# 05-2 Time

Monday, May 1, 2023 4:58 PM

# TIME SERIES VISUALIZATION AND ANALYSIS

Sungahn Ko
sako@unist.ac.kr

1

## Disclaimer

2

- The slides cannot be distributed, posted, or used outside of this class
- Slides in this course courtesy of
  - Dr. Abish Malik (Purdue)
  - Dr. Yun Jang (Sejong Univ.)
  - Dr. Ross Maciejewski (ASU)
  - Dr. Niklas Elmqvist (UMD)
  - Dr. David Ebert (Purdue)

HAiV

2

## Time-Oriented Data

3

- Time is an outstanding dimension reflected by Shneiderman's Task by Data Type Taxonomy[1]
- Time oriented data is ubiquitous,
  - stock markets, movie trends, business, medicine
- Each data case is likely an event of some kind,
  - with one variable being the date and time

W. Aigner, S. Miksch, W. Muller, H. Schumann, C. Tominski, "Visual Methods for Analyzing Time-Oriented Data", *IEEE Trans. On Visualization and Computer Graphics*, Vol. 14, No. 1, Jan.-Feb. 2008, pp. 47-60.

HAiV

3

## Time Series

4

- A random selection of 4000 graphs from 15 newspapers and magazines worldwide showed that between 1974 and 1980, 75% of these graphs were time series[1]
- What questions can we ask of these visuals?[2]
  - ???

1 – E. Tufte, "The Visual Display of Quantitative Information," 1983.
2 – A. MacEachren, "How Maps Work,: Representation, Visualization and Design," The Guilford Press, 1995.

HAiV

4

# Time

5

- Time is ?? ordered
  - We assume one time point precedes another
  - The notion of time being ordered is
    - closely bound to the notion of causality (things that happen close together in time are often related)
- Time is ?? continuous
  - We often break time into discrete components for analysis
  - Most people think of time as a flowing stream
- Time is ?? cyclical
- Time is ?? independent of location

L. Wilkinson, The Grammar of Graphics, *Springer*, 2005

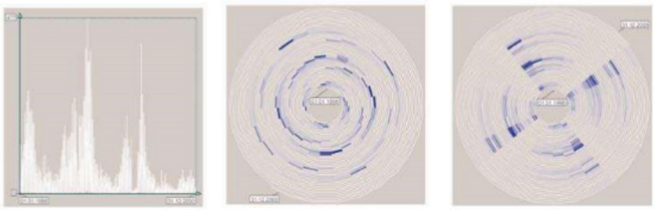
HAiV

5

# Linear vs. Cyclical Time

6

- Linear time assumes a starting point and defines the time domain with data elements from the past and future
- Natural processes may be cyclic (e.g., seasons)
  - The ordering of points in a cyclic time domain would be meaningless
  - Winter comes before summer, but also after summer



W. Aigner, S. Miksch, W. Muller, H. Schumann, C. Tominski, "Visual Methods for Analyzing Time-Oriented Data", *IEEE Trans. On Visualization and Computer Graphics*, Vol. 14, No. 1, Jan.-Feb. 2008, pp. 47-60.

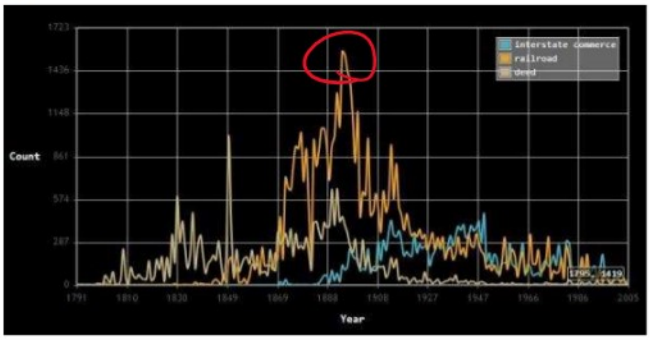
HAiV

6

# Linear Time

7

- Standard presentation
  - 2D line graph with time on x-axis and value on y-axis




**New Database Creates Time-Series Plots of Phrases in U.S. Supreme Court Opinions** December 16, 2011 | Posted by: Scott C. Idleman: http://law.marquette.edu/facultyblog/2011/12/16/new-database-creates-time-series-plots-of-phrases-in-u-s-supreme-court-opinions/

HAiV

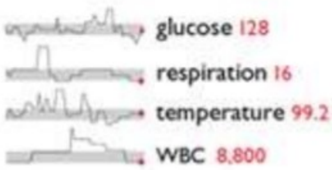
7

# Sparkline

Small line chart

8

- http://omnipotent.net/jquery.sparkline/#s-about
- "Intense, simple, wordlike graphics"



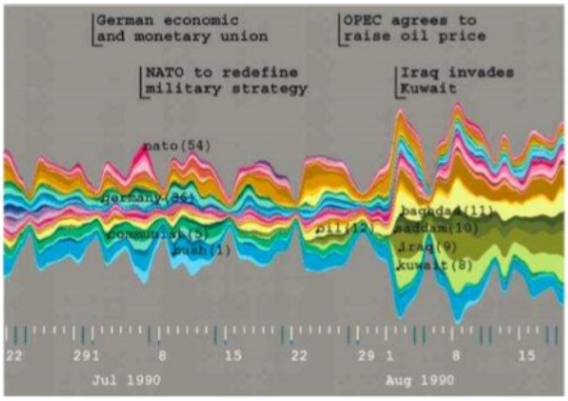

http://sparkline.org/

HAiV

8

# Theme River

9




1– S. Havre, E. Hetzler, P. Whitney, and L. Nowell. *Themeriver: Visualizing thematic changes in large document collections*. IEEE Transactions on Visualization and Computer Graphics 8(1):9-20, 2002.
* - Theme River Image from: http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html#

HAiV

9

# Linear Vs. Cyclical

10

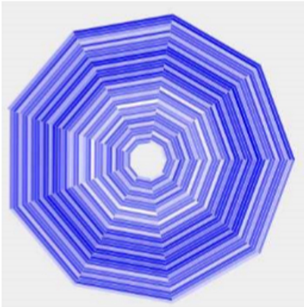
- Trends are easily seen in a linear plot, but what about repeating patterns?
  - One idea is to use spirals
    - as they easily represent the idea of repetition
  - Finding periodicity may be easier than a bar chart
    - because similar distances in a bar chart may look periodic when in fact they are not

HAiV

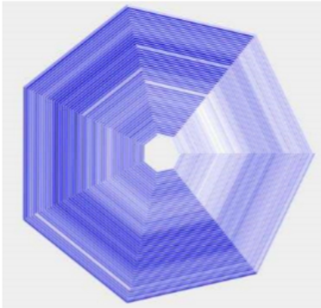
10

# Spiral Graph

11



9 Day Spiral



7 Day Spiral

Number of people born each day in 1978

1 - John V. Carlis and Joseph A. Konstan, *Interactive Visualization of Serial Periodic Data*, Proceedings User Interface Software and Technology (UIST), pp. 29-38, 1998
2 - M. Weber, M. Alexa, and W. Mu¨ ller, "Visualizing Time-Series on Spirals," Proceedings of the IEEE Symposium on Information Visualization (InfoVis '01), pp. 7-14, Oct. 2001.
3 - **Spirals for Periodic Data** *By Robert Kosara On August 7, 2011: http://eagereyes.org/techniques/spirals*

HAiV

11

# Spiral Graph

12

- Archimedean spiral – locus of points correspond to the locations over time of a point moving away from a fixed point

$$r = a + b\theta$$

- Logarithmic spiral – special kind of spiral appearing in nature

$$r = ae^{b\theta}$$

$$\theta = \frac{1}{b}\ln(\frac{r}{a})$$

- For spiral graphs, use Archimedean to map the same amount of data to the same angle

HAiV

12

# Calendar Based Visualization

13

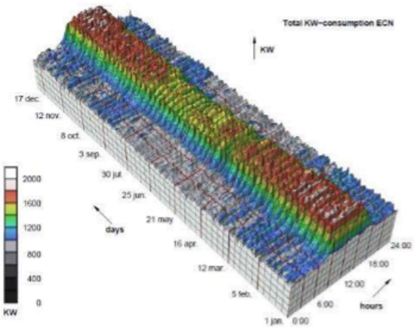



Figure 1. Power demand by ECN, displayed as a function of hours and days

1 – J. J. Van Wijk and E. R. Van Selow. *Cluster and calendar based visualization of time series data.* Proceedings of the 1999 IEEE Symposium on Information Visualization, 1999.

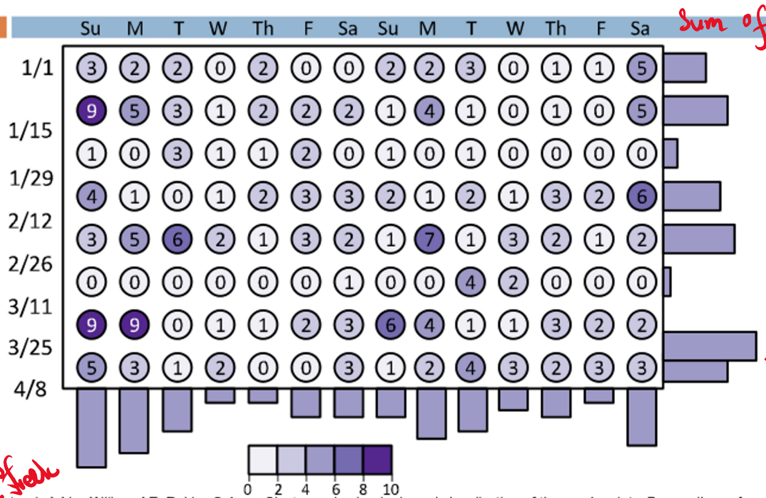
HAiV

13

# Calendar Based Visualization

14

Sum of rows




Sum of week
day per

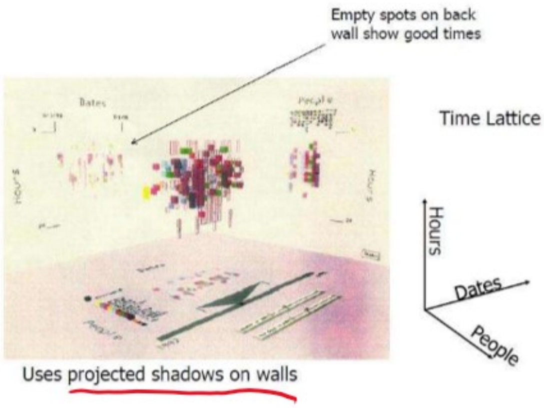
1 – J. J. Van Wijk and E. R. Van Selow. *Cluster and calendar based visualization of time series data.* Proceedings of the 1999 IEEE Symposium on Information Visualization, 1999.

HAiV

14

## Spiral Calendar

15



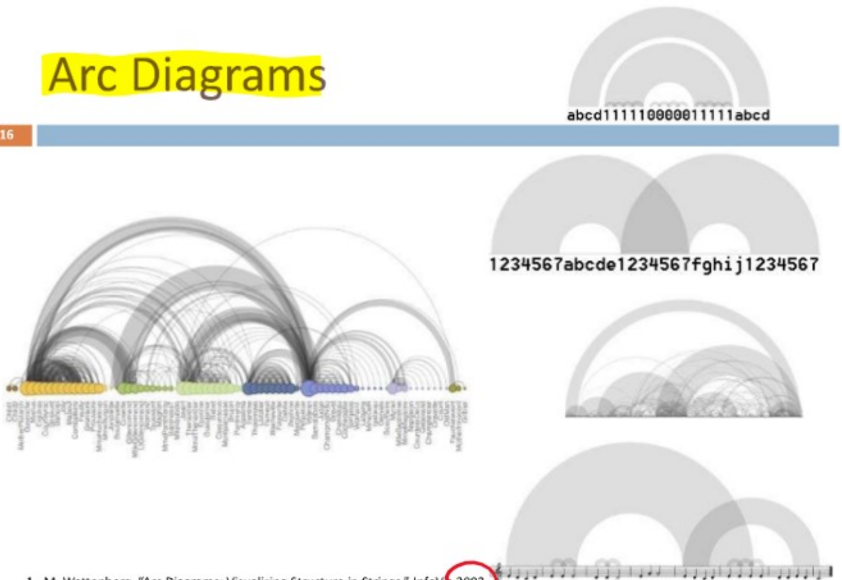

Mackinlay, J., Robertson, G. G., and DeLine, R., "Developing Calendar Visualizers for the Information Visualizer," UIST 1994.

HAiV

15

## Arc Diagrams

16




1– M. Wattenberg, "Arc Diagrams: Visualizing Structure in Strings," InfoVis 2002.
http://www.bewitched.com/song.html
http://www.turbulence.org/Works/song/gallery/gallery.html
https://www.bewitched.com/

HAiV

16

# Ordered Time vs. Branching Time

17

- Ordered time domains consider
  - things that happen one after another
- Branching time considers multiple what-if scenarios, allowing comparison of alternate scenarios
  - Can support decision making process where only one alternative is chosen

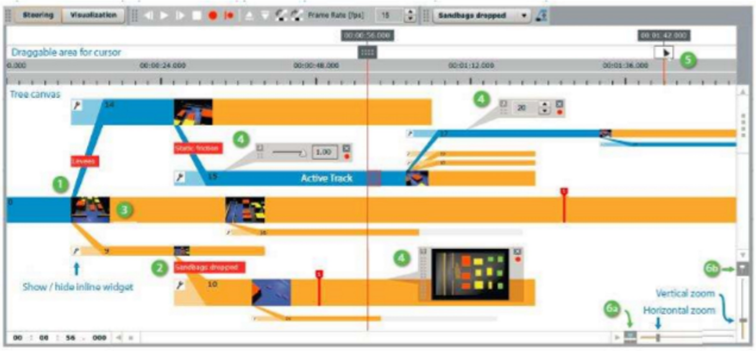
W. Aigner, S. Miksch, W. Muller, H. Schumann, C. Tominski, "Visual Methods for Analyzing Time-Oriented Data", *IEEE Trans. On Visualization and Computer Graphics*, Vol. 14, No. 1, Jan.-Feb. 2008, pp. 47-60.

HAiV

17

# World Lines

18




Creating new simulation tracks through temporal branching

http://visdom.at/media/videos/mp4/world_lines_1.mp4

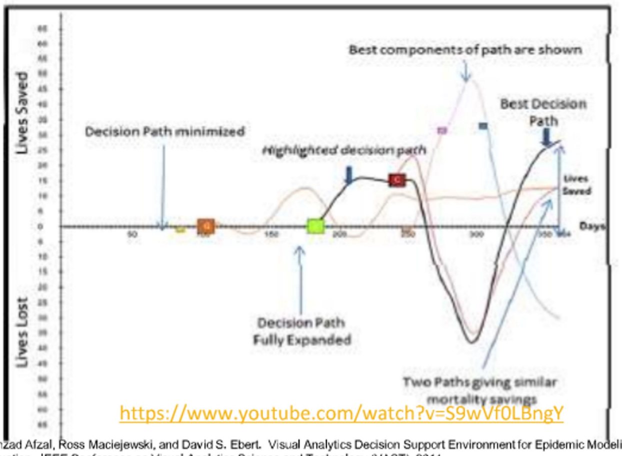
Waser, J.; Fuchs, R.; Ribičič, H.; Schindler, B.; Blöschl, G.; Gröller, E.; , "World Lines," *IEEE Transactions on Visualization and Computer Graphics*, vol.16, no.6, pp.1458-1467, Nov.-Dec. 2010

HAiV

18

## Branching for What-if, Decision Support

19




https://www.youtube.com/watch?v=S9wVf0LBngY

Shehzad Afzal, Ross Maciejewski, and David S. Ebert. Visual Analytics Decision Support Environment for Epidemic Modeling and Response Evaluation. IEEE Conference on Visual Analytics Science and Technology (VAST), 2011.

HAiV

19

☆☆ Until here

## Design Principles

20

- Show familiar visual representations whenever possible
- Provide side-by-side comparisons of small multiple views
- Spatial position is strongest visual cue
- Multiple views are more effective when coordinated through explicit linking
- Avoid abrupt visual change

Borrowed from John Stasko's Time Series Data lecture:
http://www.cc.gatech.edu/~stasko/7450/Notes/time.pdf

HAiV

20

## Time-Series Data Mining

22

- Data mining domain has techniques for examining time series, looking for patterns, looking for anomalies
- Can be used to enhance the visualizations, show us what is important
- Can also be used in exploratory analysis,
  - "I think this looks interesting, show me similar trends."

HAiV

22

## Typical Time Series Analysis

23

- Identifying Patterns
  - Trend analysis
    - A company's linear growth in sales over the years
  - Seasonality
    - Sales are higher in summer than winter
  - Forecasting
    - What is the expected sales of next quarter

HAiV

23

## Temporal Analysis Methods

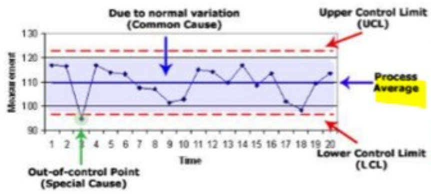
24

- For temporal data, we can find anomalies using control chart methods
- Control charts consist of a statistic representing some measurement in time (number of patients in a hospital, value of a stock, etc.)
- The mean and standard deviation of the statistic is calculated given all the available samples
- If the current value is greater than some pre-set number of standard deviations from the mean,
  - then an alert is generated (Shewart suggested three standard deviations)

HAiV

24

## Control Chart Method




Thresholds

25

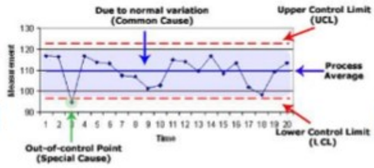
### The control chart

- A graph used to study how a process changes over time. Data are plotted in time order.
- A central line for the average, an upper line for the upper control limit and a lower line for the lower control limit.
- Lines are determined from historical data. By comparing current data to these lines, you can draw conclusions about
  - whether the process variation is consistent (in control)
  - or is unpredictable (out of control, affected by special causes of variation).

HAiV

25

## Control Chart Meth




26

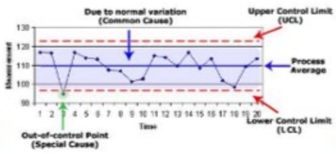
When to use a control chart?

- Controlling ongoing processes by finding and correcting problems as they occur.
- Predicting the expected range of outcomes from a process.
- Determining whether a process is stable (in statistical control).
- Analyzing patterns of process variation from special causes (non-routine events) or common causes (built into the process).
- Determining whether the quality improvement project should aim to prevent specific problems or to make fundamental changes to the process.

HAiV

26

## Control Chart Model




27

- General model for a control chart
  - Upper Control Limit = $\mu + k\sigma$
  - Center Line = $\mu$
  - Lower Control Limit = $\mu - k\sigma$

where $\mu$ is the mean of the variable, and $\sigma$ is the standard deviation of the variable.

When $k$ is set to 3, we speak of 3-sigma control charts. Historically, $k = 3$ has become an accepted standard in industry.

HAiV

27

## Moving Average/Range Charts

28

- Moving Average/Range Charts are
  - a set of control charts for variables data.
- The Moving Average chart monitors
  - the process location over time, based on the average of the current subgroup and one or more prior subgroups.
- Moving Average Charts are generally used for
  - detecting small shifts in the process mean.
  - They will detect shifts of .5 sigma to 2 sigma much faster. They are, however, slower in detecting large shifts in the process mean.

HAiV

28

## Moving Average

29

- A simple moving average (SMA) is formed by computing the average price of a measure over a specific number of periods.
- In this example we'll look at stock market closing prices
- A 5-day simple moving average is the five day sum of closing prices divided by five.
- As its name implies, a moving average is an average that moves. Old data is dropped as new data comes available.
- This causes the average to move along the time scale. Below is an example of a 5-day moving average evolving over three days.

HAiV

29

## Moving Average

30

- Daily Closing Prices: (11,12,13,14,15),16,17
  - First day of 5-day SMA: (11 + 12 + 13 + 14 + 15) / 5 = 13
  - drop 11
  - Second day of 5-day SMA: (12 + 13 + 14 + 15 + 16) / 5 = 14
  - drop 12
  - Third day of 5-day SMA: (13 + 14 + 15 + 16 + 17) / 5 = 15

HAiV

30

## Exponentially Weighted Moving Average

31

- Exponential moving averages reduce
  - the lag by applying more weight to recent prices.
- The weighting applied to the most recent price depends on the number of periods in the moving average.
- There are three steps to calculating an exponential moving average.
  - First, calculate the simple moving average.
  - An exponential moving average (EMA) has to start somewhere so a simple moving average is used as the previous period's EMA in the first calculation.
  - Second, calculate the weighting multiplier.
  - Third, calculate the exponential moving average. The formula below is for a 10-day EMA.

HAiV

31

## Exponentially Weighted Moving Average

32

- SMA (Simple Moving Average): 10 period sum / 10
- Multiplier: (2 / (Time periods + 1) ) = (2 / (10 + 1) ) = 0.1818 (18.18%)
- EMA: {Close - EMA(previous day)} x multiplier + EMA(previous day).

HAiV

32

## Finding Similarities in Time Series

33

- Indexing problem
  - Find all stocks whose fluctuations are similar to stock X
- Subsequence Similarity
  - Find out other days in which stock x had similar movements as today
- Clustering Group them based on some criteria
  thousands of stocks
  - Group regions that have similar temporal patterns
- Rule Discovery
  - Find rules like if x goes up, y goes up, etc.

HAiV

33

## Indexing Approach

34

- Extract a few key features for each time series (think PCA)
- Map each time sequence X to a point f(X) in the lower dimensional space
- This allows for faster indexing for finding similarities
- Research problem is to design solutions that are both accurate and efficient

HAiV

34

## Finding Similarities

35

- Brute force search:
  - Given query time series Q, the best match by sequential scanning is found by:

$$\min_{1 \leq i \leq N} \sum_{t=1}^{d} (X_i(t) - Q(t))^2$$

  - Method is O(nd)

HAiV

35

## Euclidean Similarity Measure

36

- View each time series as a point in n-dimensional space
- Calculate the "distance" between each point
- Easy to compute
- Scalable solution to other problems

HAiV

36

## Normalization of Sequences

37

- Normalize the mean and variance for each sequence
- Let $\mu(x)$ and $\sigma(x)$ by the mean and variance of time series X
- F(x) = ($x_i$- $\mu(x)$)/ $\sigma(x)$
- Use normalized sequence for comparisons

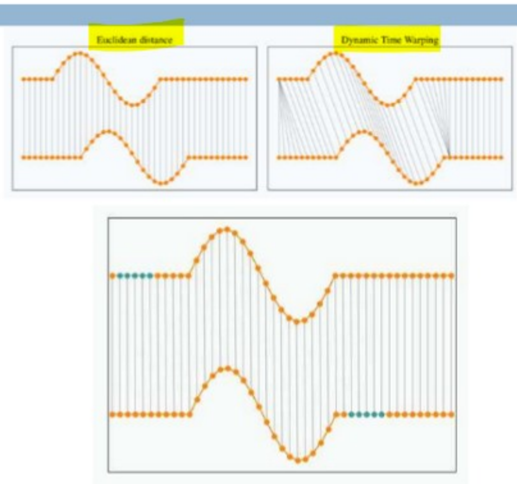
HAiV

37

# Dynamic Time Warping (DTW)

38

To find a time shift that maximize the similarity




Source: https://rtavenar.github.io/blog/fig/dtw_path.webm

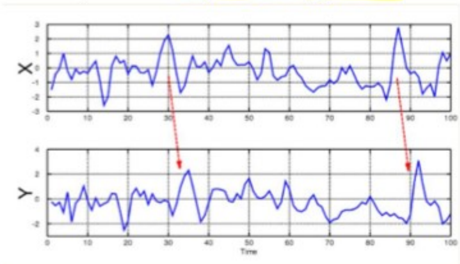
HAiV

38

# Granger Causality Test

39

- Well known temporal dependency investigation method



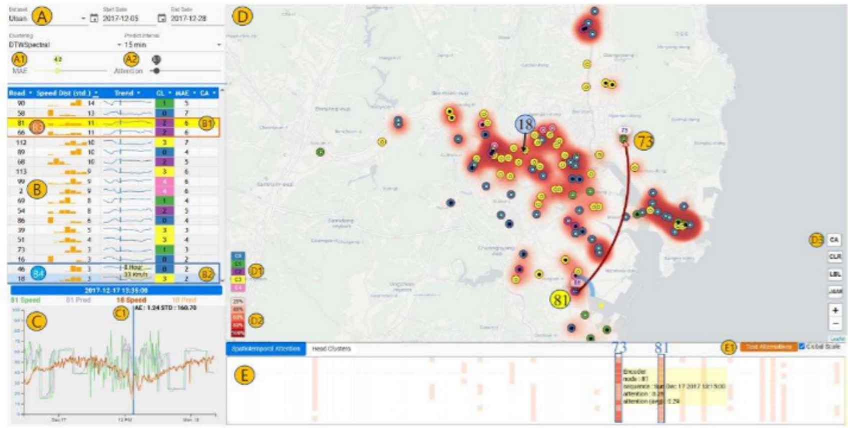

- Two principles
  - The cause
    - Happens prior to its effect
    - Has unique information about the future values of its effect

HAiV

39

## Example Use of DTW and Causality Test

Traffic Congestion



Jin et al., "A Visual Analytics System for Improving Attention-based Traffic Forecasting Models", TVCG 2023

40

## Example Use of DTW and Causality Test

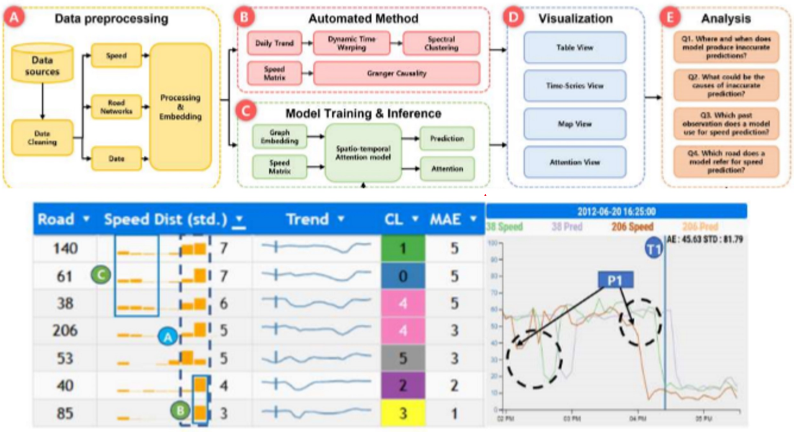



Fig. 9. Summary of tables and Causality Analysis in LA. Overall there is a positive correlation between std. and MAE. Road 206's traffic pattern precedes that of road 38 (line chart, P1).

41

# Readings

42

- Required Reading:
  - W. Aigner, S. Miksch, W. Muller, H. Schumann, C. Tominski, "Visual Methods for Analyzing Time-Oriented Data", *IEEE Trans. on Visualization and Computer Graphics*, Vol. 14, No. 1, Jan.-Feb. 2008, pp. 47-60.
  - L. Byron & M. Wattenberg, "Stacked Graphs - Geometry & Aesthetics", *IEEE Trans. on Visualization and Computer Graphics*, Vol. 14, No. 6, Nov.-Dec. 2008, pp. 1245-1252.
  - Kincaid, R.; , "SignalLens: Focus+Context Applied to Electronic Time Series," *IEEE Transactions on Visualization and Computer Graphics*, vol.16, no.6, pp.900-907, Nov.-Dec. 2010

HAiV

42