall'uomo, l'andamento periodico è presente ovunque nella vita reale*

- I dati di serie temporali possono esibire comportamenti periodici
- L'importanza dello studio delle periodicità risiede nella possibilità di effettuare stime migliori sui dati in possesso (analisi di trend, anomalie, comportamenti tipici, fenomeni isolati)

Visualizzazione di serie temporali 106

## Tecniche di Visualizzazione di Dati Ciclici

- Rappresentazione con struttura *data-grid* [Mosolov, Bederson]
  - Dati disposti uniformemente in una struttura a griglia, su righe e colonne
  - Esplorazione interattiva variando il numero di punti in ogni riga e modificando quindi anche il periodo
  - Via via che periodo visualizzato si avvicina alla reale periodicità presente nei dati, emerge un pattern diagonale molto inclinato; quando tale periodo coincide esattamente con la periodicità dei dati esplorati, emerge un pattern verticale

Visualizzazione delle informazioni sulla quantità di luce osservata nel Gennaio 2000

Visualizzazione di serie temporali 107

## Tecniche di Visualizzazione di Dati Ciclici

- Tecniche di visualizzazione a spirale [Carlis e Konstan]
  - l'**aspetto seriale** dei dati è presente osservando la forma della spirale che si diparte da un punto centrale allontanandosi via via da esso in maniera continua e costante
  - l'**aspetto periodico** è esplorato lungo i raggi
  - generalmente si ricorre alla spirale di Archimede
  - miglioramenti mediante l'uso di effetti 3D o di animazione

*Percentuale di consumo di Buphia Cappariidifolia (1980-1988)*

Visualizzazione di serie temporali 108

# TimeSearcher

**Rappresentazione di serie temporali multivariate**

**Basato sul modello di interrogazione Timebox**

**Overview con possibilità di zoomare all'interno di aree di interesse**

**Uso della SearchBox, per il ritrovamento di pattern simili ad un pattern d'interesse**

## TimeSearcher: Interfaccia

**Variables View** | **Details List** | **Items List** | **Overview**

Visualizzazione di serie temporali

## Descrizione del Prototipo: interfaccia

Visualizzazione di serie temporali

## Visualizzare i valori della serie lungo la spirale

- I valori della serie temporale determinano, per ogni ciclo un angolo di ampiezza proporzionale a tale valore
- Uso del sistema RGB (Red-Green-Blue) che esprime i colori numericamente in termini di tre coordinate, per differenziare visivamente aree e intervalli temporali, associando l'intensità del colore ai valori della serie

Visualizzazione di serie temporali

## Esplorazione interattiva di dati per l'individuazione di ciclicità

### Monitoraggio della qualità dell'aria nella città di Bari

- Osservazioni effettuate ogni ora, relative al mese di Gennaio 2005
- La serie temporale relativa alla quantità di CO presente nell'aria in zona Poggiofranco esibisce un comportamento periodico giornaliero, come risulta osservando il pattern strutturato in Fig. 4
- La quantità di CO rilevata è proporzionale all'intensità del colore con cui sono rappresentati gli intervalli temporali

Visualizzazione di serie temporali

## Analisi di dati ciclici

*Dato non pervenuto*

*Anomali, dati irregolari, fenomeni isolati*

*Quantità di CO rilevata dalle 18.00 alle 21.00 (ore di punta)*

*31 Gen 2005*

*1 Gen 2005*

*Quantità di CO rilevata dalle 7.00 alle 11.00 (ore di punta)*

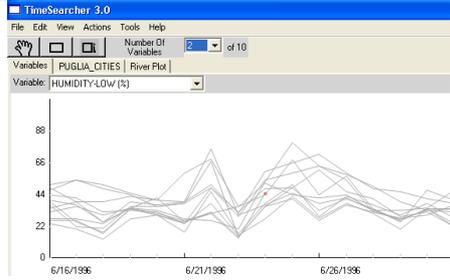
*Quantità di CO rilevata dalle 2.00 alle 3.00*

Visualizzazione di serie temporali

## Ricerca del pattern

### Analisi del problema

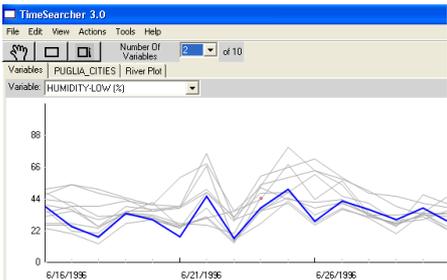
- Nella versione precedente di TimeSearcher la selezione di un pattern particolare e la ricerca di pattern simili avvengono come segue:



Visualizzazione di serie temporali

### Analisi del problema

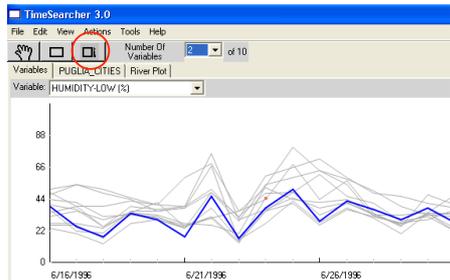
- Selezione della serie temporale da analizzare



Visualizzazione di serie temporali

### Analisi del problema

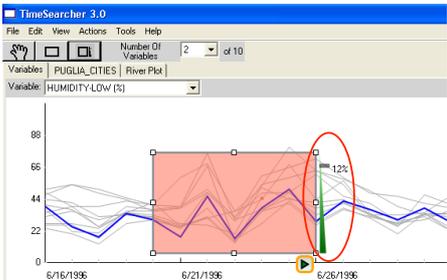
- SearchBox su un particolare pattern di interesse



Visualizzazione di serie temporali

### Analisi del problema

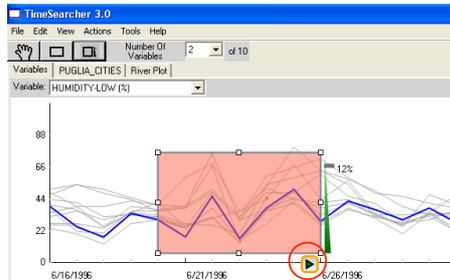
- Impostazione del valore di tolleranza



Visualizzazione di serie temporali

### Analisi del problema

- Avvio del processo di ricerca attraverso il tasto start



Visualizzazione di serie temporali

## Analisi del problema

### 5. Visualizzazione dei risultati ottenuti

Visualizzazione di serie temporali

## La nuova SearchBox

- La nuova SearchBox è caratterizzata da:
  - Visualizzazione dei punti temporali che costituiscono il pattern selezionato
  - Nuovi *bottoni* per modificare la posizione di tali punti lungo l'asse delle ordinate
  - Visualizzazione dei punti rilevanti in termini di percezione
  - Impostazioni delle opzioni di ricerca

Visualizzazione di serie temporali

## La nuova SearchBox

### ■ *Spostamento della posizione di un punto temporale*

Visualizzazione di serie temporali

## La nuova SearchBox

### ■ *Perceptually Important Points*

Visualizzazione di serie temporali

## La nuova SearchBox

### ■ *Impostazioni delle opzioni di ricerca*

Visualizzazione di serie temporali

# Aggregazione

## Aggregazione dei dati

- Processo attraverso il quale una collezione di dati o oggetti è ridotta ad un singolo oggetto (aggregato) che rappresenta la collezione
- Aggregato: media, massimo, minimo, ecc...
- Metodo tradizionale
 
$$X \quad \dots, X_i, X_{i+1}, \dots, X_{i+k}, X_{i+k+1}, X_{i+k+2}, \dots, X_{i+2k}, \dots$$

$$Y \quad \underbrace{\quad \quad \quad}_{y_i} \quad \underbrace{\quad \quad \quad}_{y_{i+1}}$$
- Metodo basato su processo ARIMA con medie mobili
 
$$X \quad \dots, X_i, X_{i+1}, X_{i+2}, \dots, X_{i+k}, X_{i+k+1}, \dots$$

$$Y \quad \underbrace{\quad \quad \quad}_{y_i} \quad \underbrace{\quad \quad \quad}_{y_{i+1}}$$

Visualizzazione di serie temporali

## TimeSearcher – Interfaccia

- Tre modalità di visualizzazione:
  - Variable View
  - Visualizzazione degli item singolarmente
  - River Plot

Visualizzazione di serie temporali

## TimeSearcher - Variable View

Visualizzazione di serie temporali

## TimeSearcher - Item

Visualizzazione di serie temporali

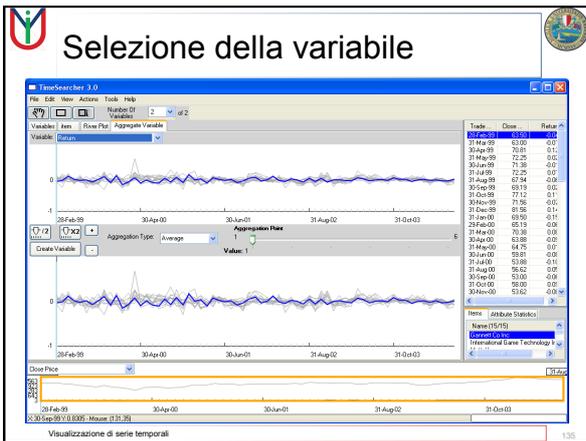
## TimeSearcher - River Plot

Visualizzazione di serie temporali

## Analisi del problema

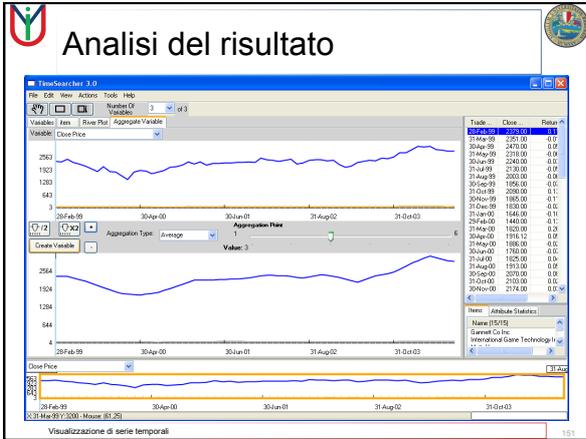
- Creazione di un' interfaccia per effettuare operazioni di aggregazione
- Task principali:
  - Selezione della variabile
  - Selezione del numero di punti di aggregazione
  - Selezione del tipo di aggregazione
  - Esecuzione dell' aggregazione
  - Analisi del risultato

Visualizzazione di serie temporali









## Template e import/export

### TQD & CSV

- Il formato proprietario di TimeSearcher 3 è il TQD, *Temporal Query Data File*, che permette di leggere e rappresentare i dati relativi alle serie temporali in esame
- Il formato CSV, *Comma Separated Values*, è riconosciuto da TimeSearcher 3.2 al fine di importare e convertire i dati in file .tqd

Visualizzazione di serie temporali

### TQD & CSV

- Esempio PrezziSupermercati.tqd

```
#Titolo
Prezzi delle mele e delle pere
#Attributi statici
Data,String
#Attributi Dinamici
Mele (euro/Kg),Float
#Numero di punti temporali
10
#Numero di elementi
3
##missingValue -1.00
#Punti temporali
01/06/2009 00:00:00,01/07/2009 00:00:00,01/08/2009 00:00:00,01/09/2009 00:00:00,01/10/2009
00:00:00,01/11/2009 00:00:00,01/12/2009 00:00:00,01/01/2010 00:00:00,01/02/2010
00:00:00,01/03/2010 00:00:00
#Dati per ciascun elemento
Auchan,
2.45,3.18,2.52,3.20,1.93,2.12,1.67,2.01,1.75,2.10,1.84,1.77,1.86,1.62,1.94,1.65,2.04,1.91,2.10
,2.00
Conad,
2.93,3.40,2.46,3.05,1.87,2.10,1.62,1.98,1.67,2.12,1.81,1.72,1.84,1.56,1.90,1.58,1.98,1.88,2.15
,2.01
Coop,
2.72,3.29,2.64,3.10,1.84,2.17,1.72,2.04,1.86,2.15,1.90,1.80,1.92,1.63,1.95,1.63,1.95,1.87,2.11
,2.02
```

Visualizzazione di serie temporali

## TQD & CSV

- Esempio PrezziSupermercati.csv

```
,Auchan,,Coop,,Conad,
Data, Mele (euro/Kg), Pere (euro/Kg), Mele (euro/Kg), Pere (euro/Kg),
Mele (euro/Kg), Pere (euro/Kg)
01/06/2009, 2.65, 3.18, 2.93, 3.4, 2.72, 3.29
01/07/2009, 2.52, 3.2, 2.46, 3.05, 2.64, 3.1
01/08/2009, 1.93, 2.12, 1.87, 2.1, 1.84, 2.17
01/09/2009, 1.67, 2.01, 1.62, 1.98, 1.72, 2.04
01/10/2009, 1.75, 2.1, 1.67, 2.12, 1.86, 2.15
01/11/2009, 1.84, 1.77, 1.81, 1.72, 1.9, 1.8
01/12/2009, 1.86, 1.62, 1.84, 1.56, 1.92, 1.63
01/01/2010, 1.94, 1.65, 1.9, 1.58, 1.95, 1.63
01/02/2010, 2.04, 1.91, 1.98, 1.88, 1.95, 1.87
01/03/2010, 2.1, 2, 2.15, 2.01, 2.11, 2.02
```

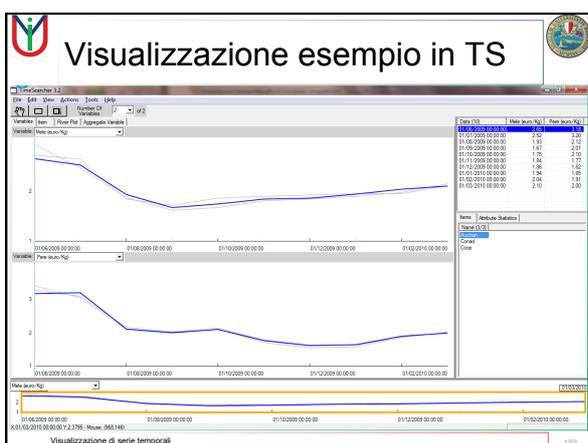
Visualizzazione di serie temporali

## TQD & CSV

- Esempio: PrezziSupermercati.csv

	A	B	C	D	E	F	G
1		Auchan		Coop		Conad	
2	Data	Mele (euro/Kg)	Pere (euro/Kg)	Mele (euro/Kg)	Pere (euro/Kg)	Mele (euro/Kg)	Pere (euro/Kg)
3	01/06/2009	2.65	3.18	2.93	3.4	2.72	3.29
4	01/07/2009	2.52	3.2	2.46	3.05	2.64	3.1
5	01/08/2009	1.93	2.12	1.87	2.1	1.84	2.17
6	01/09/2009	1.67	2.01	1.62	1.98	1.72	2.04
7	01/10/2009	1.75	2.1	1.67	2.12	1.86	2.15
8	01/11/2009	1.84	1.77	1.81	1.72	1.9	1.8
9	01/12/2009	1.86	1.62	1.84	1.56	1.92	1.63
10	01/01/2010	1.94	1.65	1.9	1.58	1.95	1.63
11	01/02/2010	2.04	1.91	1.98	1.88	1.95	1.87
12	01/03/2010	2.1	2	2.15	2.01	2.11	2.02
13							

Visualizzazione di serie temporali



## Analisi del problema

- Obiettivi:
  - Manipolare i dati della *detail list* e salvarli
  - Esportazione dei dati nel formato *file csv* nelle seguenti modalità: normale, per variabile e per elementi.
  - Miglioramento dell' importazione dei dati in TimeSearcher 3.2.
- Metodi applicati per la valutazione dell' usabilità delle nuove funzionalità
  - Interviste
  - Metodi di ispezione

Visualizzazione di serie temporali

## Manipolare i dati della *detail list*

- Funzionalità implementate
  - Cancellazione della colonna relativa ad una variabile. Questa funzionalità è nata dall' esigenza di eliminare dalla, 'detail list', variabili inutili ai fini dell' analisi del fenomeno.
  - Modifica del valore di un elemento della colonna relativa ad una variabile. Sebbene i dati non dovrebbero essere modificati, talvolta è palese per l' analista che un dato non è valido. L' analista potrebbe calcolare il dato mancante o settare un valore particolare che aiuta l' analisi
  - Salvataggio di eventuali modifiche dei dati. Nel caso in cui si effettuino modifiche nei dati si rende necessario effettuare il salvataggio dei nuovi dati, per permettere analisi future

Visualizzazione di serie temporali

## Manipolare i dati della *detail list*

- Menu contestuale sulla *detail list*

datetime (288)	>0.30	>0.32	>0.35	>0.38
10/1/2009 10:14:00	272648300.00	220533900.00	169021600.00	119006400.00
10/1/2009 10:19:00	285805600.00	232679700.00	179712100.00	127483700.00
10/1/2009 10:24:00	303451700.00	247280100.00	191912500.00	137453200.00
10/1/2009 10:29:00	295588900.00	240534500.00	186558000.00	132972300.00
10/1/2009 10:34:00	288318300.00	233833600.00	179983300.00	127323700.00
10/1/2009 10:39:00	272648300.00	220533900.00	169021600.00	119006400.00
10/1/2009 10:44:00	269872200.00	217745	Edit	00.00
10/1/2009 10:49:00	287253500.00	232519	Delete Column	00.00
10/1/2009 10:54:00	269700100.00	218411		00.00
10/1/2009 10:59:00	287303000.00	232919000.00	179552900.00	127048400.00
10/1/2009 11:04:00	255246100.00	206151100.00	157638800.00	110493600.00
10/1/2009 11:09:00	254629600.00	205211800.00	156936000.00	109810700.00
10/1/2009 11:14:00	235042700.00	183764700.00	143402400.00	99625850.00
10/1/2009 11:19:00	237612000.00	191085600.00	145591600.00	101811700.00
10/1/2009 11:24:00	237612000.00	191085600.00	145591600.00	101811700.00

Visualizzazione di serie temporali

### Manipolare i dati della *detail list*

- Modifica del dato
- Eliminazione della colonna
- Salvataggio delle modifiche

Visualizzazione di serie temporali

### Esportazione dei dati nel formato file csv

- Esportazione i dati nel formato *file csv* nelle seguenti modalità: normale, per variabile e per elementi

Visualizzazione di serie temporali

### Esportazione dei dati nel formato file csv

- Ex\_PrezziSupermercati.csv: risultato dell'esportazione normale in formato csv del file PrezziSupermercati.tqd

Time Point	A	B	C	D	E	F	G
1		Mele (euro/kg)-Pere (euro/kg)-Auchan	Mele (euro/kg)-Pere (euro/kg)-Auchan	Mele (euro/kg)-Pere (euro/kg)-Conad	Mele (euro/kg)-Pere (euro/kg)-Conad	Mele (euro/kg)-Pere (euro/kg)-Coop	Mele (euro/kg)-Pere (euro/kg)-Coop
2	01/06/2009 00:00	2.65	3.18	2.93	3.4	2.72	3.29
3	01/07/2009 00:00	2.52	3.2	2.46	3.05	2.64	3.1
4	01/08/2009 00:00	1.93	2.12	1.87	2.1	1.84	2.17
5	01/09/2009 00:00	1.67	2.01	1.62	1.98	1.72	2.04
6	01/10/2009 00:00	1.75	2.1	1.67	2.12	1.86	2.15
7	01/11/2009 00:00	1.84	1.77	1.81	1.72	1.9	1.8
8	01/12/2009 00:00	1.86	1.62	1.84	1.56	1.92	1.63
9	01/01/2010 00:00	1.94	1.65	1.9	1.58	1.95	1.63
10	01/02/2010 00:00	2.04	1.91	1.98	1.88	1.95	1.87
11	01/03/2010 00:00	2.1	2	2.15	2.01	2.11	2.02

Visualizzazione di serie temporali

### Esportazione dei dati nel formato file csv

- Auchan.csv: uno dei file csv ottenuti dall'esportazione per item del file PrezziSupermercati.tqd

Time Point	A	B	C
1	Time Point	Mele (euro/kg)	Pere (euro/kg)
2	01/06/2009 00:00	2.65	3.18
3	01/07/2009 00:00	2.52	3.2
4	01/08/2009 00:00	1.93	2.12
5	01/09/2009 00:00	1.67	2.01
6	01/10/2009 00:00	1.75	2.1
7	01/11/2009 00:00	1.84	1.77
8	01/12/2009 00:00	1.86	1.62
9	01/01/2010 00:00	1.94	1.65
10	01/02/2010 00:00	2.04	1.91
11	01/03/2010 00:00	2.1	2

Visualizzazione di serie temporali

### Esportazione dei dati nel formato file csv

- Mele.csv: uno dei file csv ottenuti dall'esportazione per variabile del file PrezziSupermercati.tqd

Time Point	A	B	C	D
1		Auchan	Conad	Coop
2	01/06/2009 00:00	2.72	2.72	2.72
3	01/07/2009 00:00	2.64	2.64	2.64
4	01/08/2009 00:00	1.84	1.84	1.84
5	01/09/2009 00:00	1.72	1.72	1.72
6	01/10/2009 00:00	1.86	1.86	1.86
7	01/11/2009 00:00	1.9	1.9	1.9
8	01/12/2009 00:00	1.92	1.92	1.92
9	01/01/2010 00:00	1.95	1.95	1.95
10	01/02/2010 00:00	1.95	1.95	1.95
11	01/03/2010 00:00	2.11	2.11	2.11

Visualizzazione di serie temporali

### Miglioramento dell'importazione dei dati in TimeSearcher

- Importazione per tipi frequenti di file con Template
- Velocizzazione dei tempi di importazione

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

- Importazione per tipi frequenti di file con Template

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

- Importazione per tipi frequenti di file con Template

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

- Controllo sui dati mancanti

	A	B	C	D	E	F	G
1	datetime	>0.30	>0.32	>0.35	>0.38	>0.41	>0.45
2	01/10/2009 00:01	357.928.884	290.564.923	223.755.590	157.397.012	97.946.888	53.742.144
3	01/10/2009 00:09	363.307.877	295.624.076	228.293.179	160.790.208	99.869.107	54.641.097
4	01/10/2009 00:18	365.493.692	298.208.154	222.518.268	156.553.107	96.986.394	52.929.966
5	01/10/2009 00:24	351.047.061	284.710.712	218.471.415	153.162.836	94.597.243	51.153.714
6	01/10/2009 00:29	350.750.296	284.139.767	217.779.039	152.152.187	93.524.027	50.907.780
7	01/10/2009 00:34	346.436.865	280.522.059	215.006.765	150.539.543	92.371.085	49.742.288
8	01/10/2009 00:39	346.802.468	280.471.372	214.698.963	149.986.684	91.977.562	49.666.934
9	01/10/2009 00:44	350.638.040	284.364.233	218.871.149	153.735.498	94.733.524	51.270.945
10	01/10/2009 00:49	347.611.821	281.410.379	215.460.244	150.786.949	92.456.751	49.890.532
11	01/10/2009 00:54	350.145.243	283.927.511	217.539.877	151.210.731	93.511.106	50.516.230
12	01/10/2009 01:04	346.170.908	280.063.659	214.206.698	149.193.787	91.626.709	49.085.573
13	01/10/2009 01:09	349.264.104	282.771.062	216.607.452	151.057.761	92.115.522	49.571.117
14	01/10/2009 01:14	344.844.224	278.984.958	213.516.339	148.833.359	90.619.977	48.244.205
15	01/10/2009 01:19	343.625.407	278.146.682	212.490.832	148.097.621	90.380.238	48.236.495
16	01/10/2009 01:24	347.751.737	281.526.403	215.932.319	150.669.740	92.138.306	49.405.732
17	01/10/2009 01:29	345.813.151	279.614.878	214.147.390	149.203.539	91.080.831	48.681.999
18	01/10/2009 01:34	344.347.415	278.596.771	212.939.877	148.482.118	90.414.209	48.283.451
19	01/10/2009 01:39	343.097.167	277.387.473	212.933.516	147.738.417	90.716.474	48.068.649

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

- Controllo sui dati mancanti

datetime	>0.30	>0.32	>0.35	>0.38	>0.41	>0.45
10/10/2009 00:00	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:05	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:10	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:15	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:20	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:25	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:30	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:35	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:40	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:45	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:50	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 00:55	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:00	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:05	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:10	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:15	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:20	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:25	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:30	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:35	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:40	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:45	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:50	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 01:55	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:00	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:05	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:10	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:15	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:20	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:25	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:30	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:35	363307900	295624000	228232000	160790200	99891000	546411000
10/10/2009 02:40	363307900	295624000	228232000	160790200	99891000	546411000

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

- Calcolo di medie orarie in modo automatico durante la fase di importazione

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

- Calcolo di medie orarie in modo automatico durante la fase di importazione

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

Visualizzazione di serie temporali

### Miglioramento dell' importazione dei dati in TimeSearcher

Visualizzazione di serie temporali