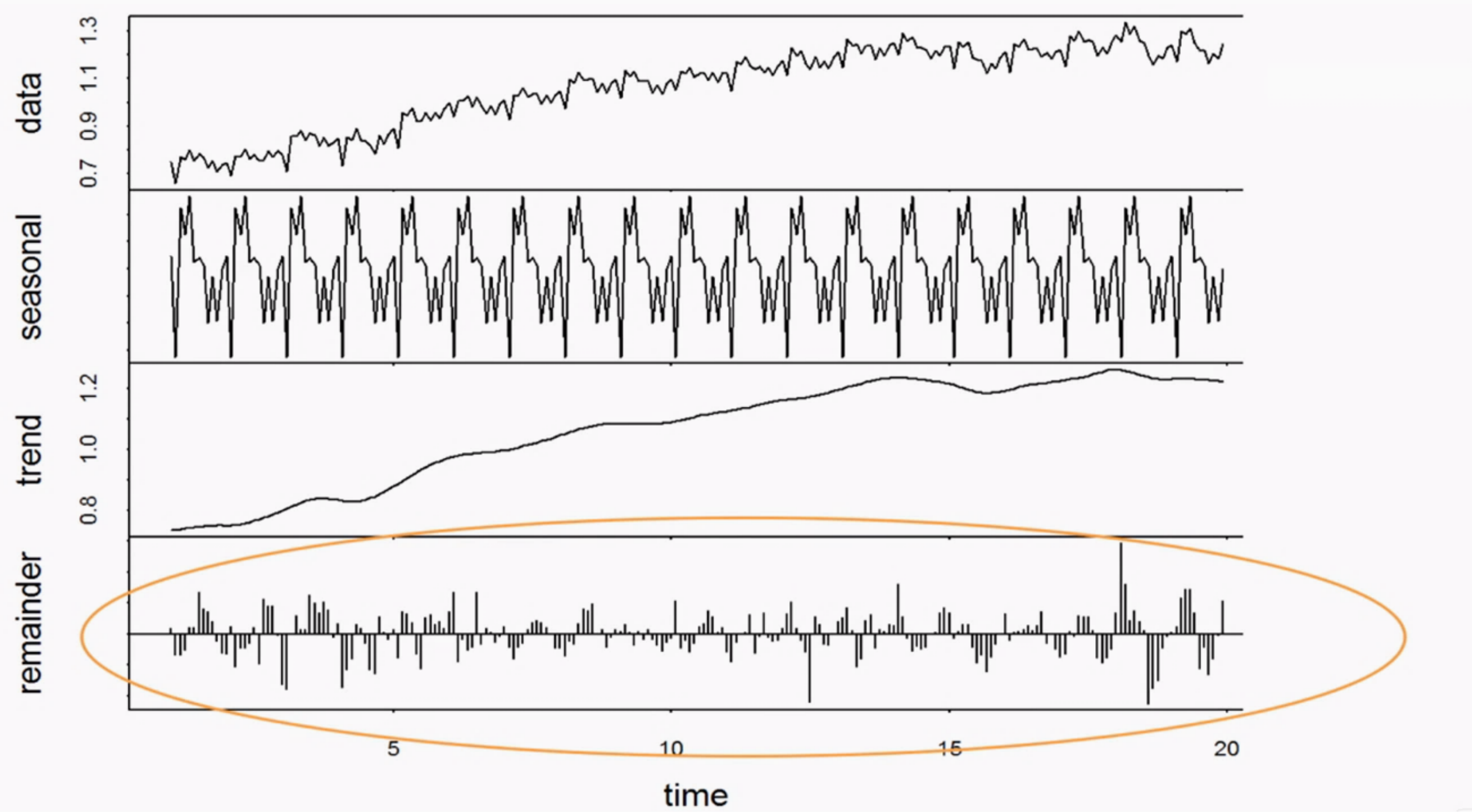
Enhanced Seasonal Decomposition of Time Series 📈 by Loess

R package ‘**stlplus**’ Date 2016-01-05

Description: Decompose a time series into seasonal, trend, and remainder components using an implementation of Seasonal Decomposition of Time Series by Loess (STL) that provides several enhancements over the STL method in the stats package. These enhancements include handling missing values, providing higher order (quadratic) loess smoothing with automated parameter choices, frequency component smoothing beyond the seasonal and trend components, and some basic plot methods for diagnostics.

STL package (abbreviated)

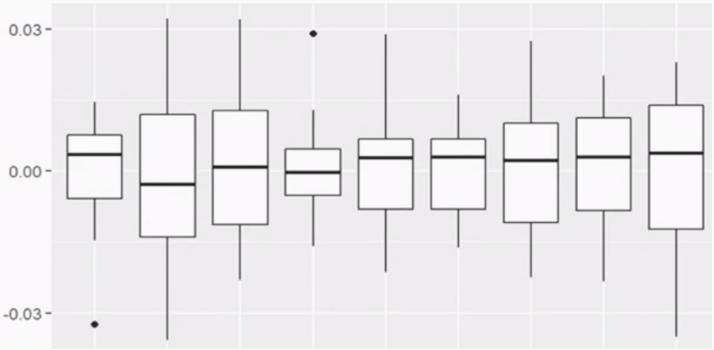
- uses loess to find a general trend T
- then subtacts the trend: $X\text{-}T$
- then uses moving average smoothing to find more trend C
- then subtracts the moving average trend: $S=X\text{-}T\text{-}C$
- the seasonal componenet, S remains (seasonal because the stationary signals are smoothed and the trend is subtracted)
- then smooths the total trend $X\text{-}S$ with loess to get total trend V (seasonal signal is removed from the original data, leaving the sum of the two remaining prior two trends $X\text{-}S = T+C$)
- the remainder is $R=X\text{-}S\text{-}V$ (remove the seasonal component, the trend V, and the remainder R remains from the original signal)

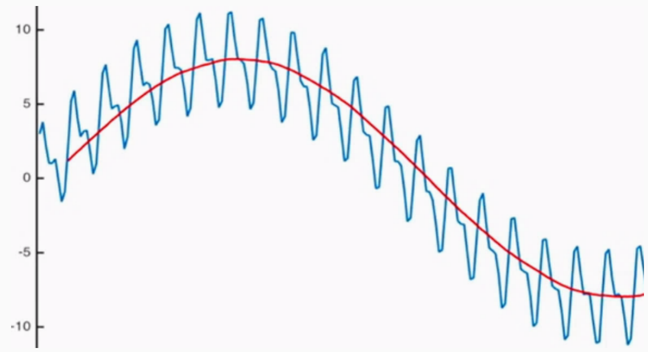


defining stationary timeseries data

- stationary timeseries have no trend
- second-order stations conditions:
  - constant mean (does not change over time)
  - constant variance (does not change over time)
  - an autocovariance that is not time-dependent

boxplots (left) and trend models (right) are both usefull for considering stationarity:





variations in box plat variances/means and existacne of a trend in a trend model indicates stationarity

Autocovariance and Autocorrelation (ACF) ("Auto" = "Self")

$$\text{Autocorrelation}_h(X_t) = \frac{\text{Autocovariance}_h(X_t)}{\text{Std}(X_t)\text{Std}(X_{t-h})}$$

$$\text{Autocovariance}_h(X_t) = Cov(X_t, X_{t-h})$$

Autocovariance is convariance of timeseries with lagged version of itself.

for clarity, the covariance of (X_t, X_{t-h}) is X a time t and a lagged version of itself $t-h$ with h functioning as the lag. X_t is a random variable true for all aspects of timeseries data. Autocorrelation is simply a normalized vertsion of the autocovariance so that the lowest value is 1.

If the signal is weakly stationary, then the varaince (and std) are constant over time:

$$\text{Autocorrelation}_h(X_t) = \frac{\text{Autocovariance}_h(X_t)}{\text{Std}(X_t)\mathbf{Std}(\boldsymbol{X_{t-h}})} \rightarrow \text{Autocorrelation}_h(X_t) = \frac{\text{Autocovariance}_h(X_t)}{\text{Std}(X_t)\mathbf{Std}(\boldsymbol{X_t})}$$

and simplifies as: $$\text{Autocorrelation}_h(X_t) = \frac{\text{Autocovariance}_h(X_t)}{\text{Var}(X_t)}$$

furthermore, if the signal is weakly stationary, autocorrelation is constant over time. The latter property allows a sample to be taken from the data due to its lack of time dependency; therefore:

$$\mathrm{Autocorrelation}_h(X_t) = \frac{\mathrm{Autocovariance}_h(X_t)}{\mathrm{Var}(X_t)}$$

$$\mathrm{SampleAutocorr}(X) = \frac{\mathrm{SampleCov}([x_h, \ldots, x_t], [x_1, \ldots, x_{t-h}])}{\mathrm{SampleVar}([x_1, \ldots, x_t])}$$

the above illustrates the use of Sample Covariance and Sample Variance in place of the Autocovariance and variance as defined formally. To compute the Sample Convariance, it is noted in the expression that the timeseries and a lagged version of itself is necessary to compute Sample Autocorrelation in the resulting depictions illustrated below: