Data Visualization Time series

Vaibhav P. Vasani
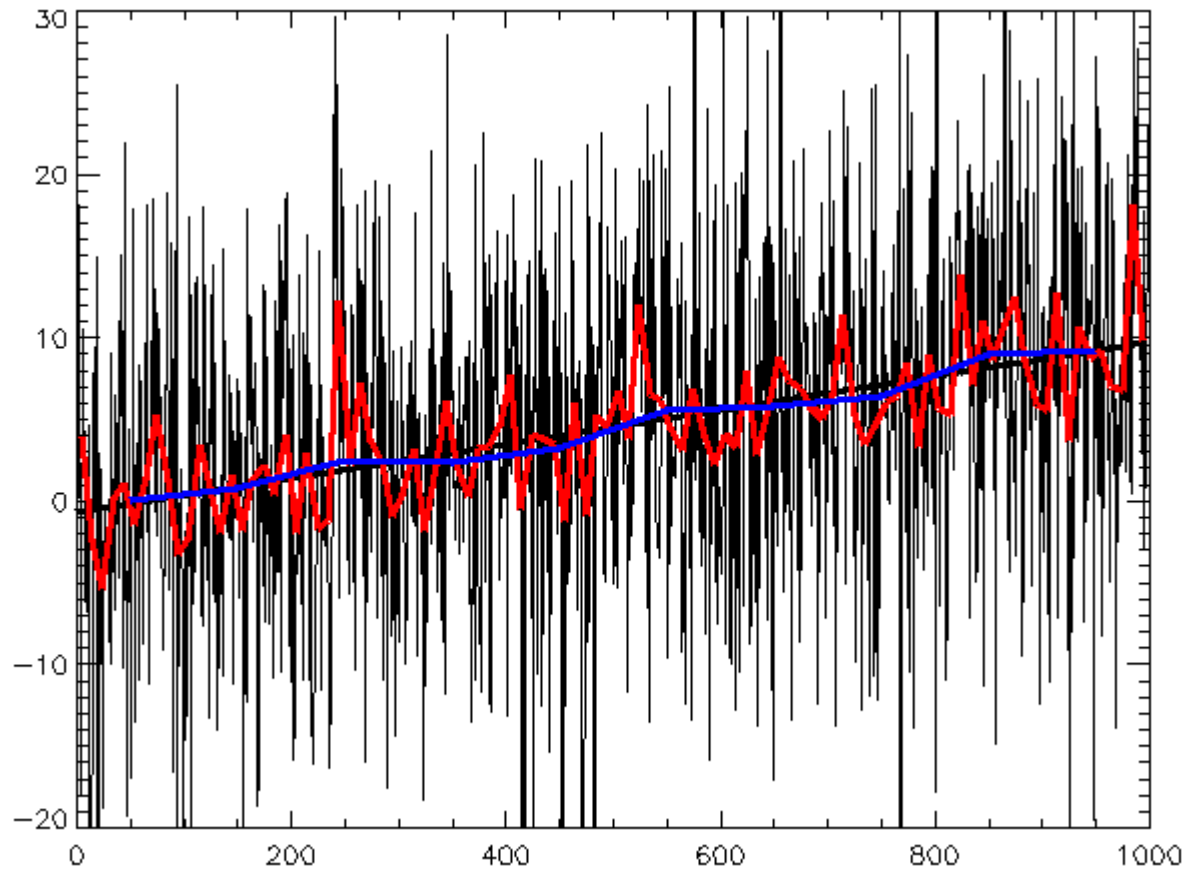
Assistant Professor

Department of Computer Engineering

K. J. Somaiya College of Engineering

Somaiya Vidyavihar University

vaibhav.vasani@gmail.com

Time series

A straightforward way to examine a regular time series is manually with a line chart.

- In mathematics, a time series is a series of data points indexed (or listed or graphed) in time order.

- Most commonly, a time series is a sequence taken at successive equally spaced points in time.

- It is a sequence of discrete-time data.

- Examples of time series are heights of ocean tides, counts of sunspots, and the daily closing value of the Industrial Average.

Web Traffic Time-series data

10am	11am	12pm	1pm	2pm	3pm
T1	T2	T3	T4	T5	T6
300	310	420	530	640	?

- A different person can have a different perspective like one can say find the mean of all observations, one can have like take mean of recent two observations, one can say like give more weightage to current observation and less to past, or one can say use interpolation.

- There are different methods to forecast the values.

- While Forecasting time series values, 3 important terms need to be taken care of and the main task of time series forecasting is to forecast these three terms.

- **1) Seasonality**
  - Seasonality is a simple term that means while predicting a time series data there are some months in a particular domain where the output value is at a peak as compared to other months. for example if you observe the data of tours and travels companies of past 3 years then you can see that in November and December the distribution will be very high due to holiday season and festival season. So while forecasting time series data we need to capture this seasonality.

- **2) Trend**
  - The trend is also one of the important factors which describe that there is certainly increasing or decreasing trend time series, which actually means the value of organization or sales over a period of time and seasonality is increasing or decreasing.

- **3) Unexpected Events**
  - Unexpected events mean some dynamic changes occur in an organization, or in the market which cannot be captured.
  - for example a current pandemic we are suffering from, and if you observe the Sensex or nifty chart of march 2020 there is a huge decrease in stock price which is an unexpected event that occurs in the surrounding.

- A stationary time series is a data that has a constant mean and constant variance.

- If I take a mean of T1 and T2 and compare it with the mean of T4 and T5 then is it the same, and if different, how much difference is there? So, constant mean means this difference should be less, and the same with variance.

- If the time series is not stationary, we have to make it stationary and then proceed with modelling. Rolling statistics is help us in making time series stationary.

- Rolling statistics calculates moving average.

- To calculate the moving average we need to define the **window size** which is basically how much past values to be considered.

- For example, if we take the window as 2 then to calculate a moving average in the above example then, at point T1 it will be blank, at point T2 it will be the mean of T1 and T2, at point T3 mean of T3 and T2, and so on. And after calculating all moving averages if you plot the line above actual values and calculated moving averages then you can see that the plot will be smooth.
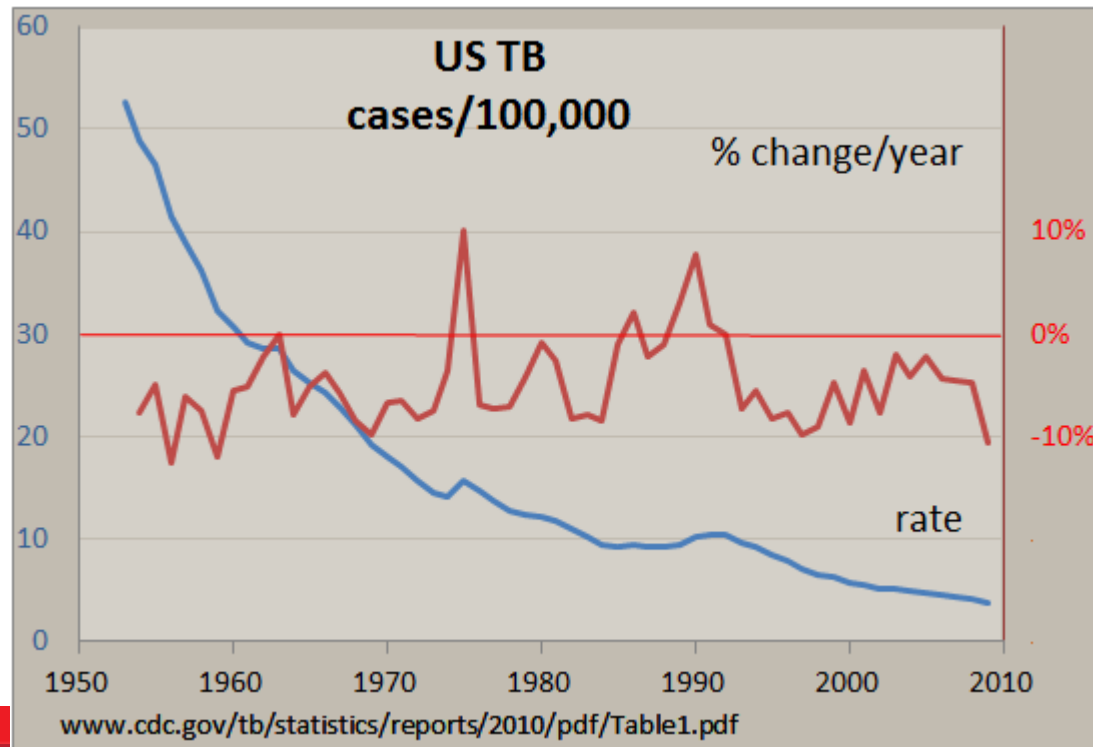
Additive and Multiplicative Time series

Additive Time Series

Value = Base Level + Trend + Seasonality + Error

Multiplicative Time Series

Value = Base Level * Trend * Seasonality * Error

- An example chart is shown on the right for tuberculosis incidence in the United States, made with a spreadsheet program. The number of cases was standardized to a rate per 100,000 and the percent change per year in this rate was calculated.

- The nearly steadily dropping line shows that the TB incidence was decreasing in most years, but the percent change in this rate varied by as much as +/- 10%, with 'surges' in 1975 and around the early 1990s. The use of both vertical axes allows the comparison of two time series in one graphic.

US TB cases/100,000

% change/year

rate

www.cdc.gov/tb/statistics/reports/2010/pdf/Table1.pdf