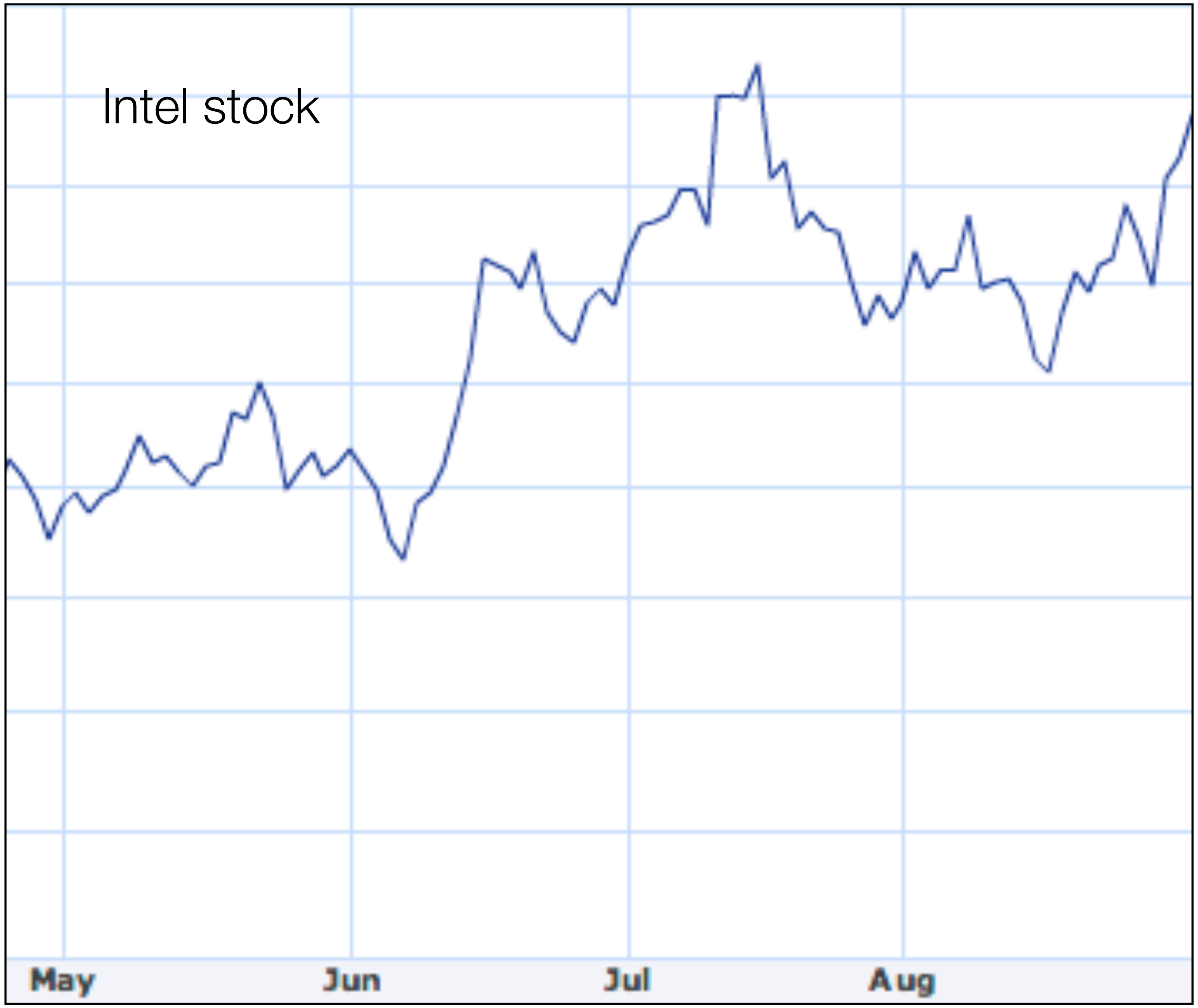
Mining Approximate Top-K Subspace Anomalies in Multi-Dimensional Time-Series Data

Xiaolei Li, Jiawei Han
University of Illinois at Urbana-Champaign

VLDB 2007

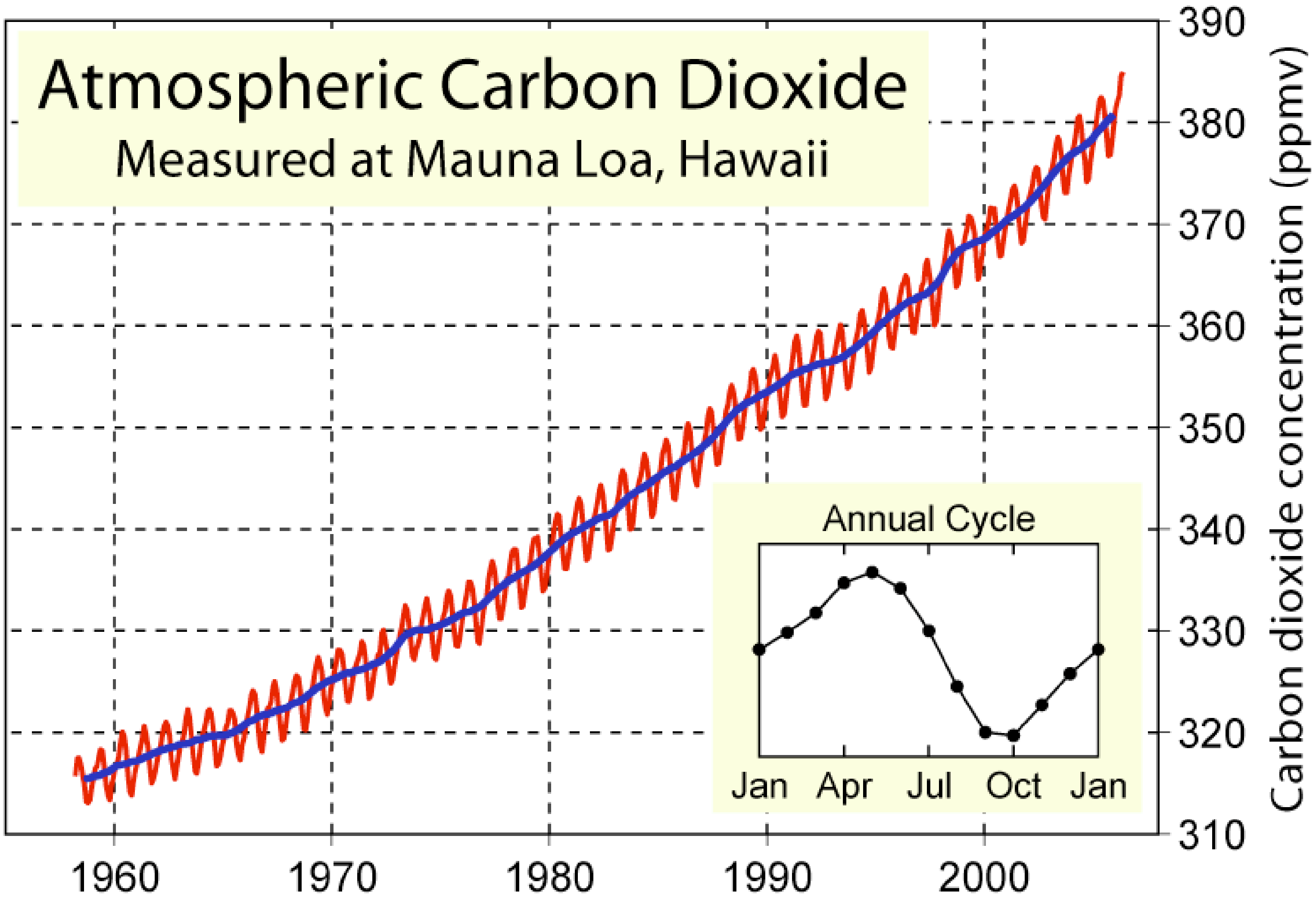
Time Series Data

- Many applications produce time series data



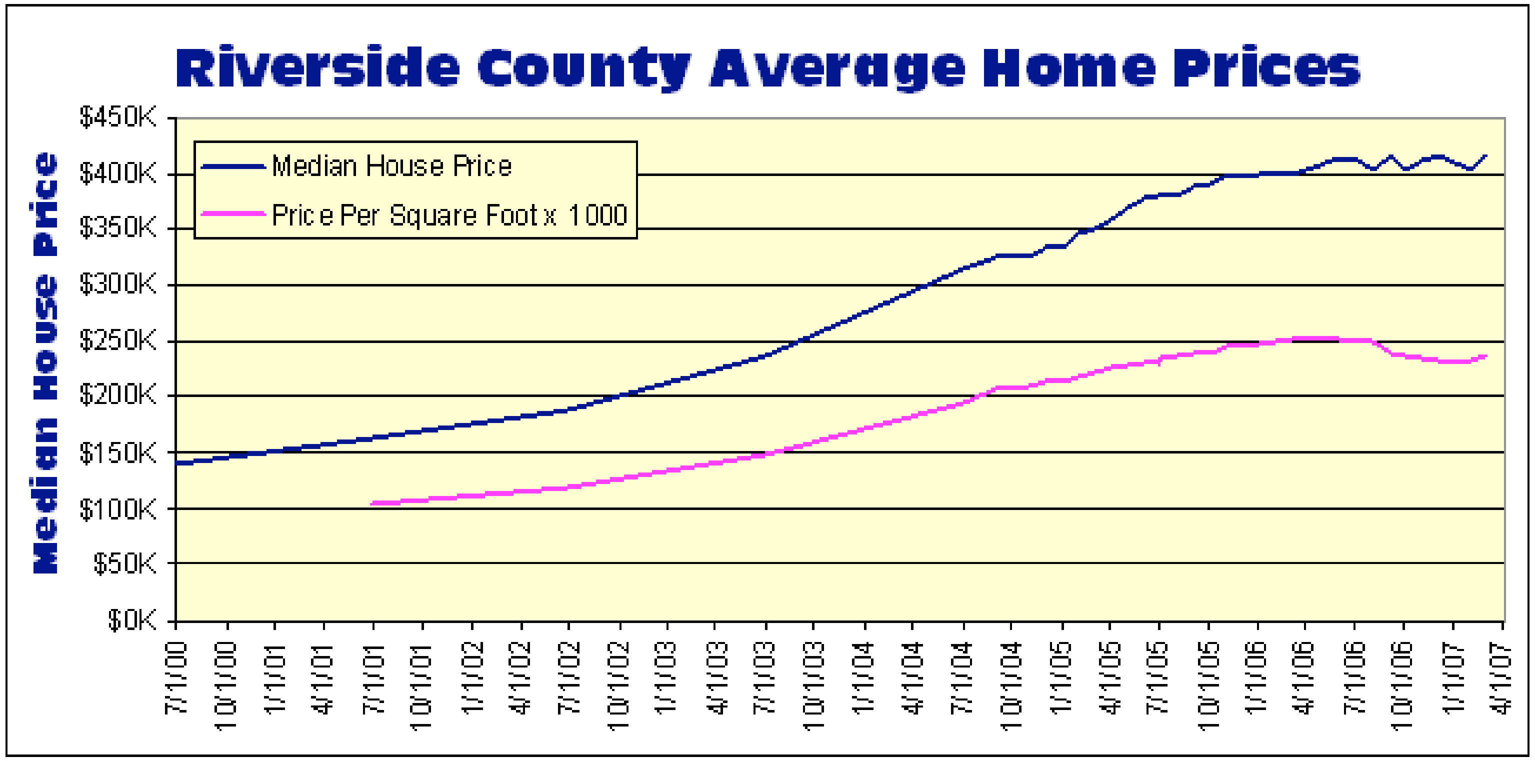
Time Series Data

- Many applications produce time series data

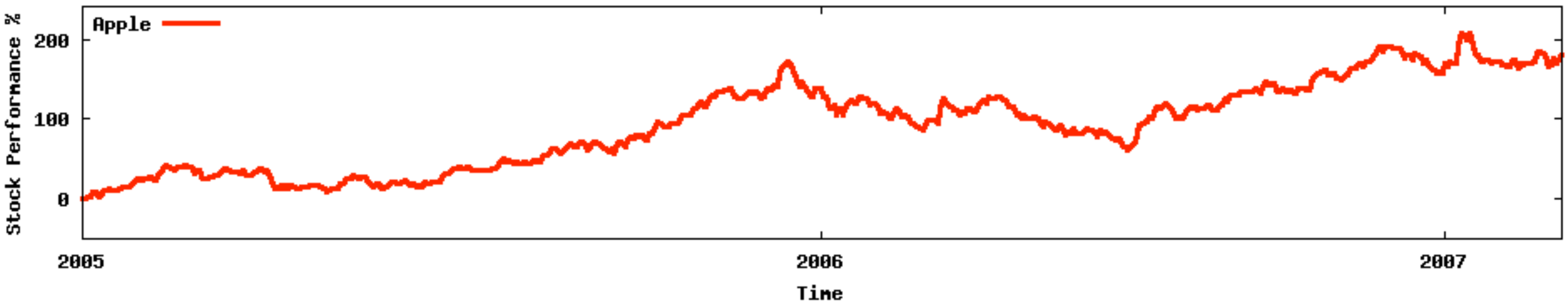


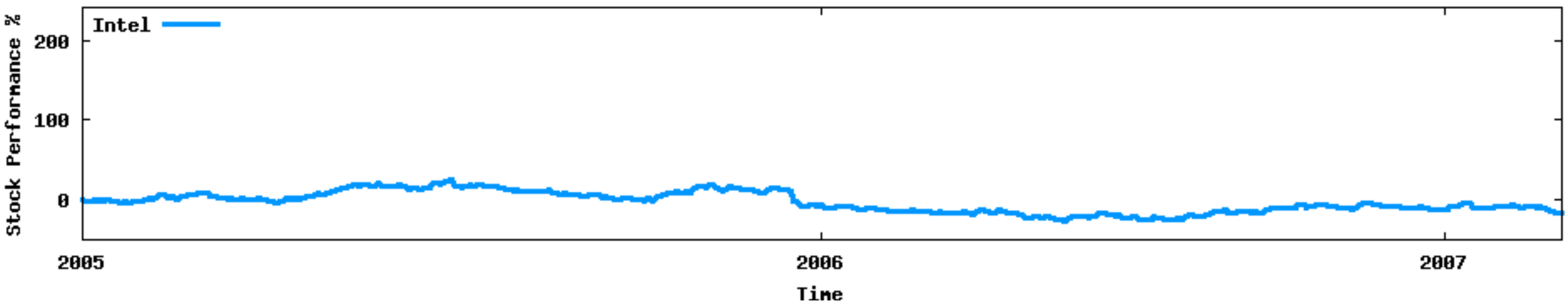
Time Series Data

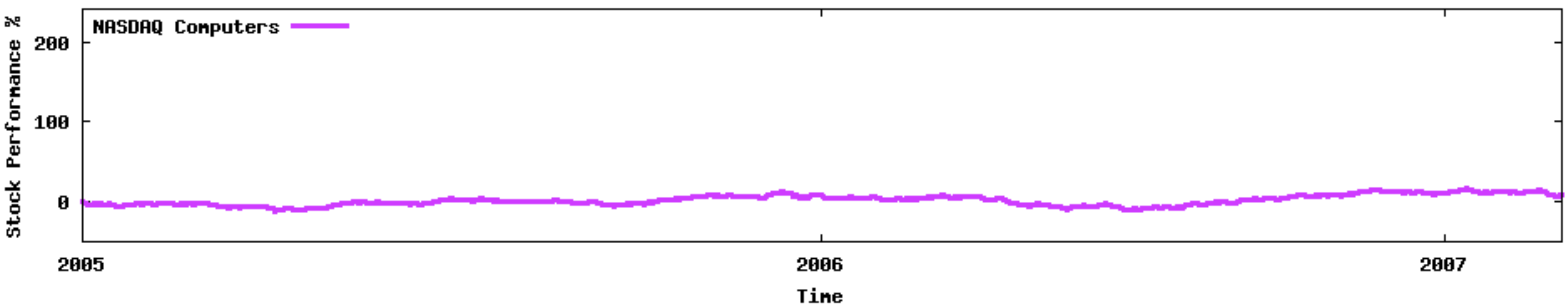
- Many applications produce time series data



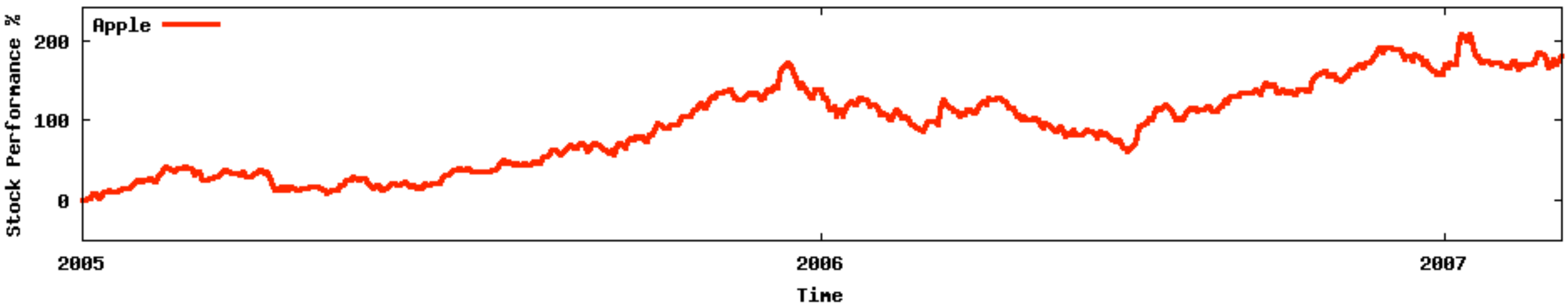
Apple, Intel, NASDAQ Computers Stock Values







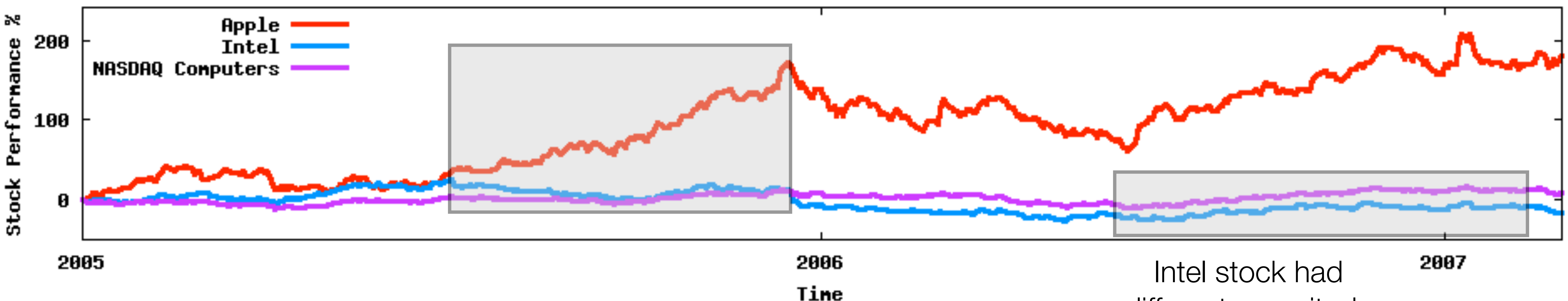
Apple, Intel, NASDAQ Computers Stock Values



Apple, Intel, NASDAQ Computers Stock Values

Compare time series to gather differences

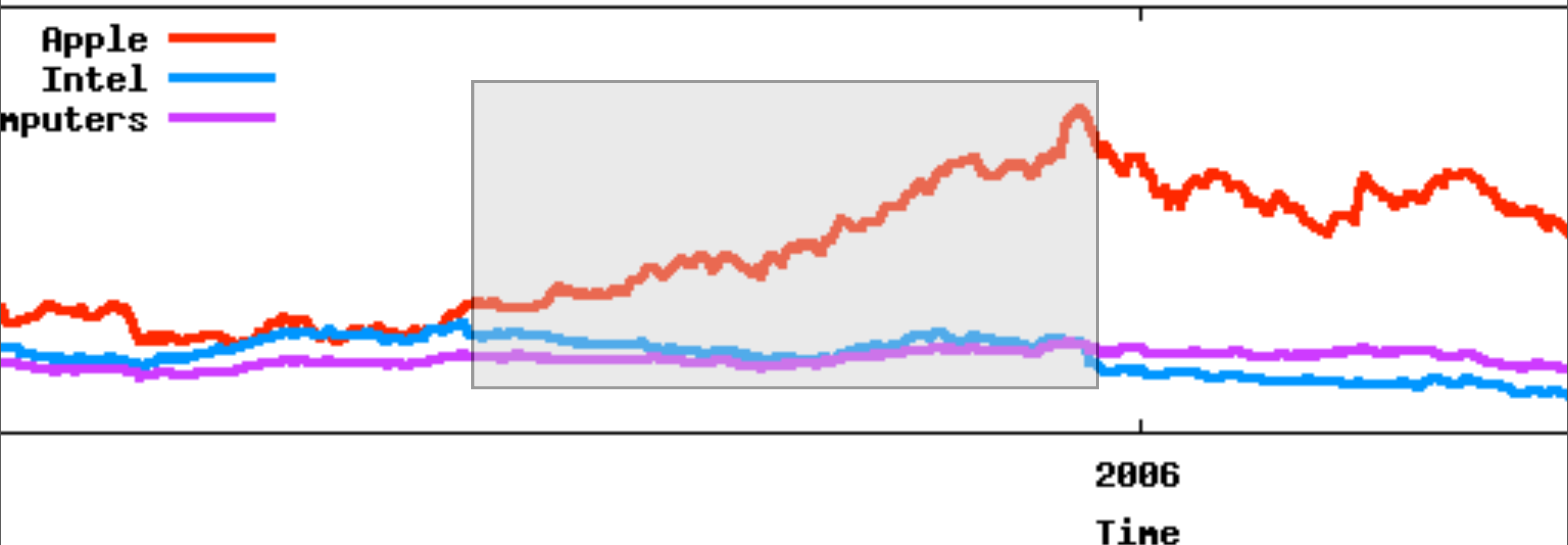
Apple stock has a
very different "trend"



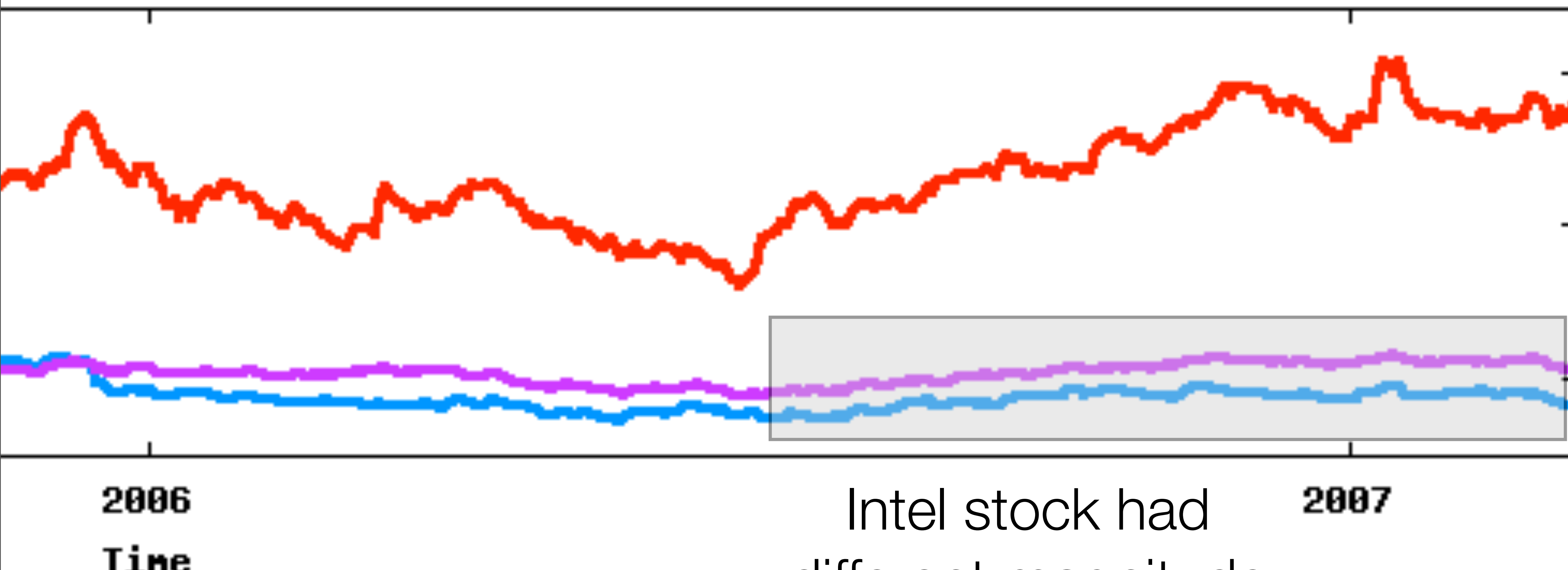
Intel stock had
different magnitude

Apple, Intel, NASDAQ Computers Stock Values

Apple stock has a
very different "trend"



Apple, Intel, NASDAQ Computers Stock Values



Intel stock had different magnitude

Problem Statement

Find **anomalies** in a

data cube of

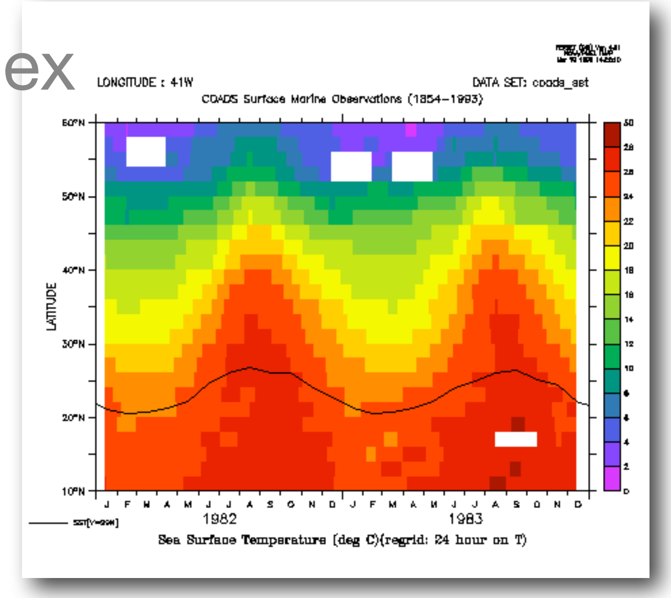
multi-dimensional

time series data

Table of Contents

Multi-Dimensional Attributes

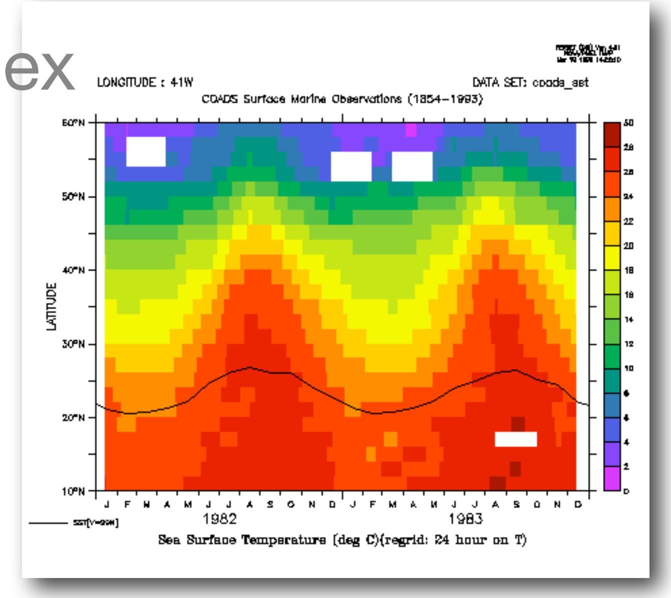
- Time series are not flat data; contains **multi-dimensional attributes**
- Stock example
  - Apple and Intel are a part of the NASDAQ Computers Index
  - Apple is hardware/software; Intel is hardware
  - Related to NASDAQ-100 Technology Stock Index



- Sales example
  - Multi-dimensional information collected for every sale (e.g., buyer age, product type, store location, purchase time)
  - Compare sales by **any combination of categories or sub-categories**: *"sales of sporting apparel to <u>males with 3+ children</u> have been declining compared to <u>overall male</u> sporting apparel sales"*

Multi-Dimensional Attributes

- Time series are not flat data; contains **multi-dimensional attributes**
- Stock example
  - Apple and Intel are a part of the NASDAQ Computers Index
  - Apple is hardware/software; Intel is hardware
  - Related to NASDAQ-100 Technology Stock Index



- Sales example
  - Multi-dimensional information collected for every sale (e.g., buyer age, product type, store location, purchase time)
  - Compare sales by **any combination of categories or sub-categories**: *"sales of sporting apparel to males with 3+ children have been declining compared to overall male sporting apparel sales"*

subset

Problem Statement

- **Find anomalies in the data cube of multi-dimensional time series data**
- Input data: relation **R** with a set of time series **S** associated with each tuple
  - Attributes of **R** form a data cube $\mathbf{C_R}$
  - Each s_i is a time series
  - Each u_i is a scalar indicating the count of the tuple

Gender	Education	Income	Product	Profit	Count
Female	Highschool	35k-45k	Food	s_1	u_1
Female	Highschool	45k-60k	Apparel	s_2	u_2
Female	College	35k-45k	Apparel	s_3	u_3
Female	College	35k-45k	Book	s_4	u_4
Female	College	45k-60k	Apparel	s_5	u_5
Female	Graduate	45k-60k	Apparel	s_6	u_6
Male	Highschool	35k-45k	Apparel	s_7	u_7
Male	College	35k-45k	Food	s_8	u_8

Problem Statement

- Find anomalies in the data cube of multi-dimensional time series data
- Input data: relation **R** with a set of time series **S** associated with each tuple
  - Attributes of **R** form a data cube **C_R**
  - Each s_i is a time series
  - Each u_i is a scalar indicating the count of the tuple

Gender	Education	Income	Product	Profit	Count
Female	Highschool	35k-45k	Food	s_1	u_1
Female	Highschool	45k-60k	Apparel	s_2	u_2
Female	College	35k-45k	Apparel	s_3	u_3
Female	College	35k-45k	Book	s_4	u_4
Female	College	45k-60k	Apparel	s_5	u_5
Female	Graduate	45k-60k	Apparel	s_6	u_6
Male	Highschool	35k-45k	Apparel	s_7	u_7
Male	College	35k-45k	Food	s_8	u_8

Problem Statement

- Find anomalies in the data cube of multi-dimensional time series data
- Input data: relation **R** with a set of time series **S** associated with each tuple
  - Attributes of **R** form a data cube $\mathbf{C_R}$
  - Each s_i is a time series
  - Each u_i is a scalar indicating the count of the tuple

Gender	Education	Income	Product	Profit	Count
Female	Highschool	35k-45k	Food	s_1	u_1
Female	Highschool	45k-60k	Apparel	s_2	u_2
Female	College	35k-45k	Apparel	s_3	u_3
Female	College	35k-45k	Book	s_4	u_4
Female	College	45k-60k	Apparel	s_5	u_5
Female	Graduate	45k-60k	Apparel	s_6	u_6
Male	Highschool	35k-45k	Apparel	s_7	u_7
Male	College	35k-45k	Food	s_8	u_8

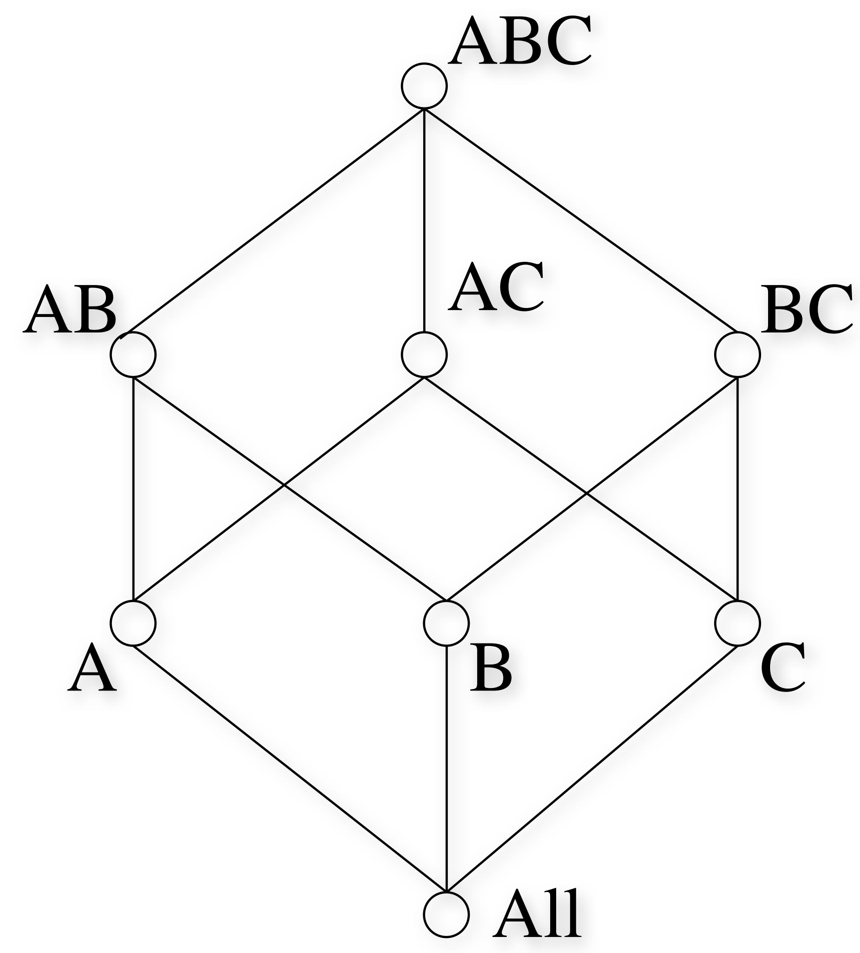
Problem Statement

- Find anomalies in the data cube of multi-dimensional time series data
- Input data: relation **R** with a set of time series **S** associated with each tuple
  - Attributes of **R** form a data cube $\mathbf{C_R}$
  - Each s_i is a time series
  - Each u_i is a scalar indicating the count of the tuple

Gender	**Education**	**Income**	**Product**	**Profit**	**Count**
Female	Highschool	35k-45k	Food	s_1	u_1
Female	Highschool	45k-60k	Apparel	s_2	u_2
Female	College	35k-45k	Apparel	s_3	u_3
Female	College	35k-45k	Book	s_4	u_4
Female	College	45k-60k	Apparel	s_5	u_5
Female	Graduate	45k-60k	Apparel	s_6	u_6
Male	Highschool	35k-45k	Apparel	s_7	u_7
Male	College	35k-45k	Food	s_8	u_8

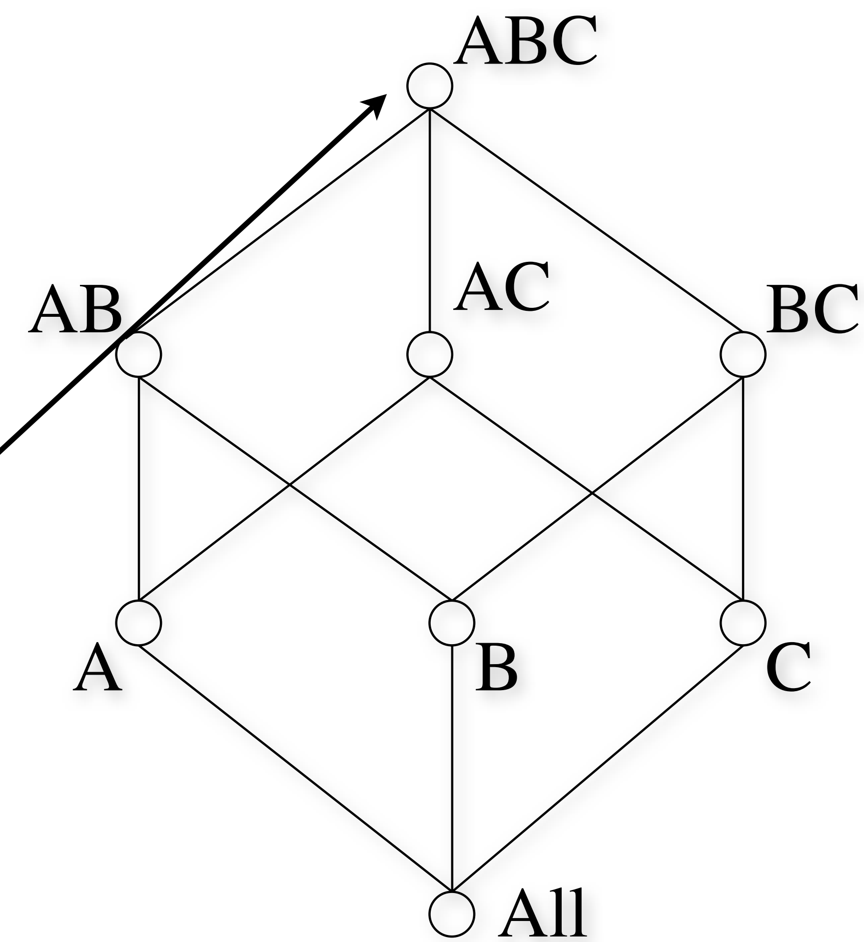
Data Cube Preliminaries

- Given a relation **R**, a data cube (denoted as **C_R**) is the set of aggregates from all possible group-by's on **R**
- In a *n*-dimensional data cube, each cell has the form **c = (a_1, a_2, ..., a_n : m)** where each a_i is the value of i^{th} attribute and m is the cube measure (e.g., profit)
- A cell is k-dimensional if there are exactly k ($\leq$ n) values amongst a_i which are not $*$ (i.e., all)
  - 2-dimensional cell: (Female, $*$, $*$, Book: x)
  - 3-dimensional cell: ($*$, College, 35k-45k, Apparel: y)
  - Base cell: none of a_i is $*$
- Parent, descendant, sibling relationships



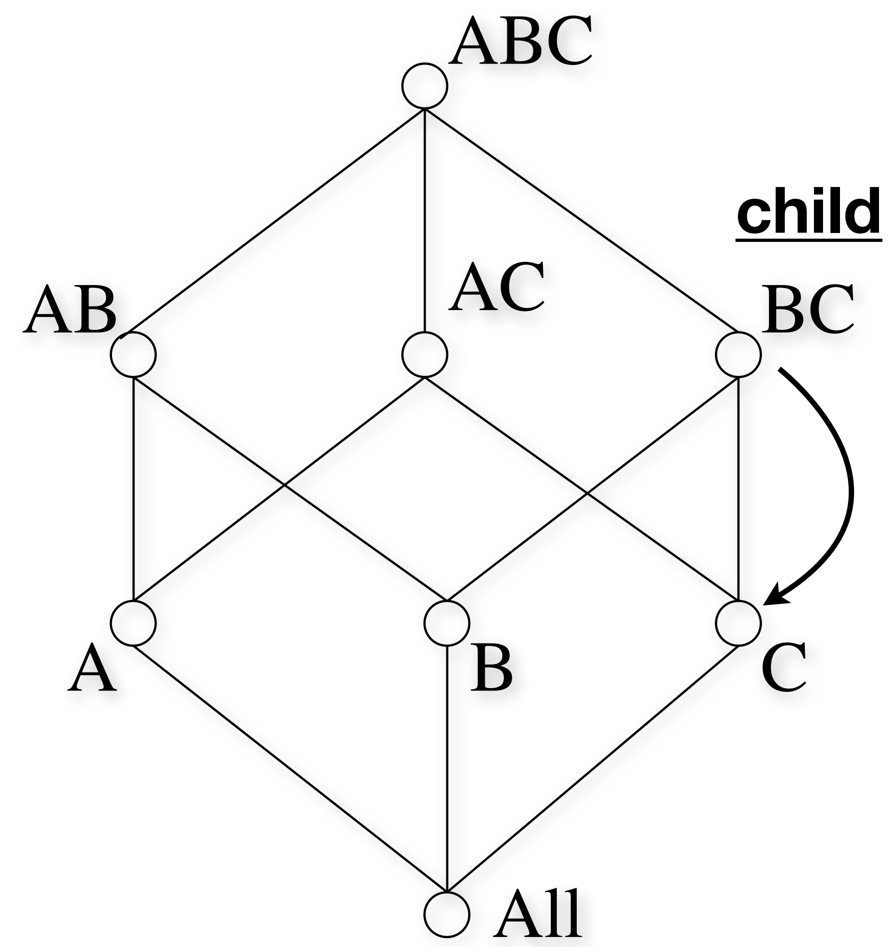
Data Cube Preliminaries

- Given a relation **R**, a data cube (denoted as $\mathbf{C_R}$) is the set of aggregates from all possible group-by's on **R**
- In a *n*-dimensional data cube, each cell has the form $\mathbf{c = (a_1, a_2, ..., a_n : m)}$ where each a_i is the value of i^{th} attribute and m is the cube measure (e.g., profit)
- A cell is k-dimensional if there are exactly k ($\leq$ n) values amongst a_i which are not $*$ (i.e., all)
  - 2-dimensional cell: (Female, $*$, $*$, Book: x)
  - 3-dimensional cell: ($*$, College, 35k-45k, Apparel: y)
  - Base cell: none of a_i is $*$
- Parent, descendant, sibling relationships



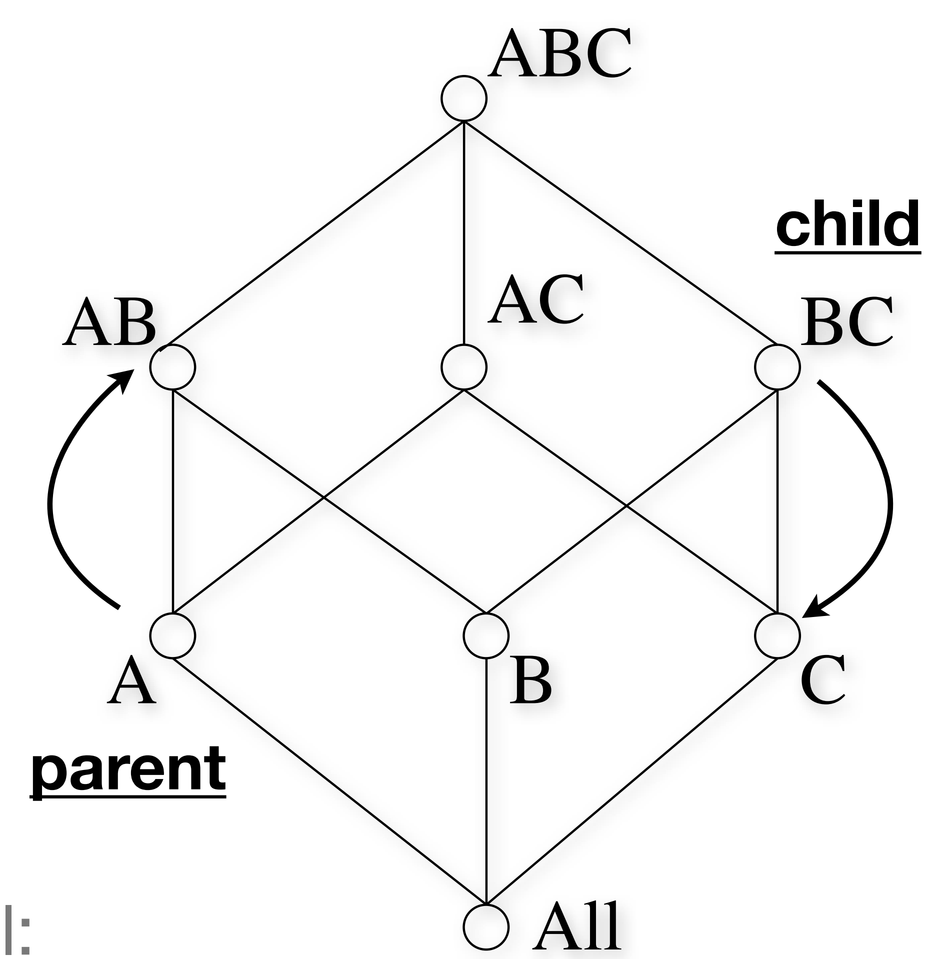
Data Cube Preliminaries

- Given a relation **R**, a data cube (denoted as **C_R**) is the set of aggregates from all possible group-by's on **R**
- In a *n*-dimensional data cube, each cell has the form **$c = (a_1, a_2, \ldots, a_n : m)$** where each a_i is the value of i^{th} attribute and m is the cube measure (e.g., profit)
- A cell is k-dimensional if there are exactly k ($\leq$ n) values amongst a_i which are not $*$ (i.e., all)
  - 2-dimensional cell: (Female, $*$, $*$, Book: x)
  - 3-dimensional cell: ($*$, College, 35k-45k, Apparel: y)
  - Base cell: none of a_i is $*$
- Parent, descendant, sibling relationships



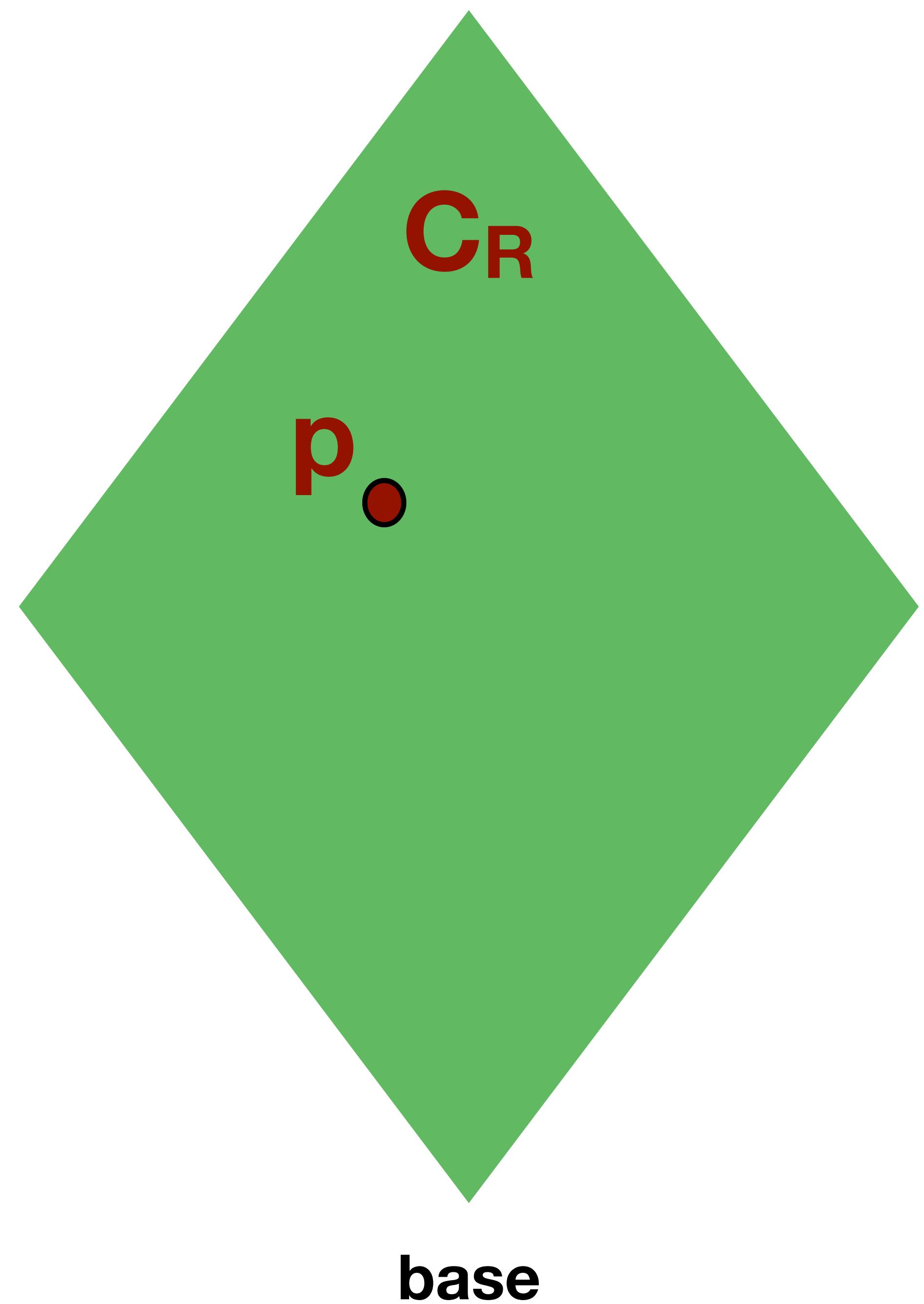
Data Cube Preliminaries

- Given a relation **R**, a data cube (denoted as **C_R**) is the set of aggregates from all possible group-by's on **R**
- In a *n*-dimensional data cube, each cell has the form **$c = (a_1, a_2, ..., a_n : m)$** where each a_i is the value of i^{th} attribute and m is the cube measure (e.g., profit)
- A cell is k-dimensional if there are exactly k ($\leq$ n) values amongst a_i which are not $*$ (i.e., all)
  - 2-dimensional cell: (Female, $*$, $*$, Book: x)
  - 3-dimensional cell: ($*$, College, 35k-45k, Apparel: y)
  - Base cell: none of a_i is $*$
- Parent, descendant, sibling relationships



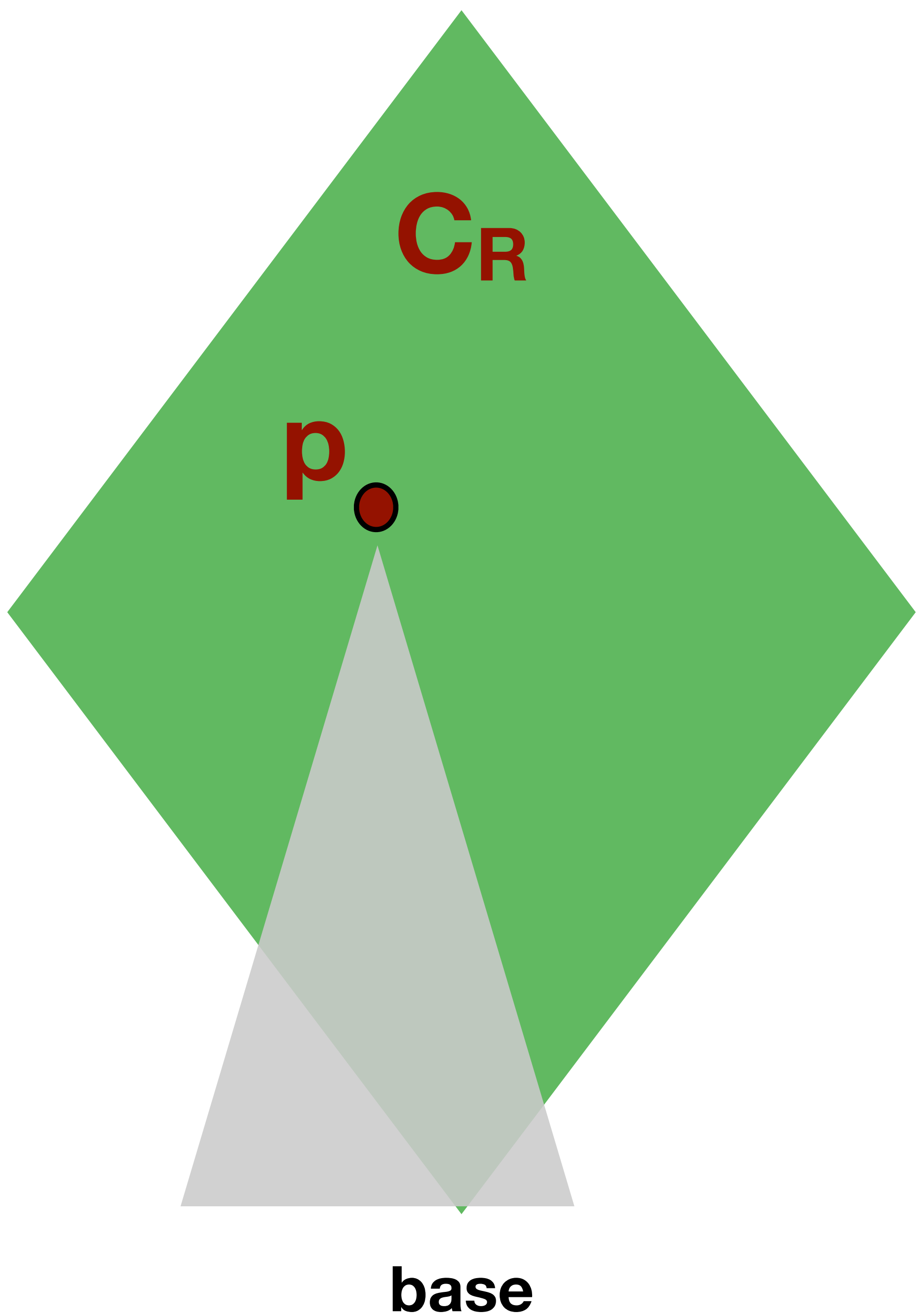
Query Model

- Given **R**, a probe cell **p** ∈ $\mathbf{C_R}$, and an anomaly function ***g***, find the **anomaly cells among descendants of p in $\mathbf{C_R}$ as measured by g**
  - Each abnormal cell must satisfy a minimum support (count) threshold
  - Anomaly does not have to hold for entire time series
  - Only the top *k* anomalies as ranked by *g* are needed



Query Model

- Given **R**, a probe cell **p** ∈ $\mathbf{C_R}$, and an anomaly function ***g***, find the **anomaly cells among descendants of p in $\mathbf{C_R}$ as measured by g**
  - Each abnormal cell must satisfy a minimum support (count) threshold
  - Anomaly does not have to hold for entire time series
  - Only the top *k* anomalies as ranked by *g* are needed



Related Work

- Exploratory Data Analysis
  - [Sarawagi SIGMOD'00] explores OLAP anomaly but necessitates full cube materialization
  - [Palpanas SSDBM'01] approximately finds interesting cells in data cube but still requires exponential calculations
  - [Imielinski DMKD'02] requires anti-monotonic measure and does not focus on time series
- Time Series Data Cube [Chen VLDB'02]
  - Only suitable for low-dimensional data
  - Requires user guidance
- General outlier detection, subspace clustering, and time series similarity search does not address OLAP-style data

Measuring Anomaly: Intuition

Measuring Anomaly: Intuition

1. For every cell, compute the expected time series (with respect to the probe cell)

Measuring Anomaly: Intuition

1. For every cell, compute the expected time series (with respect to the probe cell)

2. Compare the expected time series vs. the observed time series

Measuring Anomaly: Intuition

1. For every cell, compute the expected time series (with respect to the probe cell)

2. Compare the expected time series vs. the observed time series

3. Rank to get top *k*

Observed Time Series

- Given any cell ***c*** in $\mathbf{C_R}$, there is an associated **observed time series *s*$_c$**

- In the context of a probe cell *p*, it is computed by aggregating all time series associated with both *c* and *p*

$$s_c = \sum_{tid_i \in (c \cap \sigma_p(R))} s_i$$

Observed Time Series (2)

Gender	Education	Income	Product	Profit	Count
Female	Highschool	35k-45k	Food	s_1	u_1
Female	Highschool	45k-60k	Apparel	s_2	150
Female	College	35k-45k	Apparel	s_3	200
Female	College	35k-45k	Book	s_4	u_4
Female	College	45k-60k	Apparel	s_5	600
Female	Graduate	45k-60k	Apparel	s_6	50
Male	Highschool	35k-45k	Apparel	s_7	u_7
Male	College	35k-45k	Food	s_8	u_8

- Example: p = (Gender = "Female", Product = "Apparel")

	c		s_c	|c|
	Education	**Income**	**Profit**	**Count**
p	*	*	$s_2 + s_3 + s_5 + s_6$	1000
	Highschool	*	s_2	150
	College	*	$s_3 + s_5$	800

Expected Time Series

- Given any cell *c* that is a descendant of *p*, there is also an **expected time series $\hat{s}_c$**
- Intuition: A descendant cell of *p* is a subset of *p*. Assuming that market segments behave proportionally by its size, one can calculate the expected time series from *p*'s time series

$$\hat{s}_c = \left(\frac{|c|}{|p|}\right) s_p$$

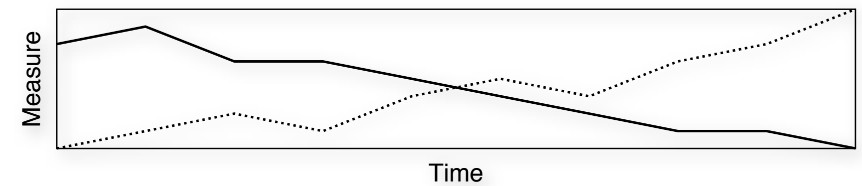
c		s_c	$\hat{s}_c$	|c|
Education	**Income**	**Profit**		**Count**
*	*	$s_2 + s_3 + s_5 + s_6 = s_p$	n/a	1000
Highschool	*	s_2	150 / 1000 x s_p	150
College	*	$s_3 + s_5$	800 / 1000 x s_p	800

Anomaly Definition

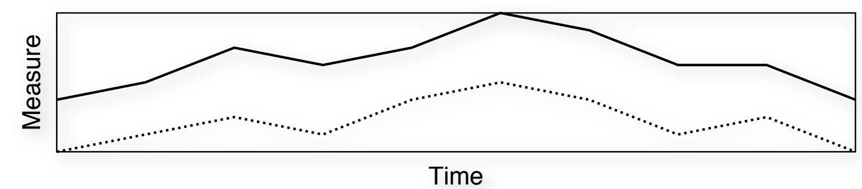
- General idea: **$g(s_c, \hat{s}_c) \Rightarrow R$**

Anomaly Definition

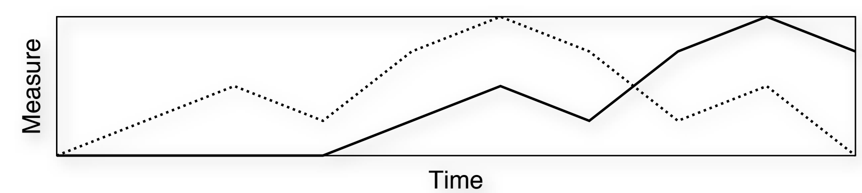
- General idea: $\mathbf{g(s_c, \hat{s}_c) \Rightarrow R}$
- Four types of anomalies
  - Trend
  - Magnitude
  - Phase
  - Miscellaneous



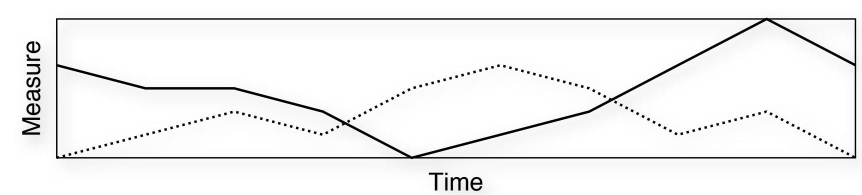
(a) Trend Anomaly



(b) Magnitude Anomaly



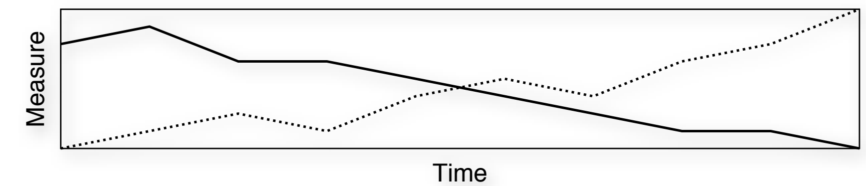
(c) Phase Anomaly



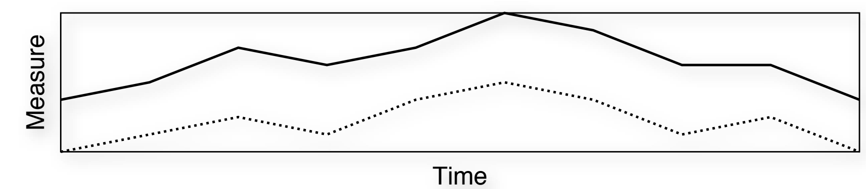
(d) Miscellaneous Anomaly

Anomaly Definition

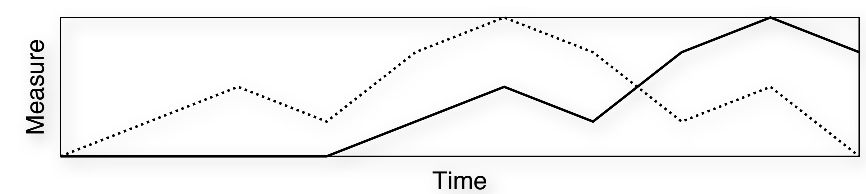
- General idea: **$g(s_c, \hat{s}_c) \Rightarrow R$**
- Four types of anomalies
  - Trend
  - Magnitude
  - Phase
  - Miscellaneous
- Measured via **first-order linear regression**
  - Simple and efficient (direct cube aggregation of parameters [Chen VLDB'02])
  - Effective at catching obvious anomalies



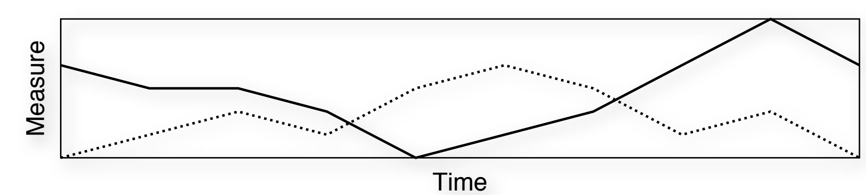
(a) Trend Anomaly



(b) Magnitude Anomaly



(c) Phase Anomaly



(d) Miscellaneous Anomaly

Mining Top-K Anomalies in Data Cube

Algorithm 1 Naïve Top-k Anomalies

Input: Relation R, time-series data S, query probe cell p, anomaly function g, parameter k, minimum support m

Output: Top-k scoring cells in C_p as ranked by g and satisfies m

1. Retrieve data for $\sigma_p(R)$
2. Compute the data cube C_p with $\sigma_p(R)$ as the fact table with m as the iceberg parameter
3. Return top k anomaly cells in C_p for each g

Mining Top-K Anomalies in Data Cube

Algorithm 1 Naïve Top-k Anomalies

Input: Relation R, time-series data S, query probe cell p, anomaly function g, parameter k, minimum support m

Output: Top-k scoring cells in C_p as ranked by g and satisfies m

1. Retrieve data for $\sigma_p(R)$
2. Compute the data cube C_p with $\sigma_p(R)$ as the fact table with m as the iceberg parameter
3. Return top k anomaly cells in C_p for each g

1. Expensive to compute C_p (exponential in number of dimensions)

Mining Top-K Anomalies in Data Cube

Algorithm 1 Naïve Top-k Anomalies

Input: Relation R, time-series data S, query probe cell p, anomaly function g, parameter k, minimum support m

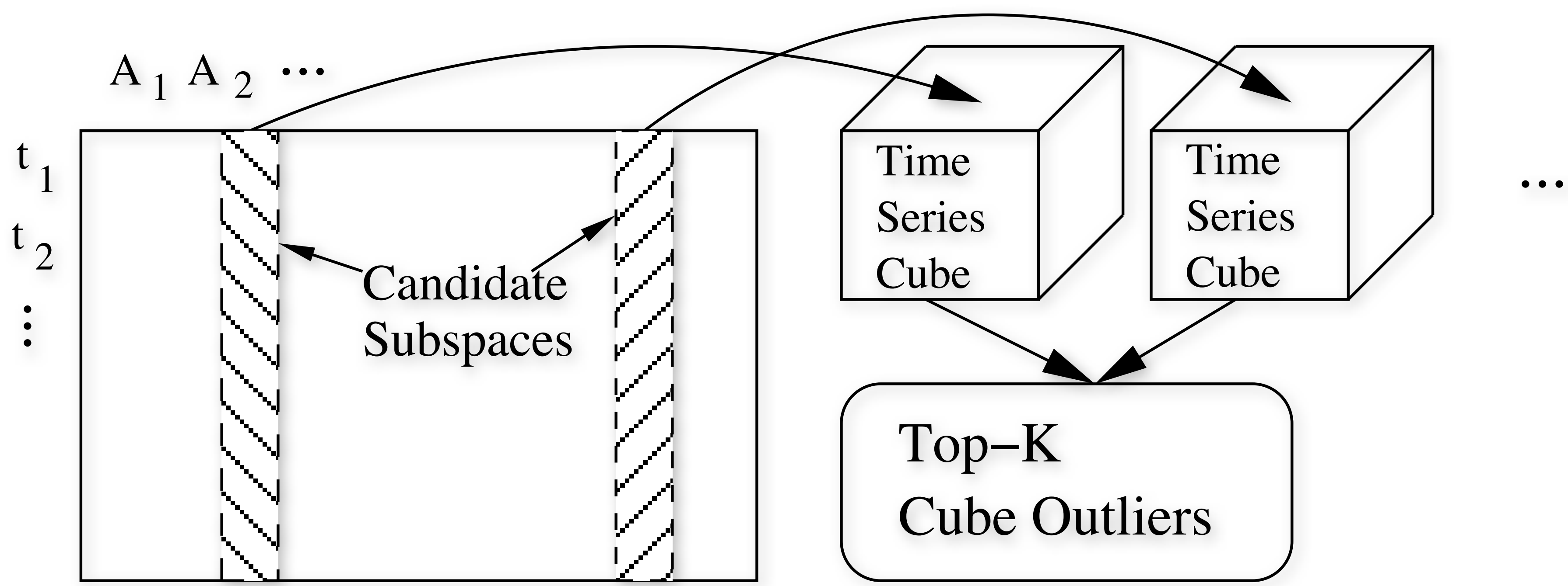
Output: Top-k scoring cells in C_p as ranked by g and satisfies m

1. Retrieve data for $\sigma_p(R)$
2. Compute the data cube C_p with $\sigma_p(R)$ as the fact table with m as the iceberg parameter
3. Return top k anomaly cells in C_p for each g

1. Expensive to compute C_p (exponential in number of dimensions)
2. Finds all anomalies before collecting top-*k*

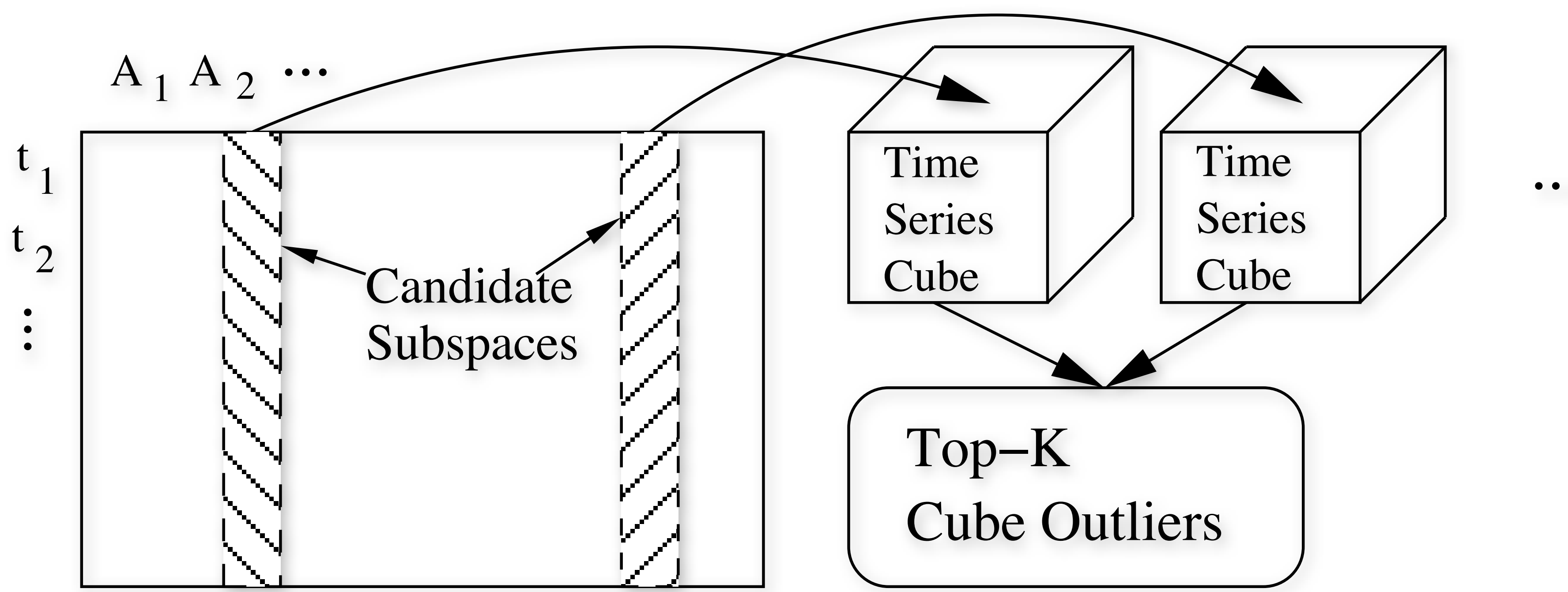
SUITS Framework

- **Subspace Iterative Time Series Anomaly Search (SUITS)**



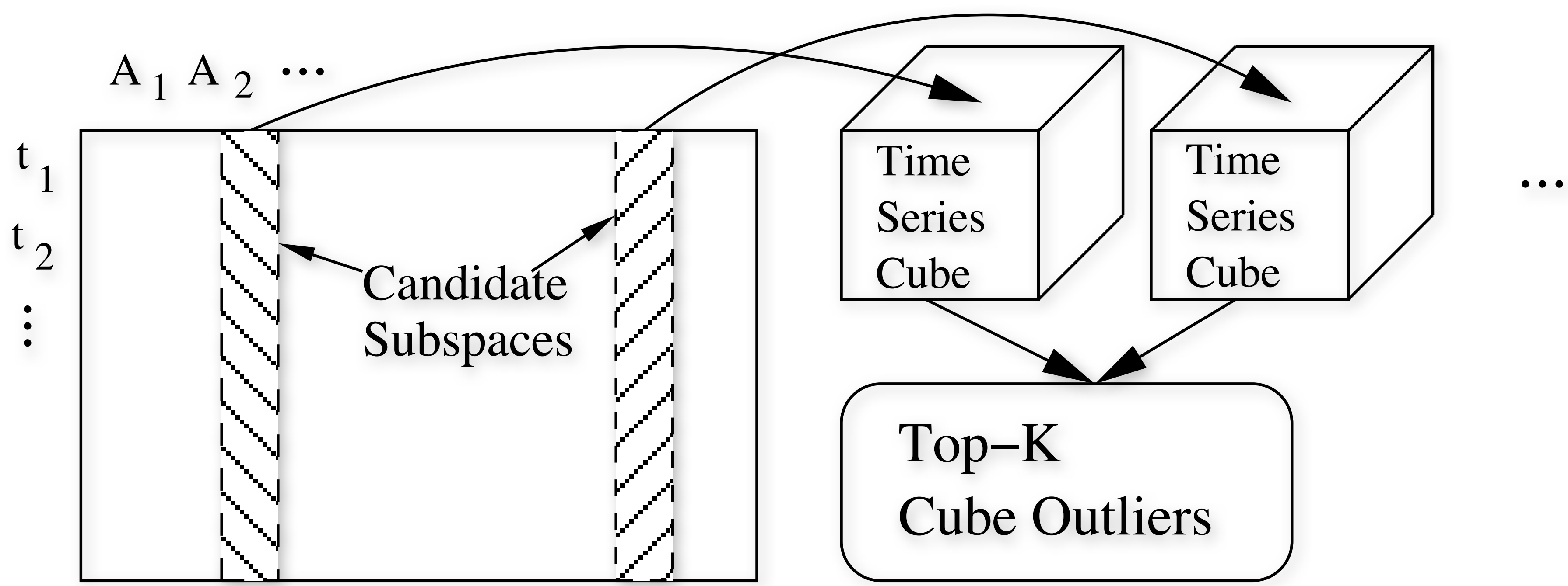
SUITS Framework

- **Subspace Iterative Time Series Anomaly Search (SUITS)**
- Iteratively select subspaces out of the 2^n total subspaces



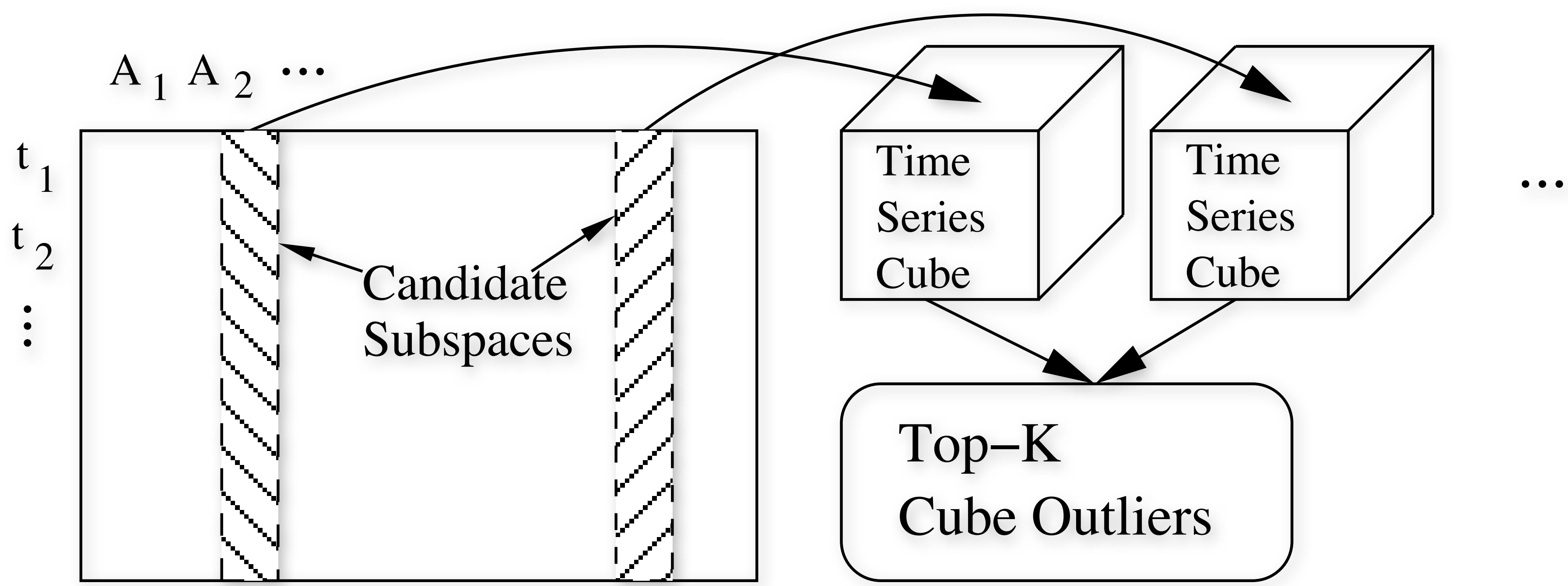
SUITS Framework

- **Subspace Iterative Time Series Anomaly Search (SUITS)**
- Iteratively select subspaces out of the 2^n total subspaces
- Compute anomalies within subspaces



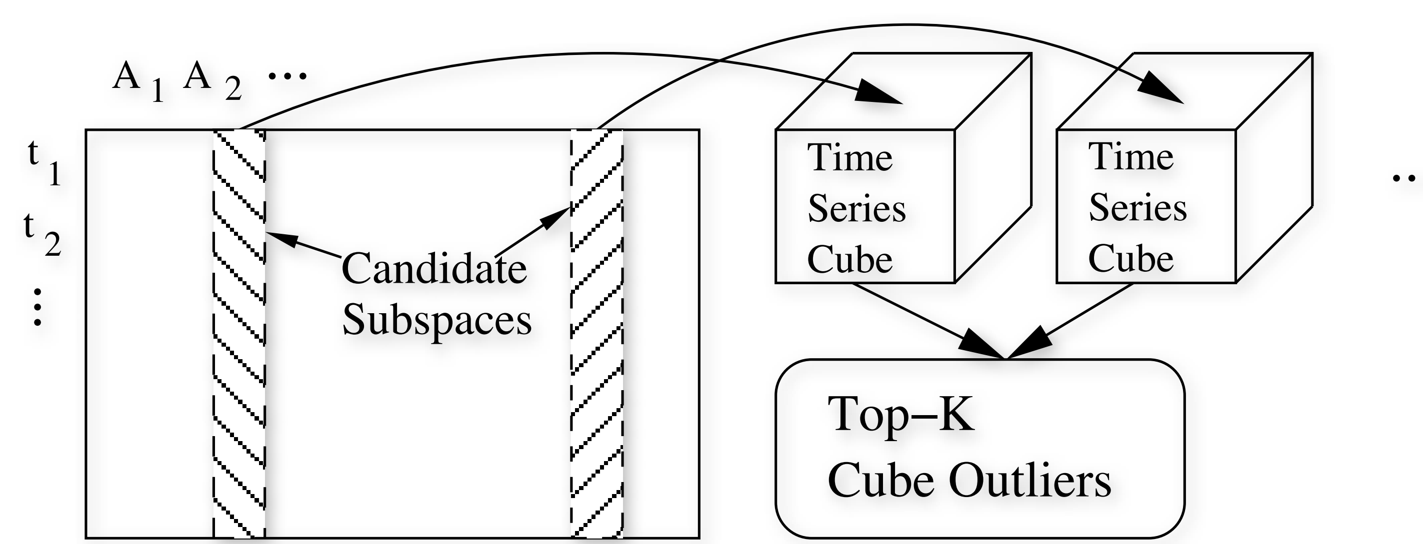
SUITS Framework

- **Subspace Iterative Time Series Anomaly Search (SUITS)**
- Iteratively select subspaces out of the 2^n total subspaces
- Compute anomalies within subspaces
- Combine to form overall anomalies



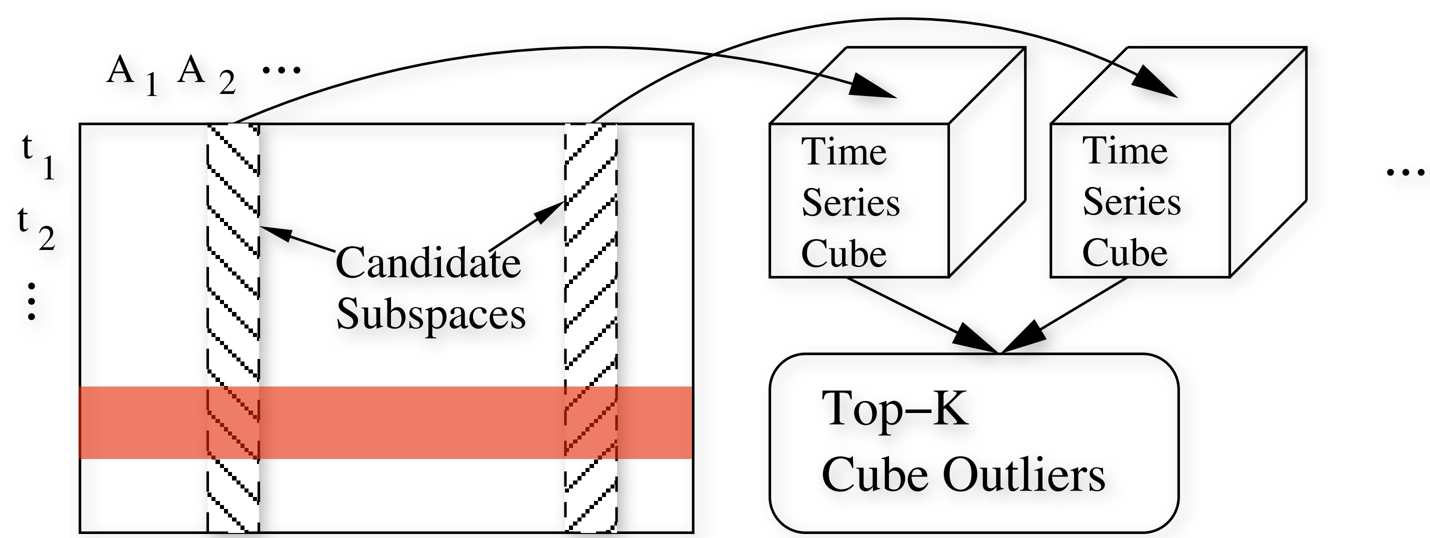
How to Choose Candidate Subspaces

- Intuition
    - By definition, anomalies are rare and most of the 2^n subspaces do not contain any
    - Descendant cells stemming from the same anomalies (in some ancestor cell) should exhibit similar abnormal behavior
- Procedure



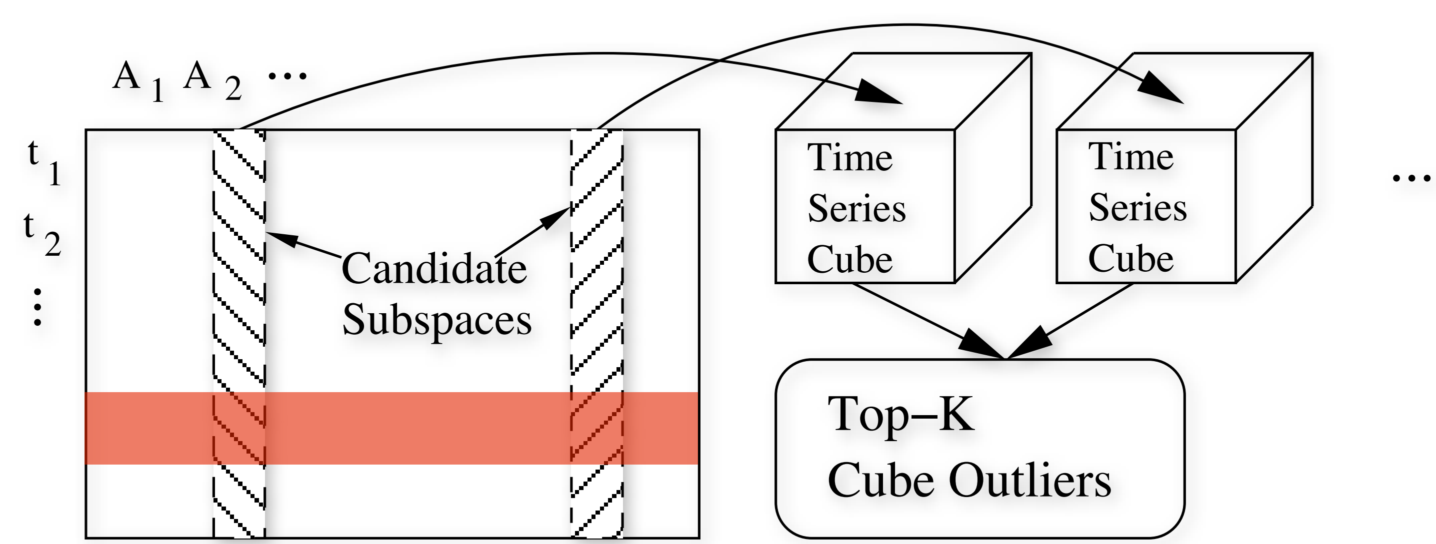
How to Choose Candidate Subspaces

- Intuition
  - By definition, anomalies are rare and most of the 2^n subspaces do not contain any
  - Descendant cells stemming from the same anomalies (in some ancestor cell) should exhibit similar abnormal behavior
- Procedure
  1. Search for a group of similar anomalies in the set of base cells



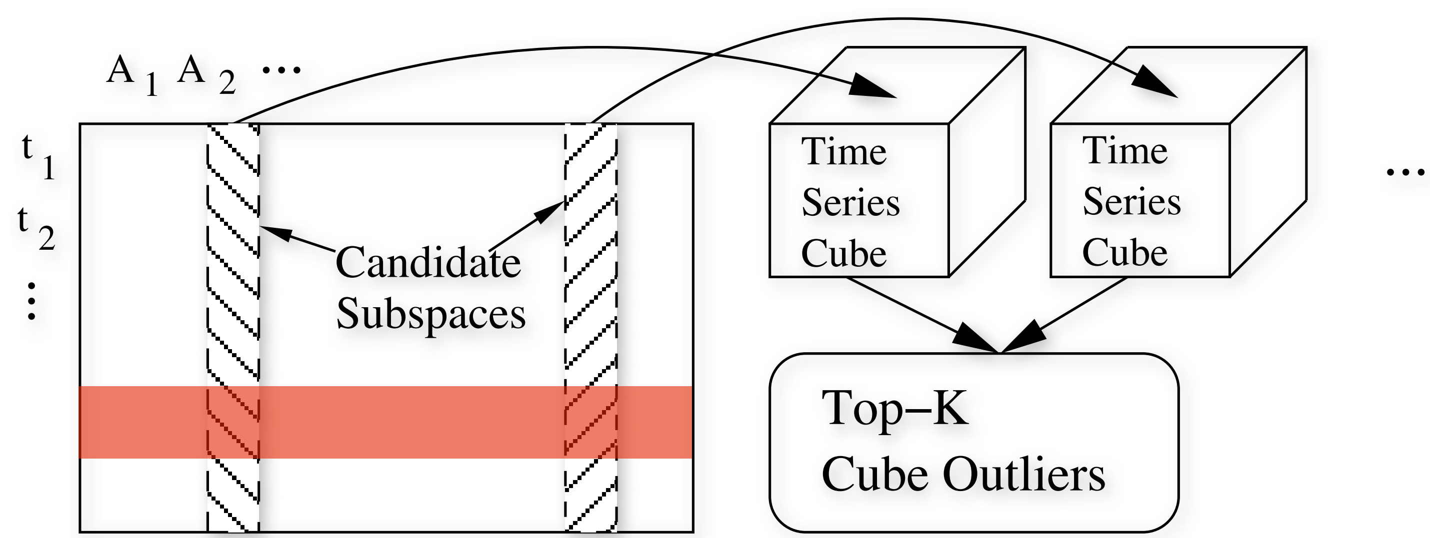
How to Choose Candidate Subspaces

- Intuition
  - By definition, anomalies are rare and most of the 2^n subspaces do not contain any
  - Descendant cells stemming from the same anomalies (in some ancestor cell) should exhibit similar abnormal behavior
- Procedure
  1. Search for a group of similar anomalies in the set of base cells
  2. Find a subspace correlated with the group



How to Choose Candidate Subspaces

- Intuition
  - By definition, anomalies are rare and most of the 2^n subspaces do not contain any
  - Descendant cells stemming from the same anomalies (in some ancestor cell) should exhibit similar abnormal behavior
- Procedure
  1. Search for a group of similar anomalies in the set of base cells
  2. Find a subspace correlated with the group
  3. Compute the local top-*k* anomalies in the subspace



How to Choose Candidate Subspaces (2)

- Time Anomaly Matrix

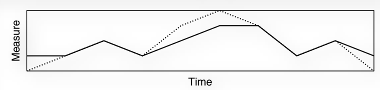
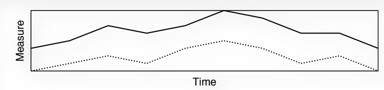
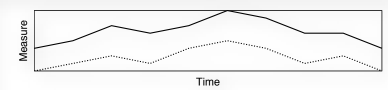
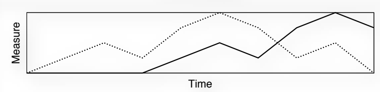
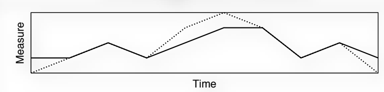
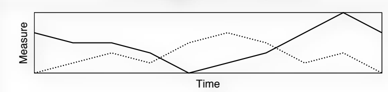
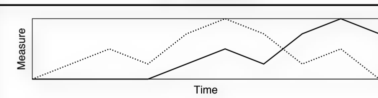
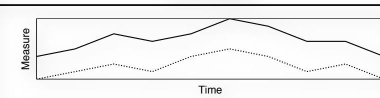
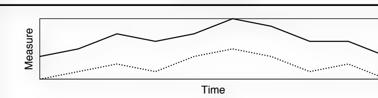
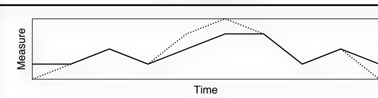
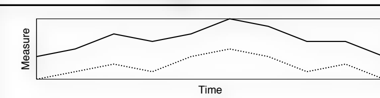
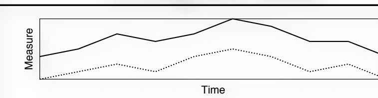
Education	Income	$S[1]$	$S[2]$	$S[3]$
Highschool	45k–60k	 None	 **Magnitude**	 **Magnitude**
College	35k–45k	 Phase	 None	 Misc
College	45k–60k	 Phase	 **Magnitude**	 **Magnitude**
Graduate	45k–60k	 None	 **Magnitude**	 **Magnitude**

Table 4: Time Anomaly Matrix

  - Partition each observed and expected time series into subsequences and compute anomalies
  - Group anomalies by type and also amount
  - Iteratively select groups of similar anomaly cells from matrix

How to Choose Candidate Subspaces (2)

- Time Anomaly Matrix

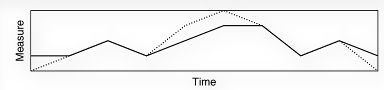
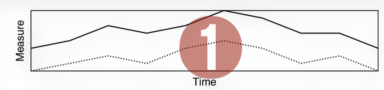
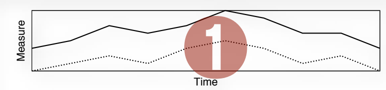
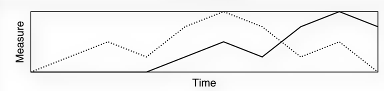
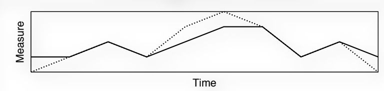
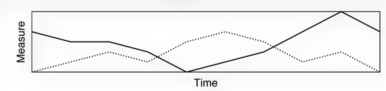
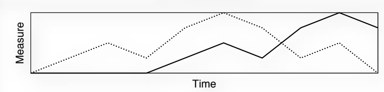
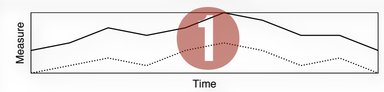
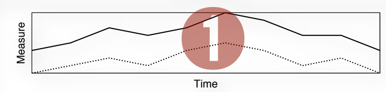
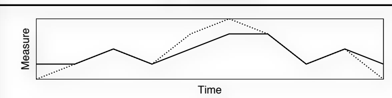
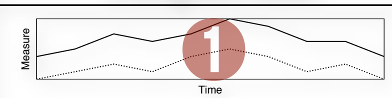
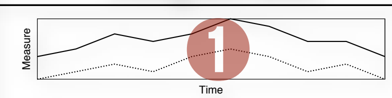
Education	Income	$S[1]$	$S[2]$	$S[3]$
Highschool	45k–60k	 None	 **Magnitude**	 **Magnitude**
College	35k–45k	 Phase	 None	 Misc
College	45k–60k	 Phase	 **Magnitude**	 **Magnitude**
Graduate	45k–60k	 None	 **Magnitude**	 **Magnitude**

Table 4: Time Anomaly Matrix

- Partition each observed and expected time series into subsequences and compute anomalies
- Group anomalies by type and also amount
- Iteratively select groups of similar anomaly cells from matrix

How to Choose Candidate Subspaces (2)

- Time Anomaly Matrix

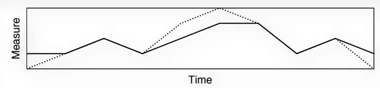
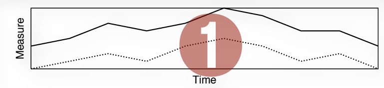
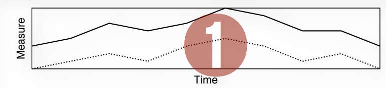
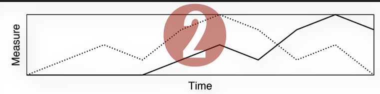
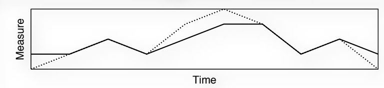
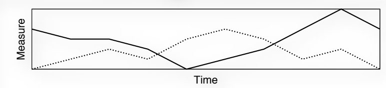
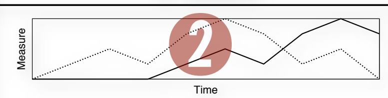
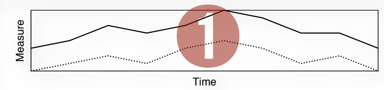
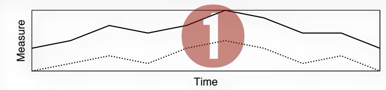
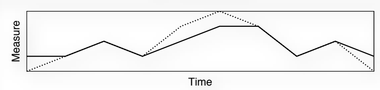
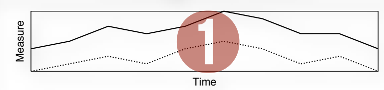
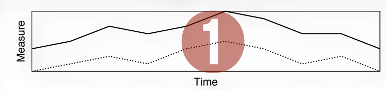
Education	Income	$S[1]$	$S[2]$	$S[3]$
Highschool	45k–60k	 None	 **Magnitude**	 **Magnitude**
College	35k–45k	 Phase	 None	 Misc
College	45k–60k	 Phase	 **Magnitude**	 **Magnitude**
Graduate	45k–60k	 None	 **Magnitude**	 **Magnitude**

Table 4: Time Anomaly Matrix

- Partition each observed and expected time series into subsequences and compute anomalies
- Group anomalies by type and also amount
- Iteratively select groups of similar anomaly cells from matrix

How to Choose Candidate Subspaces (3)

- Given a group in the Time Anomaly Matrix, select its correlated subspace
- Rank attribute-value pairs by **Anomaly Likelihood** (AL) score
  - Attribute values that occur very frequently and within a homogenous dimension have high AL scores
  - AL = (Frequency of Attribute-Value) x (Entropy of Attribute)$^{-1}$
- Select the top few and form the candidate subspace

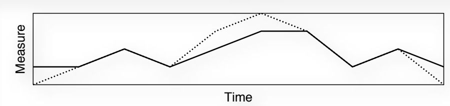
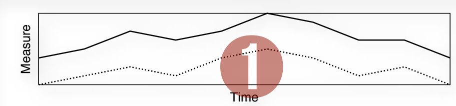
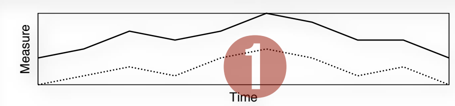
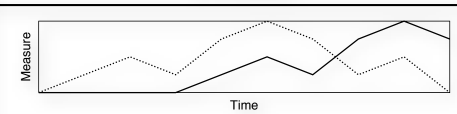
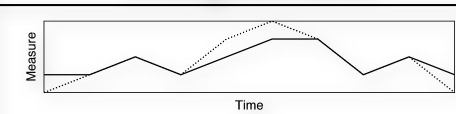
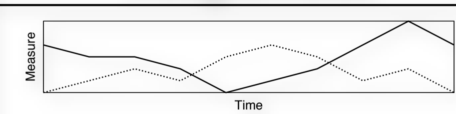
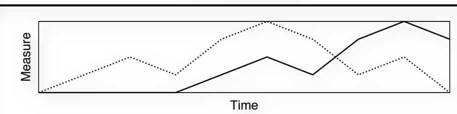
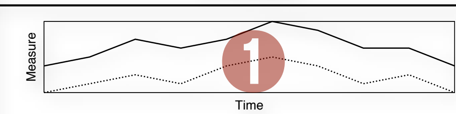
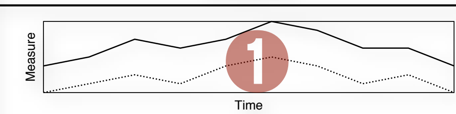
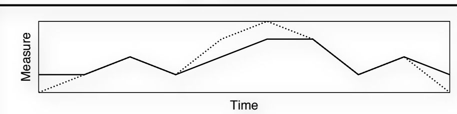
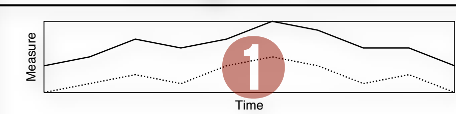
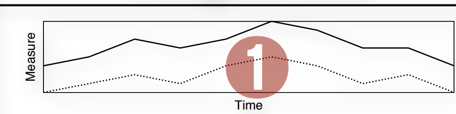
Education	Income	$S[1]$	$S[2]$	$S[3]$
Highschool	45k–60k	 None	 **Magnitude**	 **Magnitude**
College	35k–45k	 Phase	 None	 Misc
College	45k–60k	 Phase	 **Magnitude**	 **Magnitude**
Graduate	45k–60k	 None	 **Magnitude**	 **Magnitude**

Table 4: Time Anomaly Matrix

How to Choose Candidate Subspaces (3)

- Given a group in the Time Anomaly Matrix, select its correlated subspace
- Rank attribute-value pairs by **Anomaly Likelihood** (AL) score
  - Attribute values that occur very frequently and within a homogenous dimension have high AL scores
  - AL = (Frequency of Attribute-Value) x (Entropy of Attribute)$^{-1}$
- Select the top few and form the candidate subspace

Attribute Value	Frequency	AL Score
Income = 45k–60k	3	∞
Education = Highschool	1	1.58
Education = College	1	1.58
Education = Graduate	1	1.58

Education	Income	$S[1]$	$S[2]$	$S[3]$
Highschool	45k–60k	None	Magnitude	Magnitude
College	35k–45k	Phase	None	Misc
College	45k–60k	Phase	Magnitude	Magnitude
Graduate	45k–60k	None	Magnitude	Magnitude

Table 4: Time Anomaly Matrix

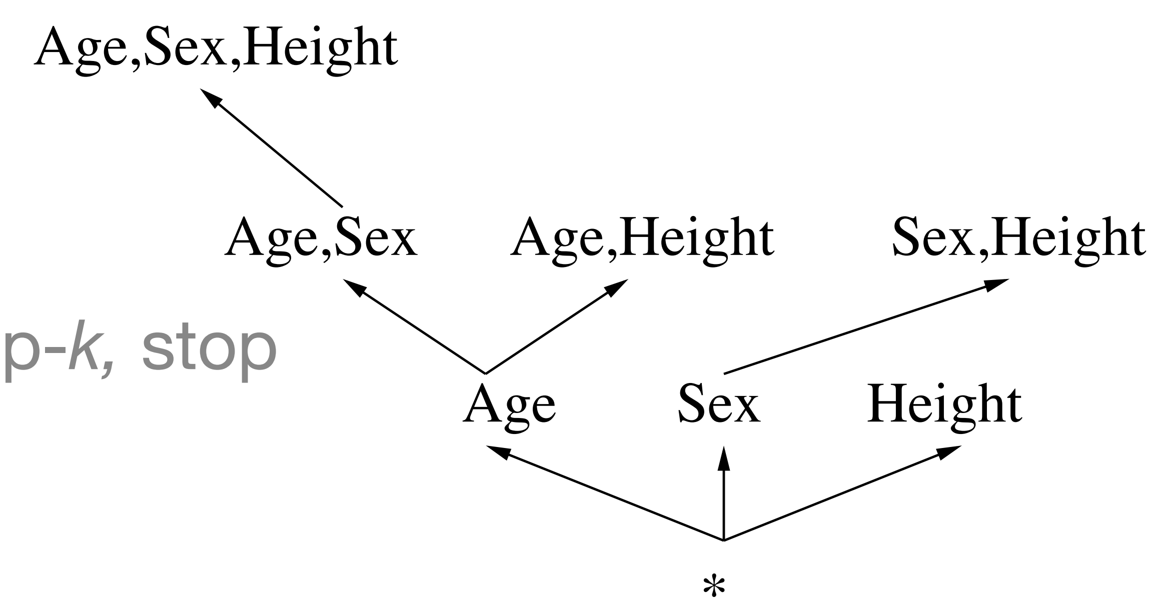
Table of Contents

Discovering Top-K Anomaly Cells

- Each subspace is small enough (~5 dimensions) for full cube materialization
- Efficient Regression Calculation
  - **Linear regression** needed for anomaly calculation (comparisons between parameters of observed and expected time series regression)
  - Regression parameters can be **aggregated losslessly** [Chen VLDB'02]
  - Only need to perform regression calculation once in the base cuboid
  - Higher level cuboids' regression parameters can be calculated via simple aggregation

Discovering Top-*K* Anomaly Cells (2)

- More efficient top-*k* anomaly detection (i.e., avoid computing the whole data cube)
- Intuition: calculate anomaly upper bounds during cubing and prune branches if upper bound is below current top-*k*
- Procedure
  - Bottom-up cube calculation [Beyer SIGMOD'99]
  - Keep track of current top-*k*
  - Calculate anomaly upper bound
  - If upper bound is below the worst in top-*k,* stop



SUITS Algorithm in Summary

Algorithm 2 SUITS

Input & Output: Same as Algorithm 1

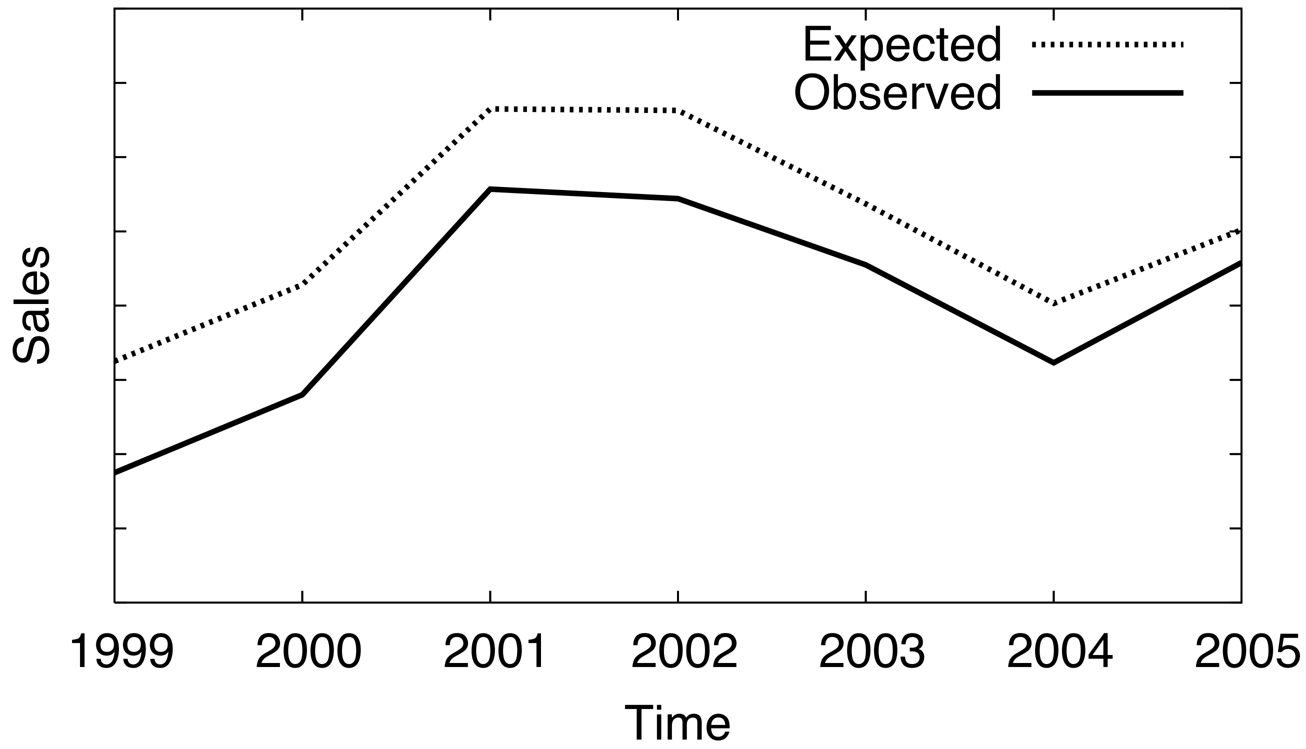
1. Retrieve data for $\sigma_p(R)$
2. Repeat until global answer set contains global top-k
3. $B \leftarrow$ candidate attribute values from $\{A_1, \ldots A_n\}$
4.   Retrieve top k anomaly cells from C_B using g and m
5.   Add top k cells to global answer set
6.   Remove discovered anomalies from input
7. Return top k cells in global answer set

- Final top-*k* is approximation of true global top-*k*
- Top-*k* pruning relies on monotonic properties of upper bound. If not satisfied, need to compute full subspace cube

Experiments

- Real market sales data from industry partner
- Time series data from 1999 to 2005
- Nearly 1 million sales and 600 dimensions

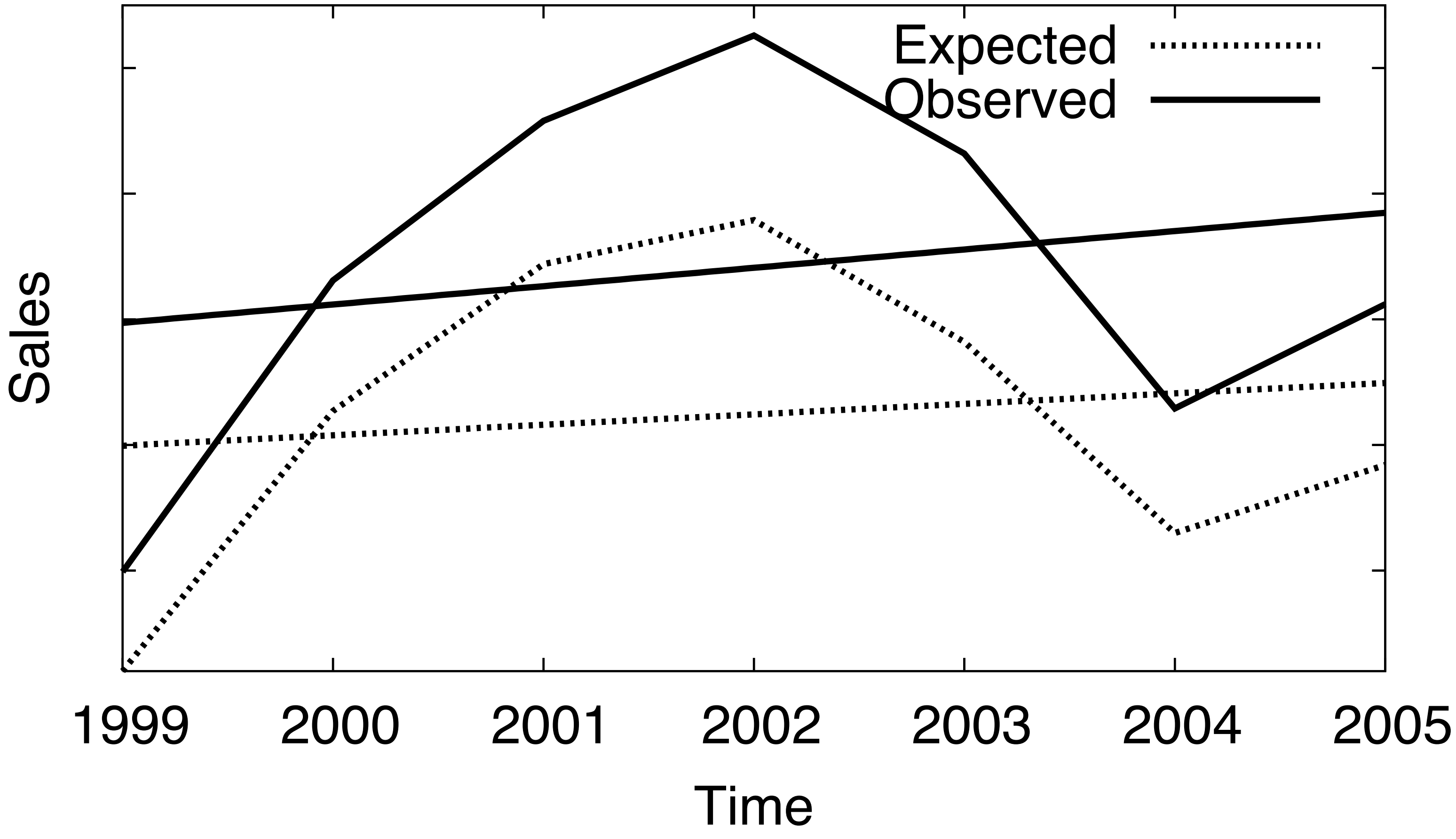
Sample Query 1



- **Probe**: Gender = “Male” ^ Marital = “Single” ^ Product = luxury item
- **Greatest anomaly**: Generation = “Post-Boomer” : less than expected
- **Explanation**: “Post-Boomer” are young and do not have enough money yet to purchase luxury item

Sample Query 2

- **Probe**: Gender = “Female” ^ Education = “Post-Graduate” ^ Product = cheap item
- **Greatest anomaly**:
  1. Employment = “Full-Time” ⇒ less
  2. Occupation = “Manager/Professional” ⇒ less
  3. <u>Number of Children Under 16 = 0 ⇒ more</u>
- **Explanation**: Number of Children Under 16 = 0 ⇔ “Young” ⇔ not enough accumulated wealth



Query Efficiency

Probe	$\|R\|$	Naïve	SUITS$_0$		SUITS		Common Top-10
		Time	Time	% Improve	Time	% Improve	
Male, Single	10	14	5.9	58%	5.4	61%	9
Male, Married	10	299	95	68%	60	80%	10
Male, Divorced	10	3.6	2.8	22%	2.8	22%	10
Female, Single	10	15	8.2	46%	7.0	53%	9
Female, Married	10	114	31.0	73%	23.0	80%	8
Female, Divorced	10	5.5	3.8	31%	3.7	33%	10
Post-Boomer, Children=0	11	68.8	39.6	43%	32.1	53%	10
Post-Boomer, Children=1	11	16.8	5.4	68%	4.8	71%	10
Post-Boomer, Children=2	11	15.5	7.8	50%	6.7	57%	10
Boomer, Children=0	11	108.9	75.7	30%	52.4	52%	10
Boomer, Children=1	11	120.3	68.9	43%	58.0	52%	10
Boomer, Children=2	11	46.6	27.2	42%	23.6	49%	10
			Average	48%		55%	9.6

Table 8: Run times of trend anomaly query with low dimensional data ($10 \le |R| \le 11$)

Dimensionality Efficiency

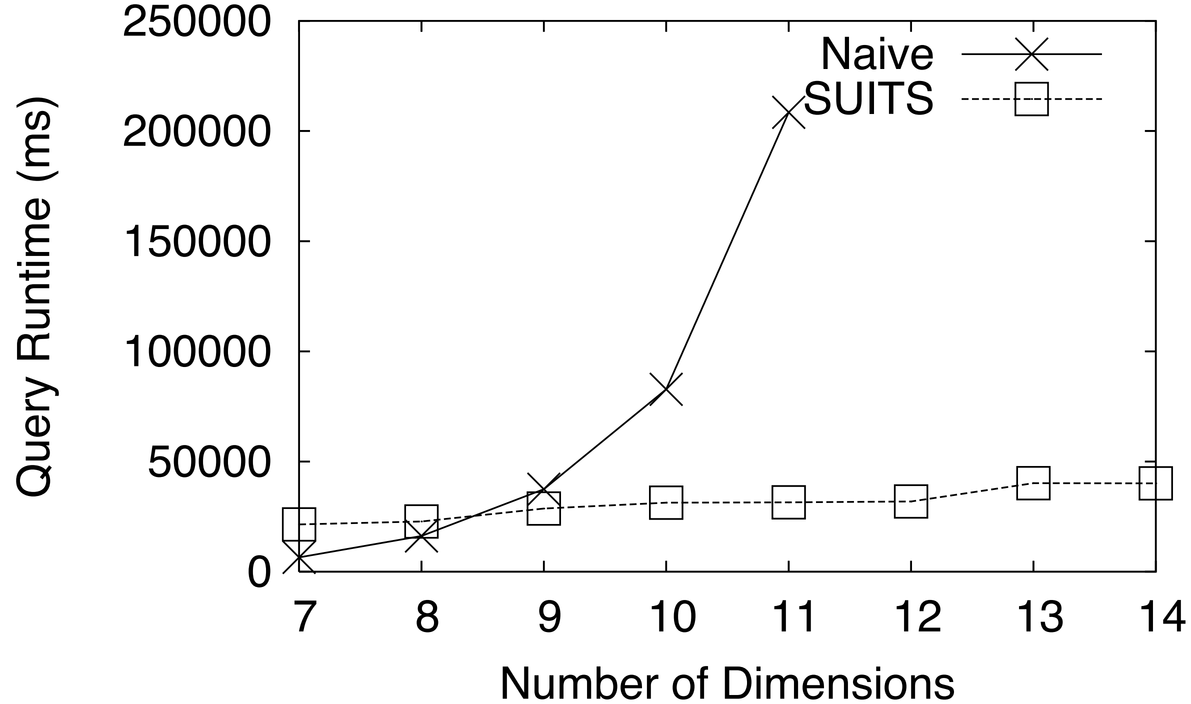


Figure 9: Running time vs. number of dimensions

Conclusion

- Detecting anomalies in data cube of time series data
- Iterative subspace search
- Efficient top-*k* anomaly detection
- Experiments with real data

Thank You!