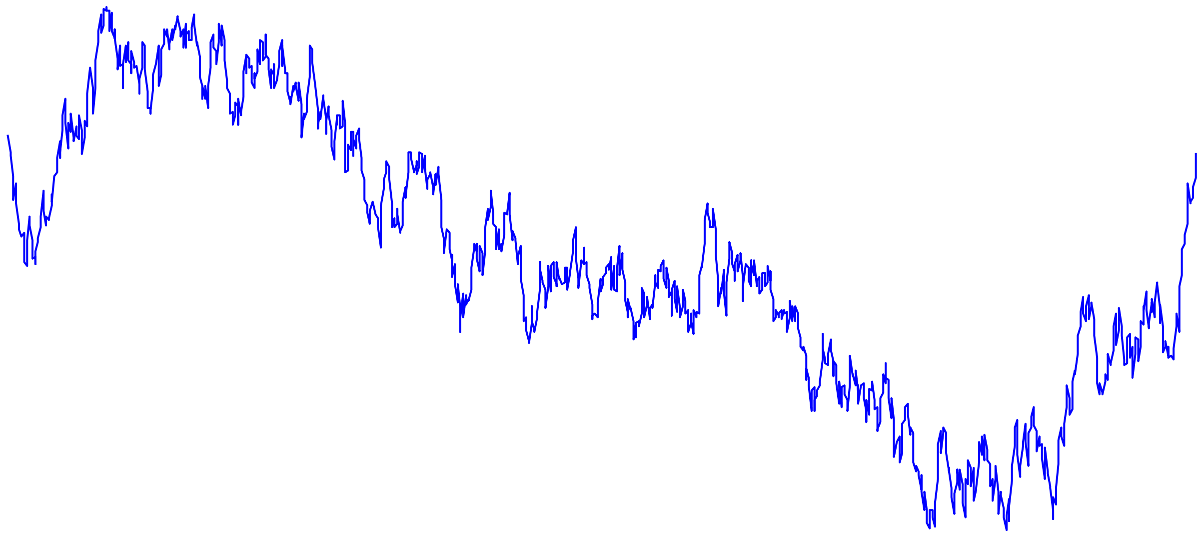
# Introduction and Preprocessing

Francesco Spinnato, Riccardo Guidotti

# Syllabus

- Time Series, Datasets and Problems
- Missing Values
- Anomalies
- Normalizations
- Components
- Stationarity
- Autocorrelation



# What is Time Series Analysis?

- Time Series Analysis is the **study of data points collected over time**, typically at consistent intervals across a defined period. Unlike random or intermittent data collection, **TSA treats time as a central variable**.

- TSA is useful because it shows **how variables evolve over time** and captures dependencies between observations. Since each data point may be related to previous values, **the temporal order of the data matters**.

- **A reliable TSA usually requires many observations**. A larger sample helps reduce the effect of noise, improves consistency, and makes it easier to identify trends, patterns, and seasonal variation.

- Common applications include **forecasting future values** from historical data and using past time-dependent patterns for **classification or regression tasks,** including predicting exogenous variables.

# Why Companies Benefits From TSA?

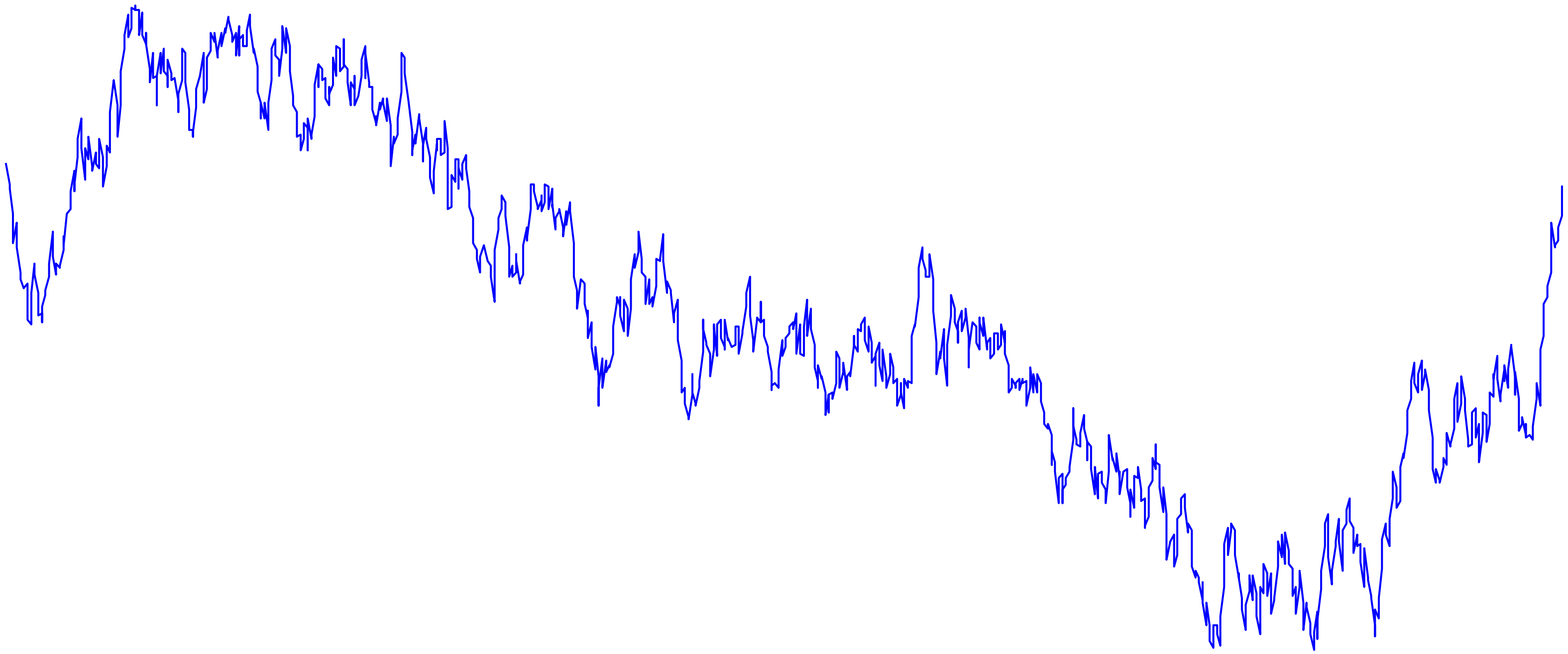
- Time Series Analysis helps **identify underlying trends, systemic patterns, and changes over time**. By using visualizations, analysts can detect seasonal trends, cyclic behavior, and **possible causes behind observed patterns**. Modern analytics platforms also support more advanced visualizations beyond basic line graphs.

- **TSA is especially useful for prediction**. Because it analyzes data collected at consistent intervals, it can support forecasting, estimate the likelihood of future events, and improve understanding of recurring patterns such as seasonality and cycles.

- **Modern technology has made TSA more practical and powerful**. Organizations can now collect large volumes of data daily, making it easier to build consistent datasets for comprehensive and reliable analysis.

# What is a Time Series?

A time series $[x_1, \ldots, x_T]$ is a collection of $T$ observations $x_i$ made sequentially in time, generally at constant time intervals. $[t_1, \ldots, t_T]$ is the timestamp vector.



| |
|---|
| 25.1750 |
| 25.2250 |
| 25.2500 |
| 25.2500 |
| 25.2750 |
| 25.3250 |
| 25.3500 |
| 25.3500 |
| 25.4000 |
| 25.4000 |
| 25.3250 |
| 25.2250 |
| 25.2000 |
| 25.1750 |
| … |
| 24.6250 |
| 24.6750 |
| 24.6750 |
| 24.6250 |
| 24.6250 |
| 24.6250 |
| 24.6750 |
| 24.7500 |

# Series in general

Time series concern time, but in general a series of observations is taken at defined steps along any ordered dimension:

- Time: Milliseconds, Seconds, Minutes, Hours, Days, Months, Years, …
- Spatial: Locations, Positions, …
- Relative: one cm left, two cm left, …

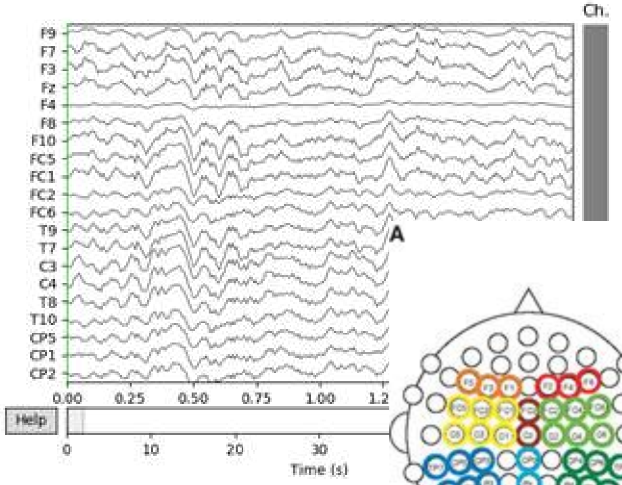
The important point is to have an ordered variable that provides a direction for its values.

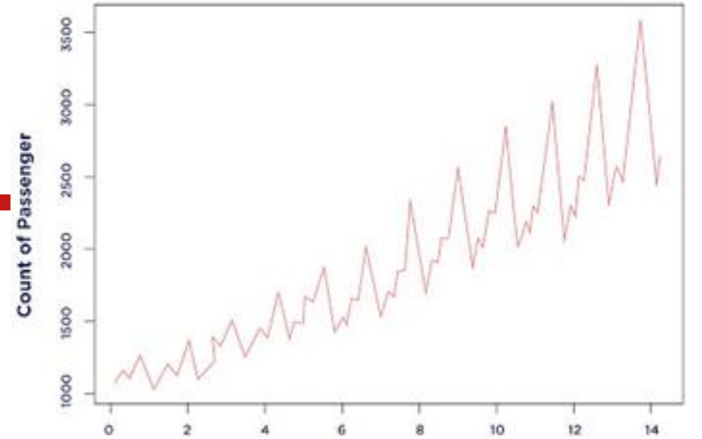
Then from a given observation $x_i$ it is easy to identify the "past", i.e., what came before $x_i$, and the "future", i.e., what comes after $x_i$

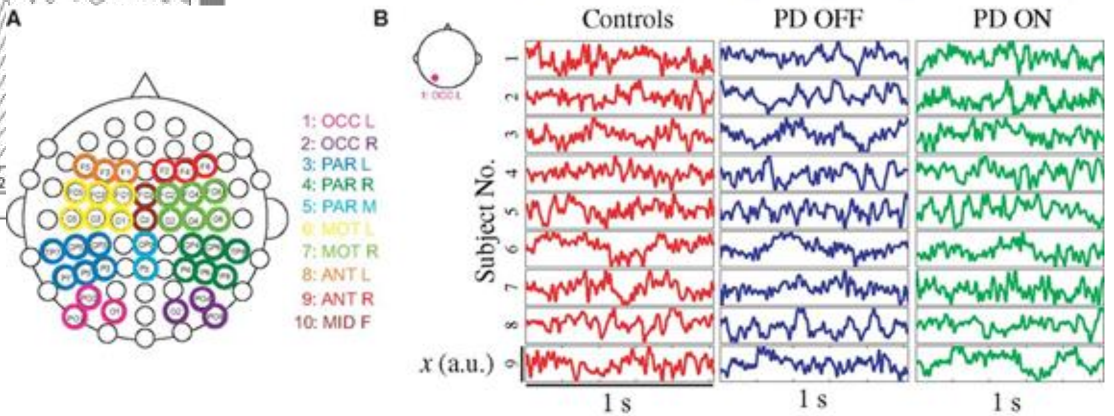
| |
|---|
| 25.1750 |
| 25.2250 |
| 25.2500 |
| 25.2500 |
| 25.2750 |
| 25.3250 |
| 25.3500 |
| 25.3500 |
| 25.4000 |
| 25.4000 |
| 25.3250 |
| 25.2250 |
| 25.2000 |
| 25.1750 |
| ... |
| 24.6250 |
| 24.6750 |
| 24.6750 |
| 24.6250 |
| 24.6250 |
| 24.6250 |
| 24.6750 |
| 24.7500 |

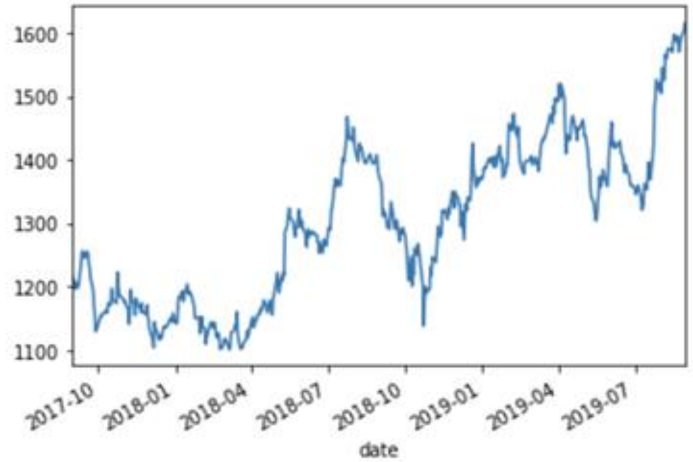
# Time Series are Ubiquitous

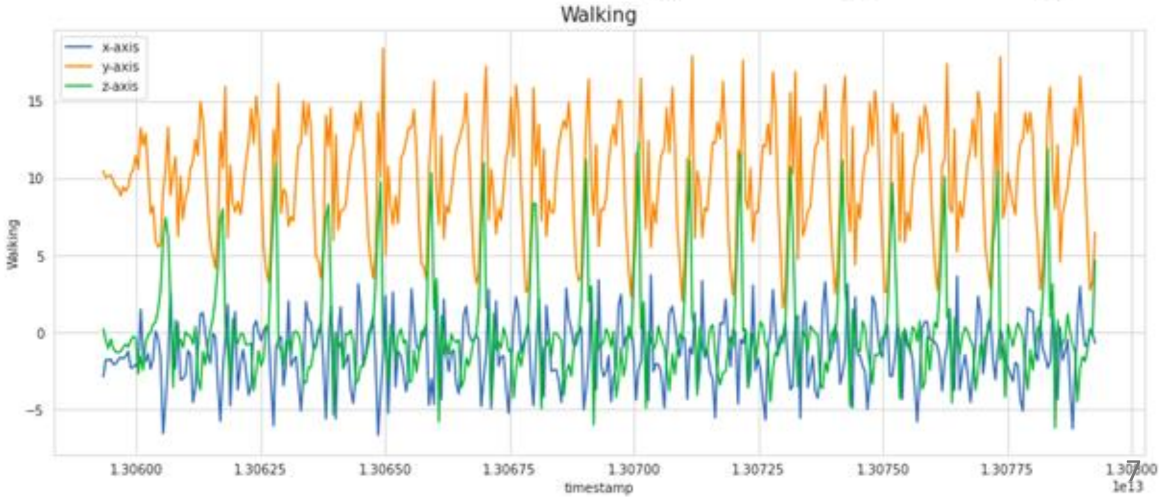
- Blood pressure
- Politics popularity rating
- The annual rainfall in Pisa
- Passengers of a company
- Accelerations on different axes
- The value of your stocks
- EEG and ECG



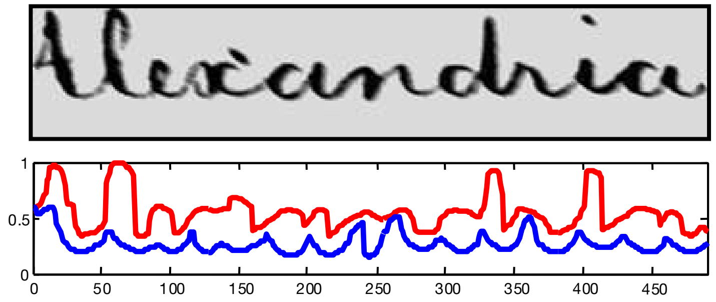



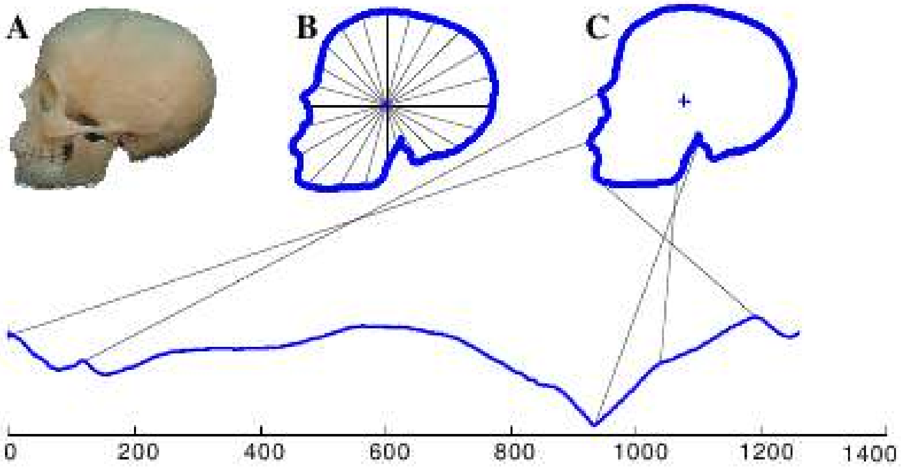



# Time Series are Ubiquitous

- Other data types can be modeled as time series
  - Text data: words count
  - Images: edges displacement
  - Videos: object positioning

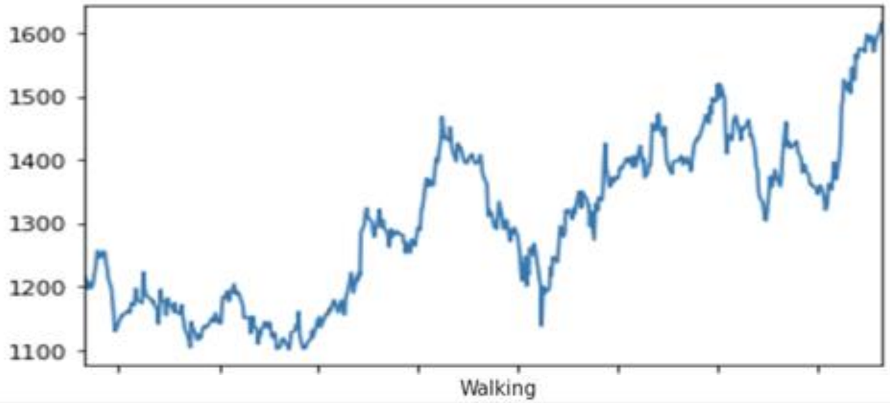


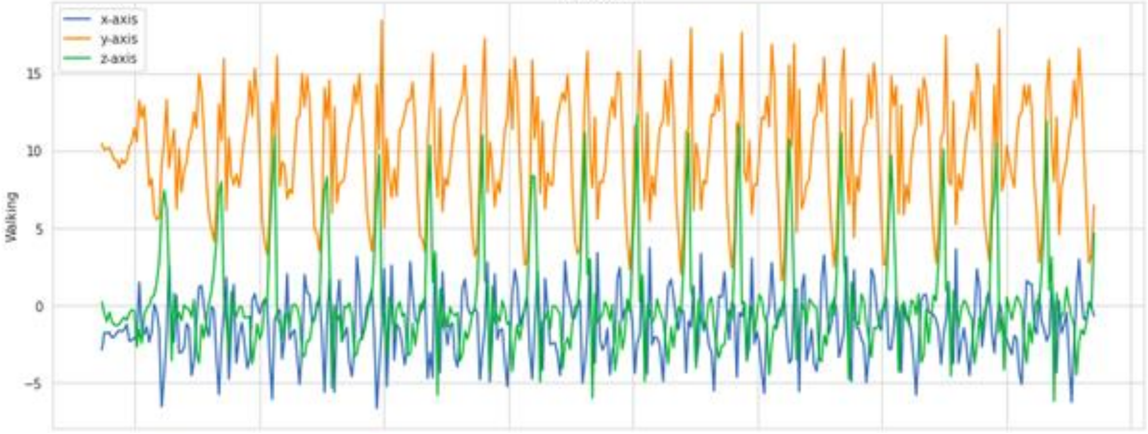

# Time Series Data

- Time Series $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x_C}]$ is a collection of *C* signals (or channels) each one with *T* observations $x_i$, i.e., $\mathbf{x_i} = [x_1, \ldots, x_T]$

- Univariate Time Series *(C = 1)*

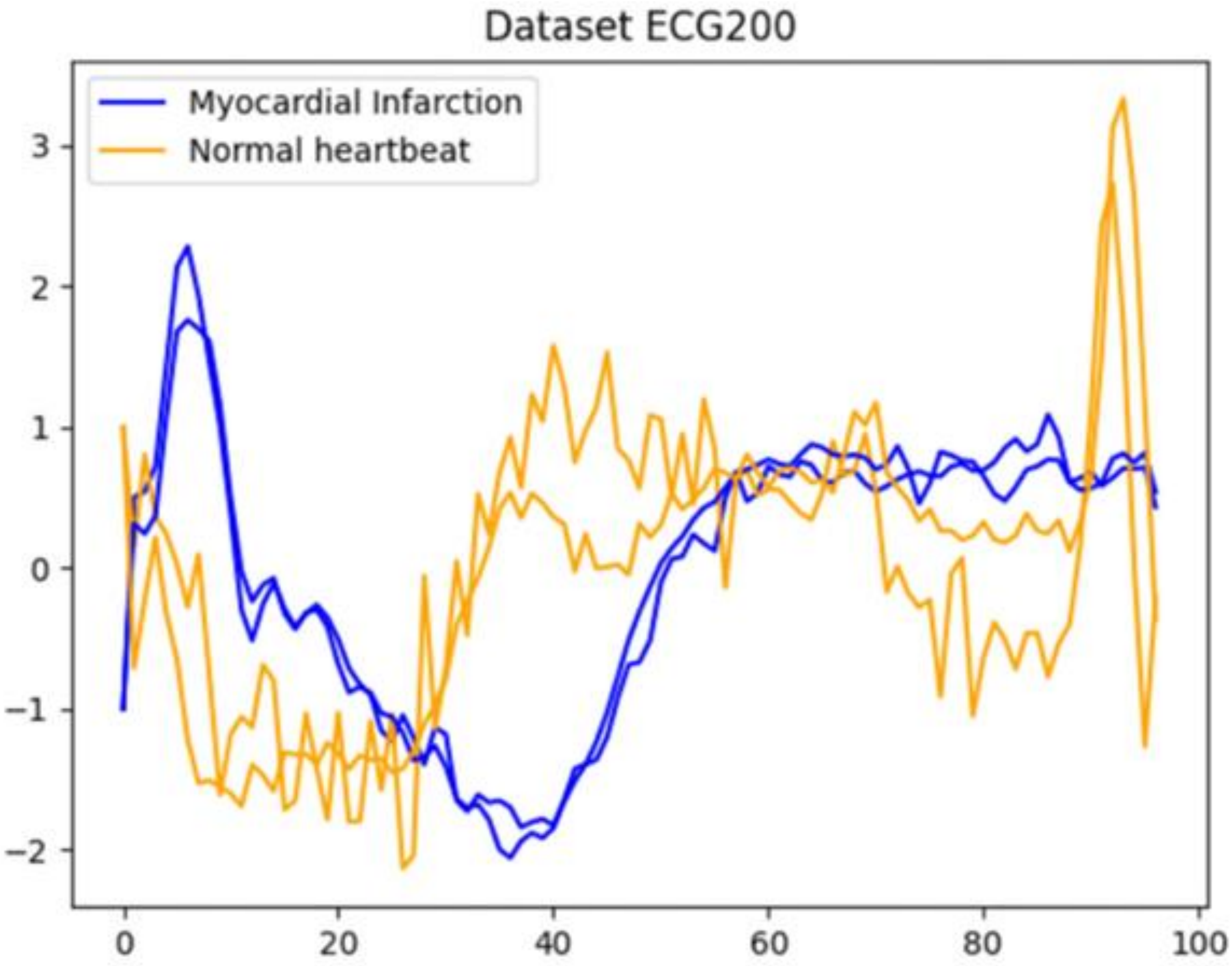



- Multivariate Time Serie *(C > 1)*



# Time Series Data

- A Time Series Dataset is a collection of time series $\mathcal{X} = \{\mathbf{X_1}, \dots, \mathbf{X_N}\}$.
- Associated to each time series we can have exogenous variables
  - Categorical Features
    - Accelerometers: Type of Movement, Crash vs NoCrash, Car Model, etc.
    - Machine Sensors: Failure vs NoFailure, Product in Production, Type of Machine, etc.
    - EEG: Syntoms, Seizure vs NoSeizure
    - Students Marks: Student Name, Course, Degree, University, Background
  - Continuous Features
    - Accellerometers: Age, Weight, Engine Temperature, etc.
    - Machine Sensors: Number of Products Realized, System Temperature, etc.
    - EEG: Age, Weight, Height, etc.
    - Students Marks: Age, Family Income, Weight, etc.


Dataset ECG200
Myocardial Infarction
Normal heartbeat
3
2
1
0
−1
−2
0
20
40
60
80
100

# Time Series in Datasets

- Stored in a 2d matrix (univariate TS)
- Shape is (n_ts, n_timestamps)

| id | $t_0$ | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ | $t_8$ | $t_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| w4r5 | 111 | 110 | 111 | 112 | 120 | 123 | 116 | 118 | 119 | 123 |
| e2t6 | 89 | 76 | 79 | 85 | 67 | 56 | 78 | 97 | 45 | 78 |
| q1e4 | 91 | 95 | 89 | 87 | 110 | 120 | 125 | 130 | 111 | 90 |
| w23t | 110 | 111 | 112 | 122 | 123 | 112 | 118 | 119 | 123 | 119 |

# Time Series in Datasets

- Stored in a wide format (id, time, all other features)

| id | time | temp | speed | rot | product | failure | sys temp |
|---|---|---|---|---|---|---|---|
| w4r5 | $t_0$ | 20.1 | 111 | 3 | A | 0 | 32.5 |
| w4r5 | $t_1$ | 18.6 | 110 | 4 | A | 0 | 32.5 |
| w4r5 | $t_2$ | 19.4 | 111 | 3 | A | 1 | 32.5 |
| w4r5 | $t_3$ | 20.4 | 112 | 5 | A | 0 | 32.5 |
| w4r5 | $t_4$ | 21.5 | 120 | 6 | A | 0 | 32.5 |
| e2t6 | $t_0$ | 12.7 | 89 | 29 | B | 0 | 34.6 |
| e2t6 | $t_1$ | 19.8 | 76 | 45 | B | 0 | 34.6 |
| e2t6 | $t_2$ | 17.4 | 69 | 34 | B | 0 | 34.6 |
| e2t6 | $t_3$ | 8.4 | 85 | 22 | B | 1 | 34.6 |
| e2t6 | $t_4$ | 7.9 | 65 | 19 | B | 1 | 34.6 |

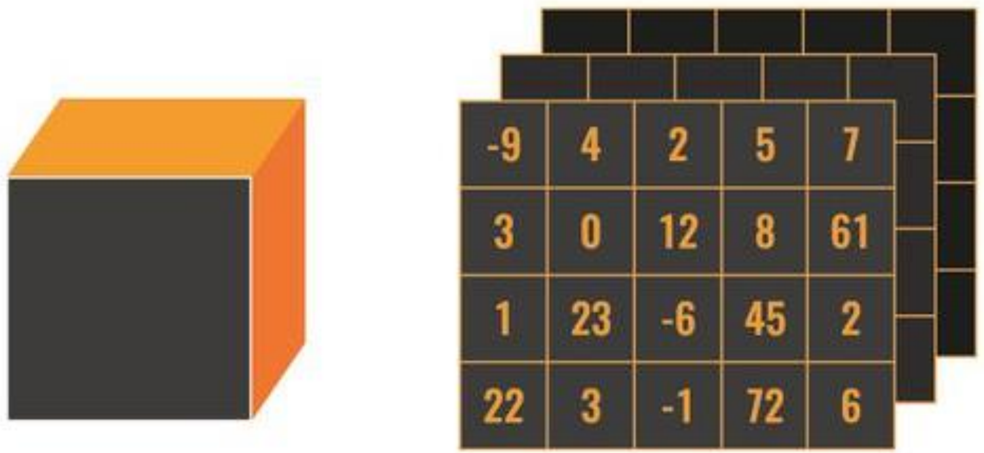
# Time Series in Datasets

- Stored in a long format (id, time, feat, value)

| id | time | feat | value |
|---|---|---|---|
| w4r5 | $t_0$ | temp | 20.1 |
| w4r5 | $t_1$ | temp | 18.6 |
| w4r5 | $t_2$ | temp | 19.4 |
| w4r5 | $t_3$ | temp | 20.4 |
| w4r5 | $t_4$ | temp | 21.5 |
| w4r5 | $t_0$ | speed | 111 |
| w4r5 | $t_1$ | speed | 110 |
| w4r5 | $t_2$ | speed | 111 |
| w4r5 | $t_3$ | speed | 112 |
| w4r5 | $t_4$ | speed | 120 |
| … | … | … | … |

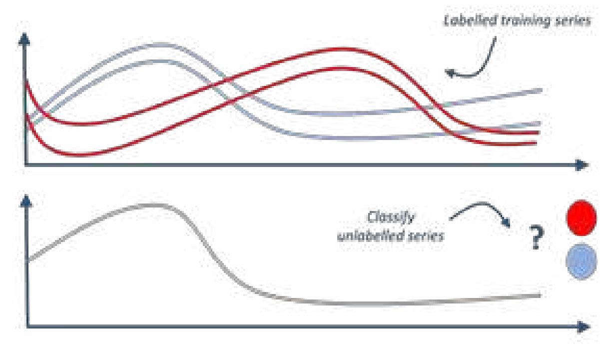
# Time Series in Datasets

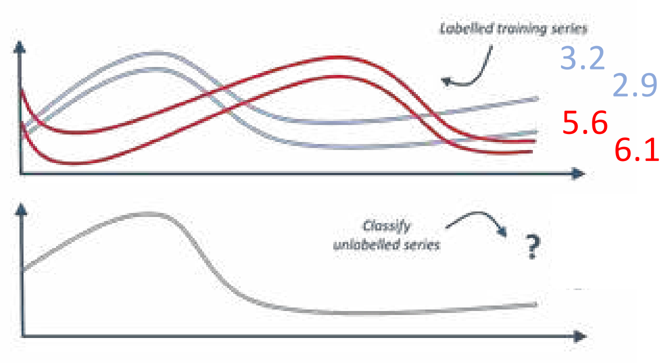
- Stored in a tensor (n_ts, n_signals, n_timestamps)
- Or sometimes (n_ts, n_timestamps, n_signals)

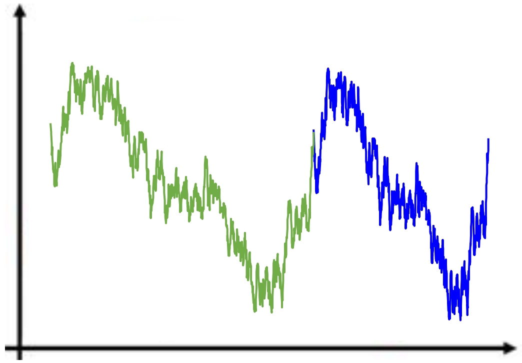


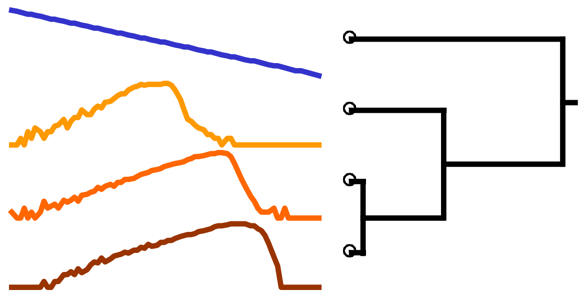
# Time Series Analytics Tasks

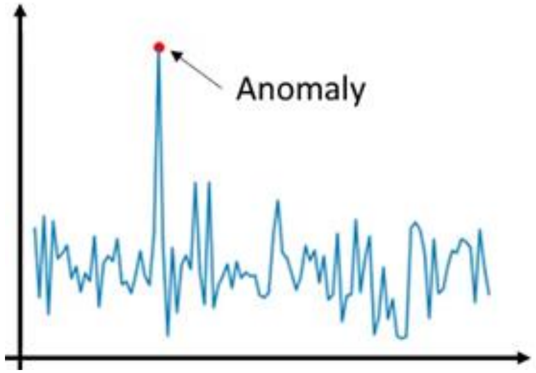
- Classification
- Regression
- Forecasting
- Clustering
- Anomaly Detection
- Pattern Mining

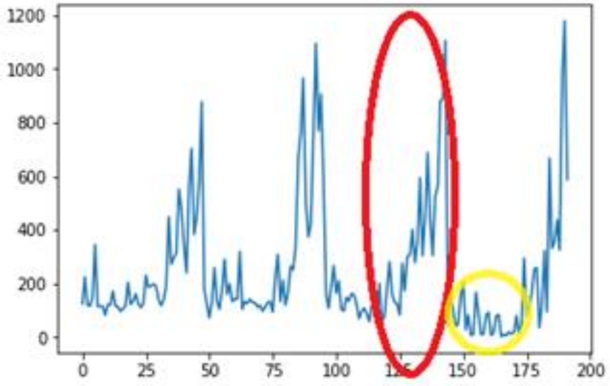


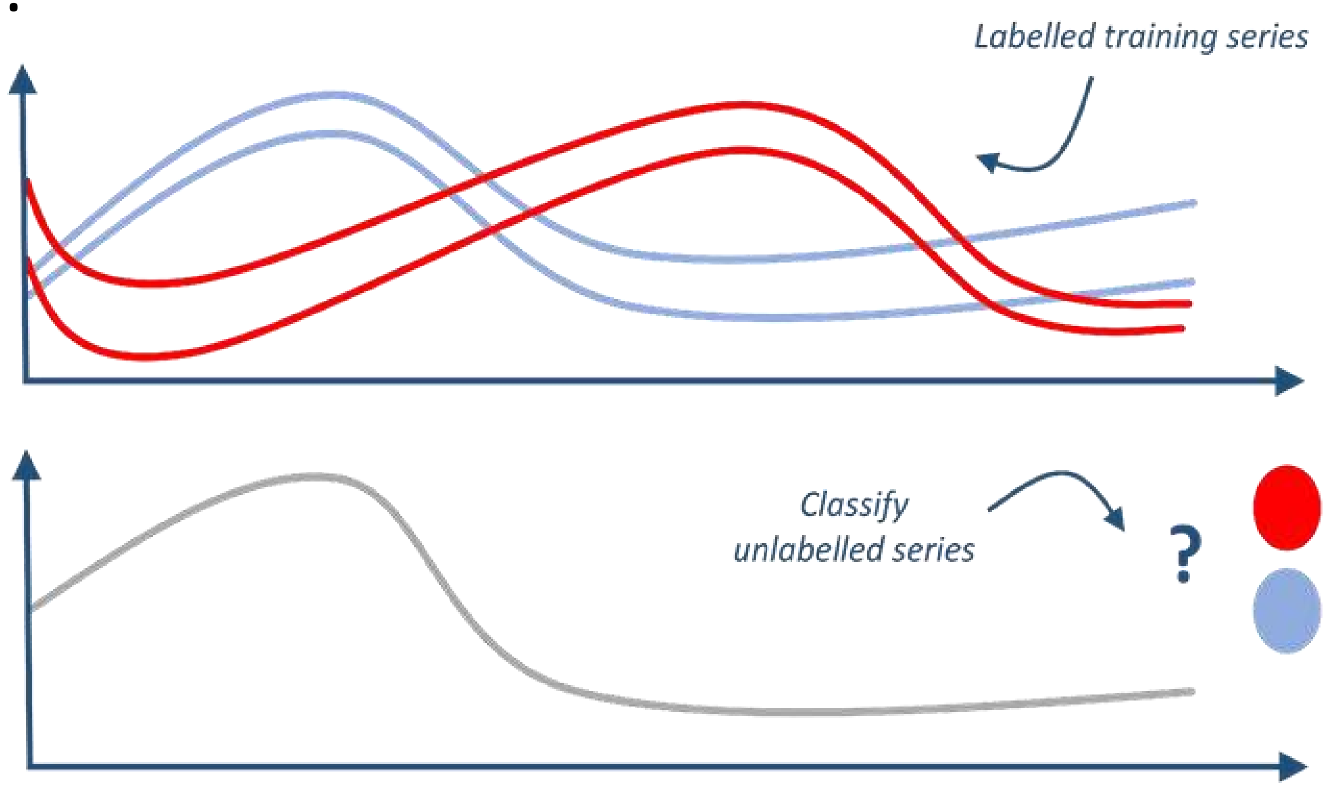




# Time Series Classification - TSC

- Given a dataset $\mathcal{X} = \{\mathbf{X_1}, \ldots, \mathbf{X_N}\}$, TSC is the task of training a model *f* to predict an exogenous <u>categorical</u> output *y* for each time series $\mathbf{X}$, i.e., $f(\mathbf{X}) = y$.

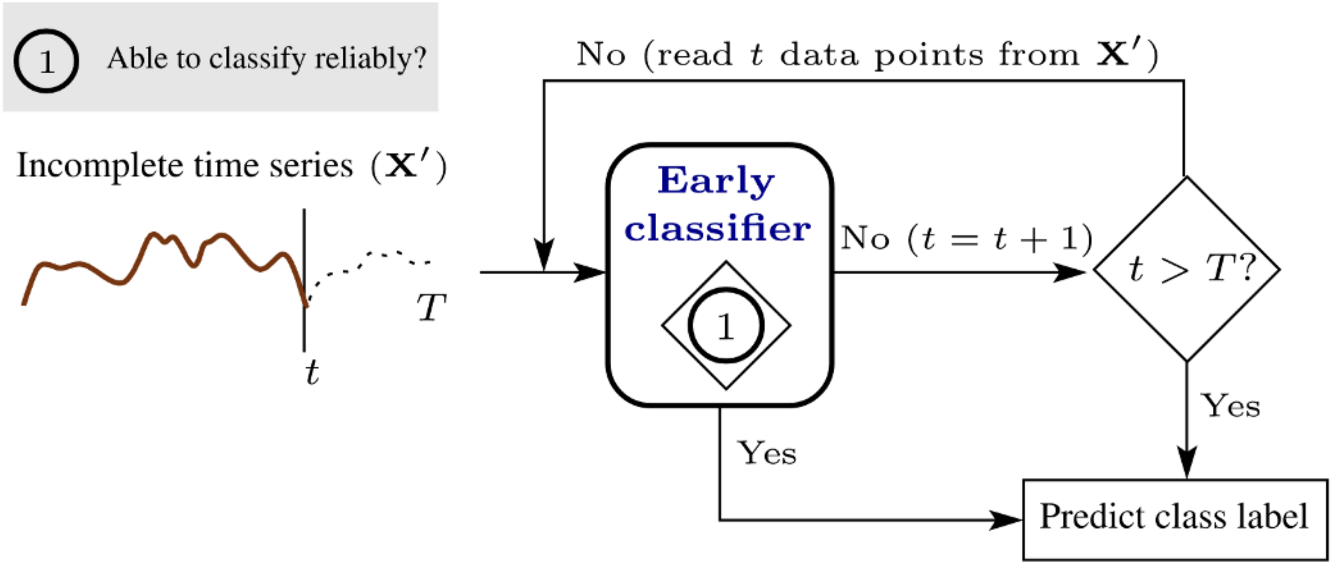


# Time Series Classification - Examples

- multivariate time series as accelerometers on the x, y, z axes and speed coming from cars black boxes. ***Target variable***: means of transport, crash vs noCrash, car model, etc.
- multivariate time series as accelerometers on the x, y, z axes, heart rate and number of steps per sec coming from personal smart devices. ***Target variable***: type of movement, type of device, next location.
- multivariate time series as machine sensors such has temperature, humidity, number of rotations, accelerations, etc. ***Target variable***: failure vs nofailure, product in production, type of machine, etc.
- multivariate time series as EEG or ECG. ***Target variable***: symptoms, disease, treatment, etc.
- univariate or multivariate time series as students' marks. ***Target variable***: student background, university, etc.

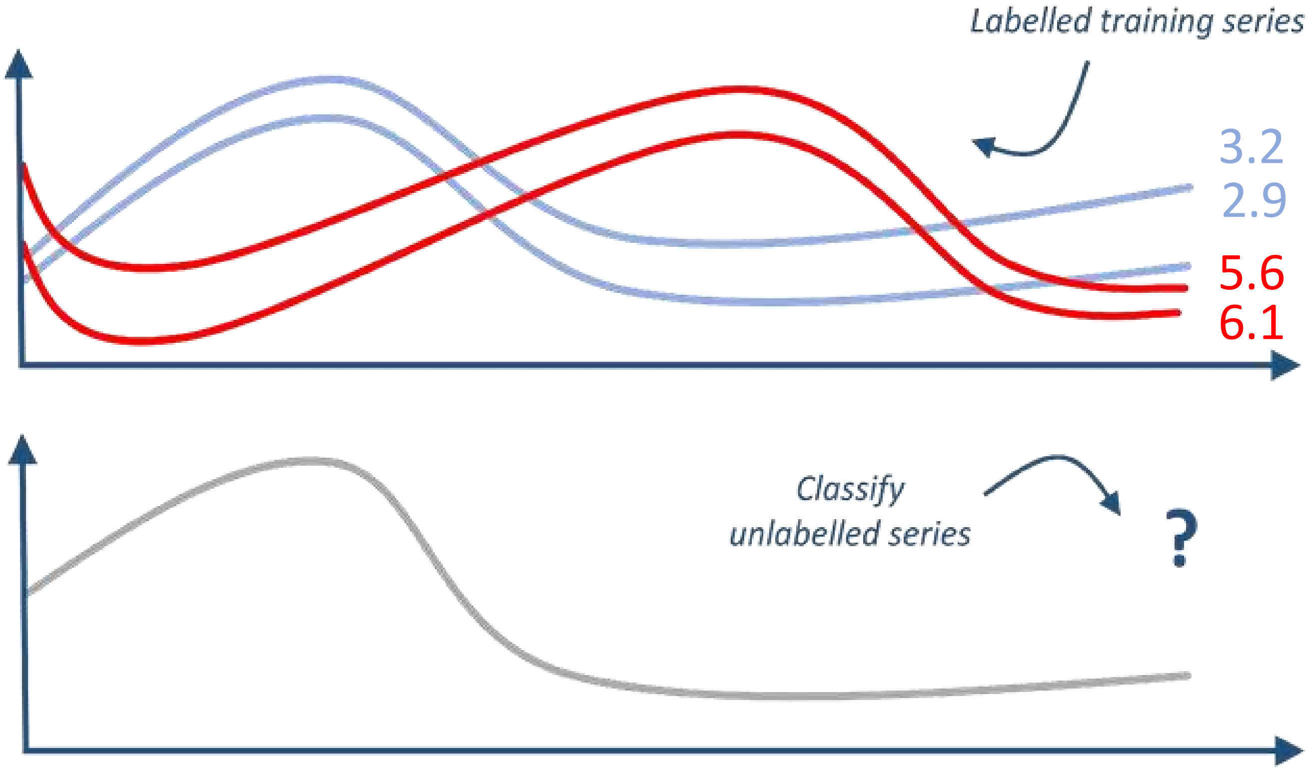
# Time Series Early Classification

- Given a dataset $\mathcal{X} = \{\mathbf{X_1}, \dots, \mathbf{X_N}\}$, time series early classification is the task of training a model *f* to predict an exogenous <u>categorical</u> output *y* for each time series $\mathbf{X}$, as early as possible, with some desired level of accuracy.



# Time Series Extrinsic Regression - TSER

- Given a dataset $\mathcal{X} = \{\mathbf{X_1}, \ldots, \mathbf{X_N}\}$, TSER is the task of training a model $f$ to predict an exogenous continuous output $y$ for each time series $\mathbf{X}$, i.e., $f(\mathbf{X}) = y$.

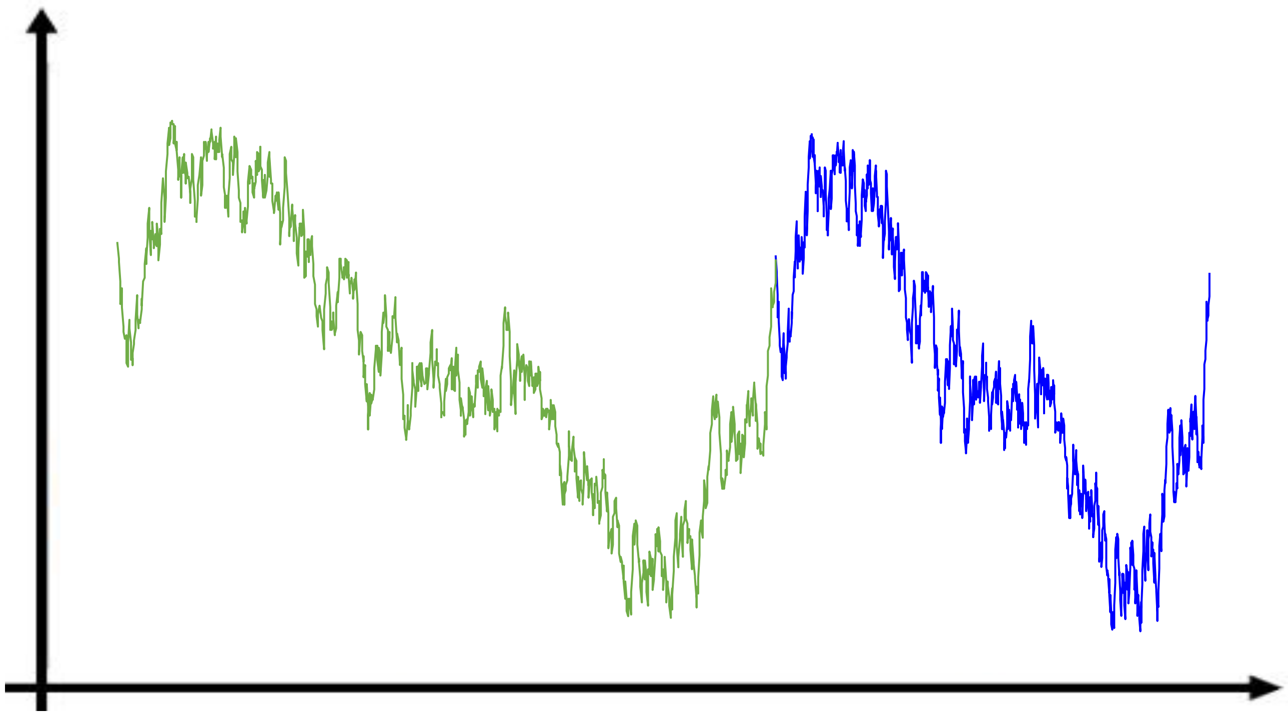


# Time Series Regression - Examples

- multivariate time series as accelerometers on the x, y, z axes and speed coming from cars black boxes. **Target variable**: engine temperature, number of car repairs, etc.
- multivariate time series as accelerometers on the x, y, z axes, heart rate and number of steps per sec coming from personal smart devices. ***Target variable***: age, number of physical activity sessions.
- multivariate time series as machine sensors such has temperature, humidity, number of rotations, accelerations, etc. ***Target variable***: number of products realized, number of failures.
- multivariate time series as EEG or ECG. ***Target variable***: patient age, patient weight, etc.
- univariate or multivariate time series as students' marks. ***Target variable***: student age, family income.

# Time Series Forecasting - TSF

- Given a time series, $\mathbf{X}$, TSF is the task of training a model *f* to predict an endogenous continuous output *y* for each time series $\mathbf{X}$, i.e., $f(\mathbf{X}) = y$.

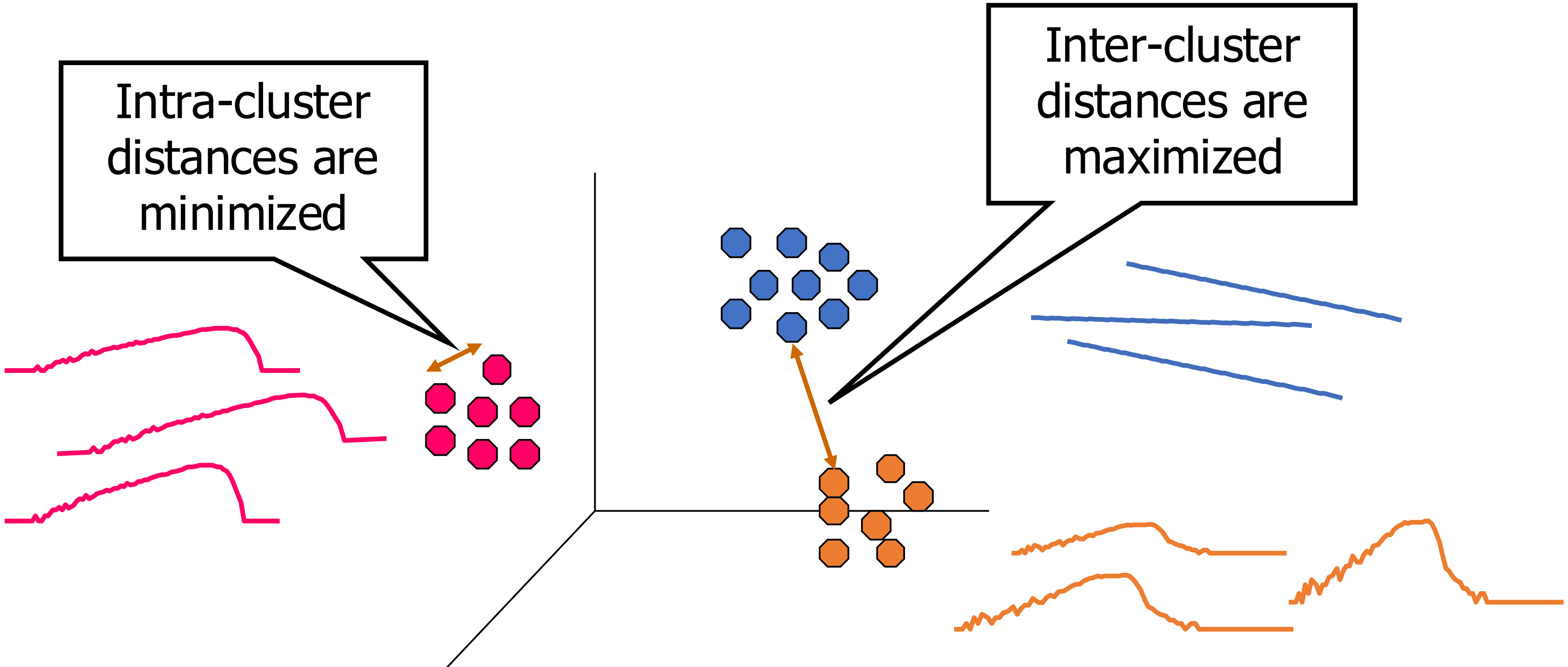


# Time Series Forecasting - Examples

- multivariate time series as accelerometers on the x, y, z axes and speed coming from cars black boxes. ***Target variable***: time series as accelerometers on the x, y-, z axes and speed.
- multivariate time series as accelerometers on the x, y, z axes, heart rate and number of steps per sec coming from personal smart devices. ***Target variable***: time series as accelerometers on the x, y, z axes, heart rate and number of steps per sec .
- multivariate time series as machine sensors such has temperature, humidity, number of rotations, accelerations, etc. ***Target variable***: temperature, humidity, number of rotations, accelerations.
- multivariate time series as EEG or ECG. ***Target variable***: future values of EEG or ECG, etc.
- univariate or multivariate time series as students' marks. ***Target variable***: future students' marks.

# Time Series Clustering

- Given a dataset $\mathcal{X} = \{\mathbf{X_1}, \dots, \mathbf{X_N}\}$ , Time Series Clustering is the task of grouping similar time series such that those in a group are similar to one another and different from the time series in other groups.

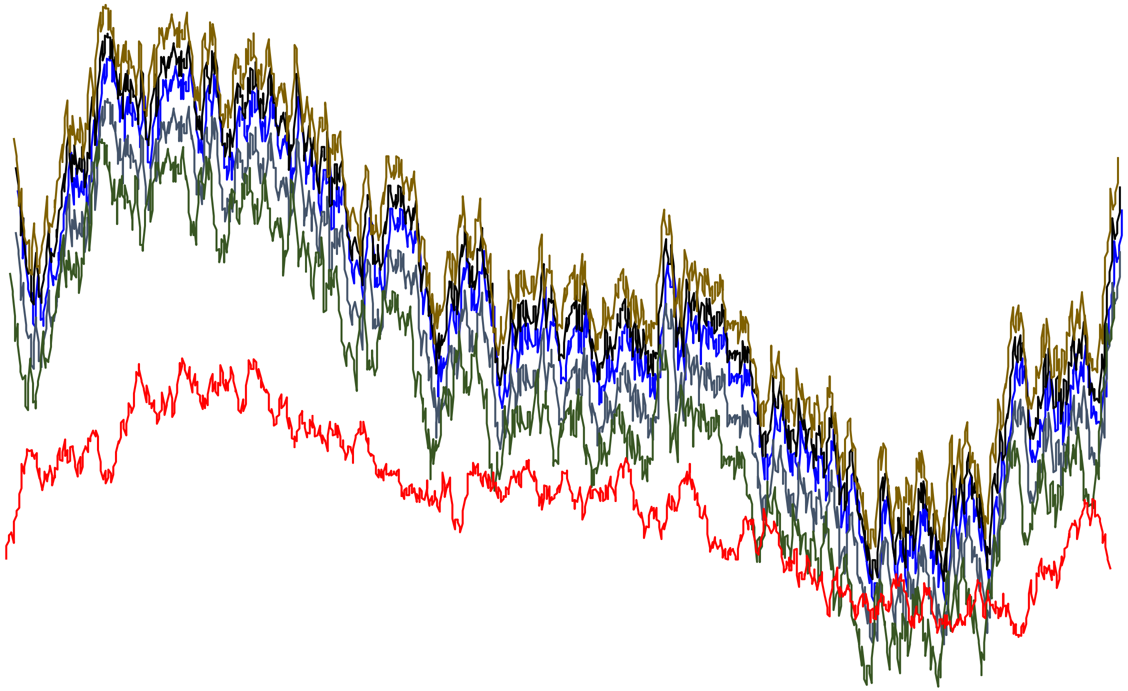


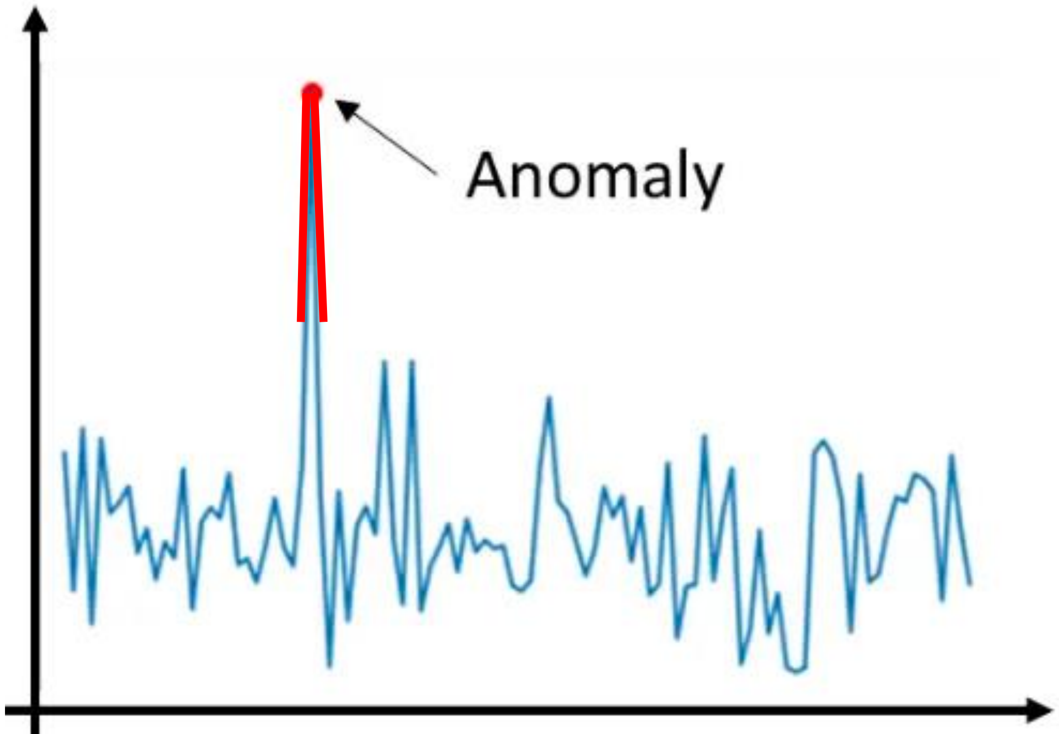
# Time Series Clustering - Examples

- multivariate time series as accelerometers on the x, y, z axes and speed coming from cars black boxes. ***Result***: groups of similar cars.
- multivariate time series as accelerometers on the x, y, z axes, heart rate and number of steps per sec coming from personal smart devices. ***Result***: groups of similar users.
- multivariate time series as machine sensors such has temperature, humidity, number of rotations, accelerations, etc. ***Results***: groups of similar products produced.
- multivariate time series as EEG or ECG. ***Results***: groups of similar patients.
- univariate or multivariate time series as students' marks. ***Results***: groups of similar students.

# Anomaly Detection

- Given a dataset $\mathcal{X} = \{\mathbf{X_1}, \ldots, \mathbf{X_N}\}$, Anomaly Detection is the task of:
  a) identifying anomalous time series within the set $X$
  b) identifying anomalous time stamps for each time series $T$



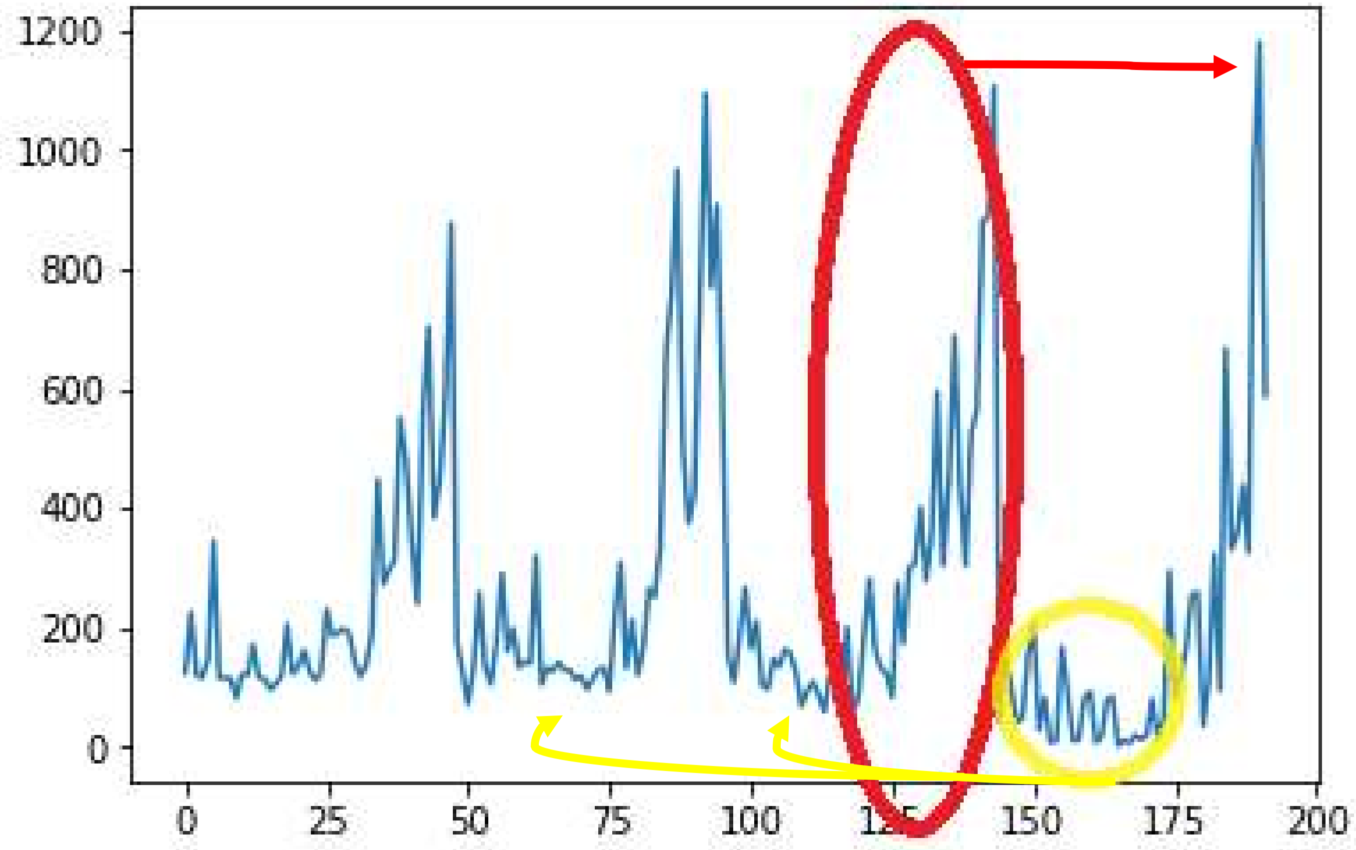


# Time Series Anomaly Detection - Examples

- multivariate time series as accelerometers on the x, y, z axes and speed coming from cars black boxes. ***Result***: anomalous cars or anomalous time stamps.
- multivariate time series as accelerometers on the x, y, z axes, heart rate and number of steps per sec coming from personal smart devices. ***Result***: anomalous users or anomalous time stamps.
- multivariate time series as machine sensors such has temperature, humidity, number of rotations, accelerations, etc. ***Results***: anomalous products/machines or anomalous time stamps.
- multivariate time series as EEG or ECG. ***Results***: anomalous patients or anomalous time stamps.
- univariate or multivariate time series as students' marks. ***Results***: anomalous students or anomalous time stamps.

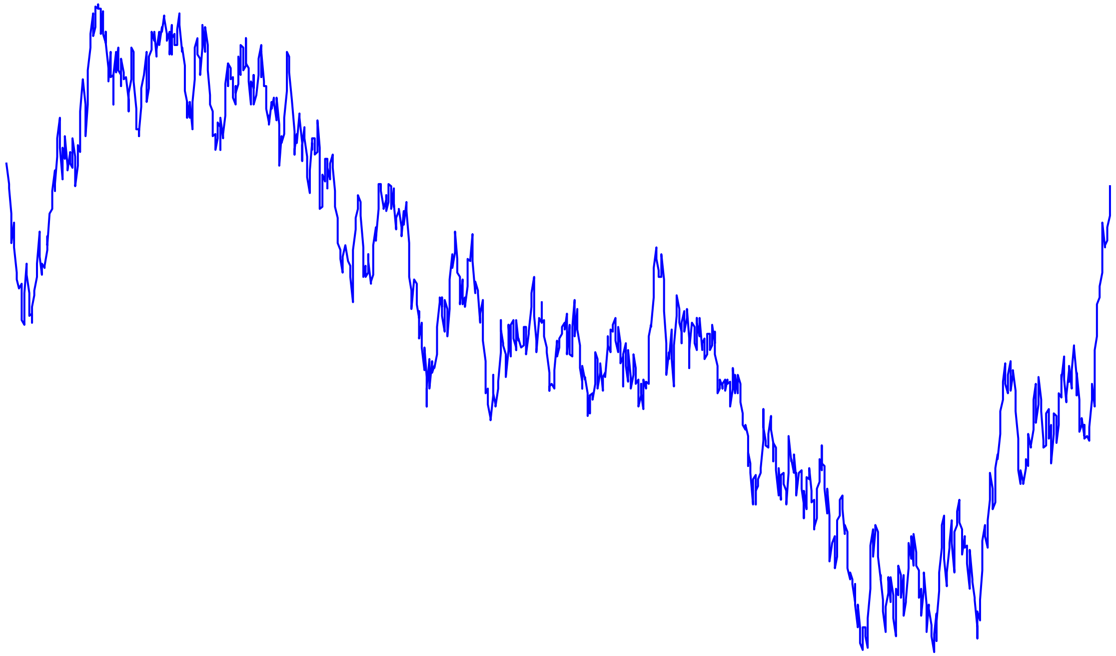
# Pattern Mining

- Given a dataset $\mathcal{X} = \{\mathbf{X_1}, \dots, \mathbf{X_N}\}$ , Pattern Mining is the task of identifying repeated subsequences in each time series *T*.



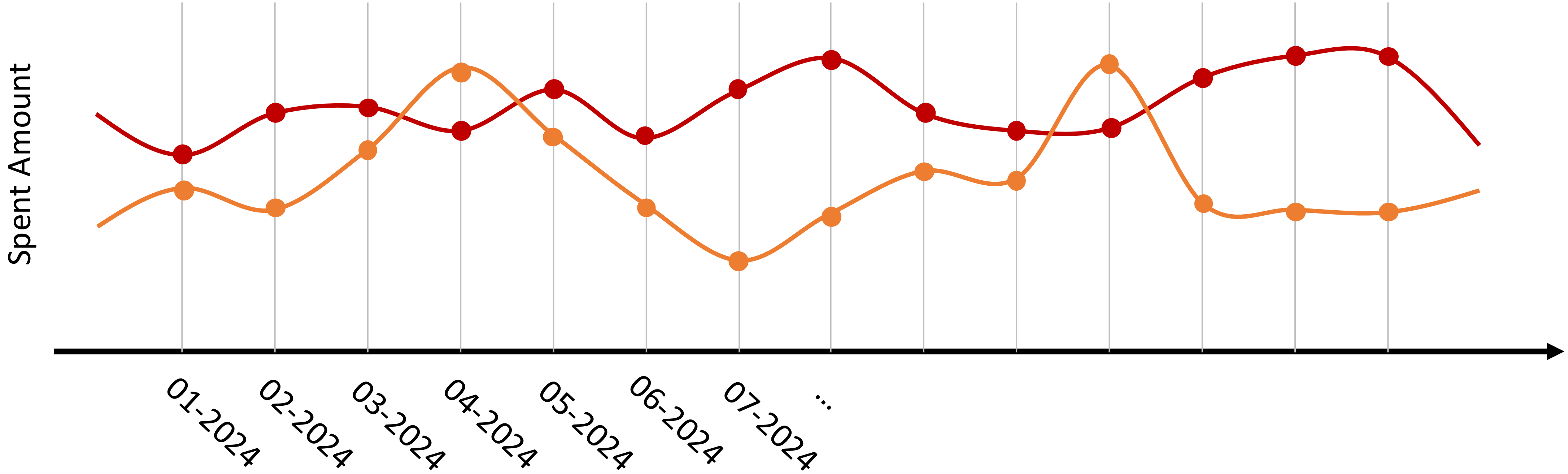
# Problems in Working with Time Series

- Large amount of data
- Similarity is not easy to estimate
- Different data formats
- Different sampling rates
- Noise
- Missing values
- …



# Assumption

We assume that for a given dataset $\mathcal{X} = \{\mathbf{X_1}, \ldots, \mathbf{X_N}\}$ all the time series have aligned time stamps, i.e., the *i-th* time stamp of a time series $\mathbf{X}_\mathrm{a}$ corresponds to the *i-th* time stamp of another time series $\mathbf{X}_\mathrm{b}$ and this is true for all the time series in *X*.



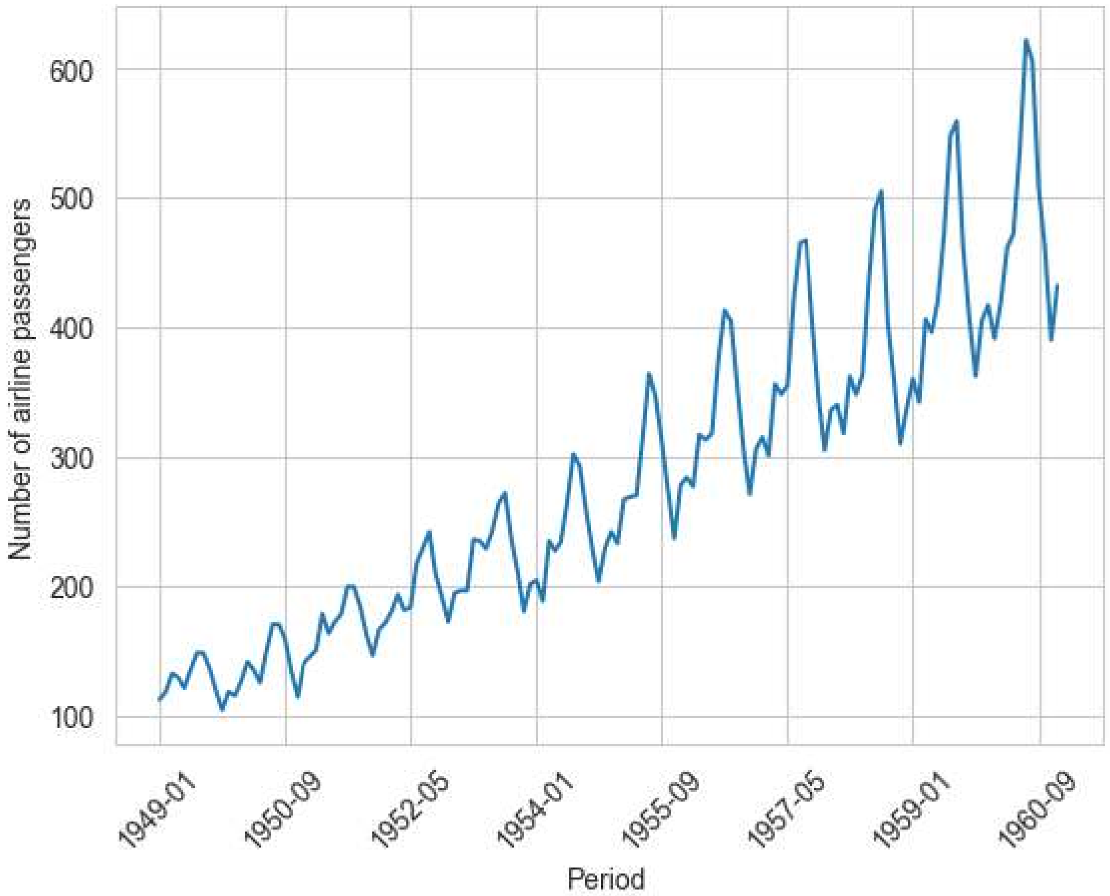
# Time Series Visualization

A plot can reveal

- **Trend**: upward or downward pattern
- **Periodicity**: repetition of behavior in a regular pattern
- **Seasonality**: periodic behavior with a known period (hourly, weekly, monthly...)
- **Heteroskedasticity**: changing variance
- **Dependence**: positive (successive observations are similar) or negative (successive observations are dissimilar)
- **Outliers**: anomalous time stamps, anomalous subsequences
- **Missing data**: missing values at a certain time stamp or longer subsequences

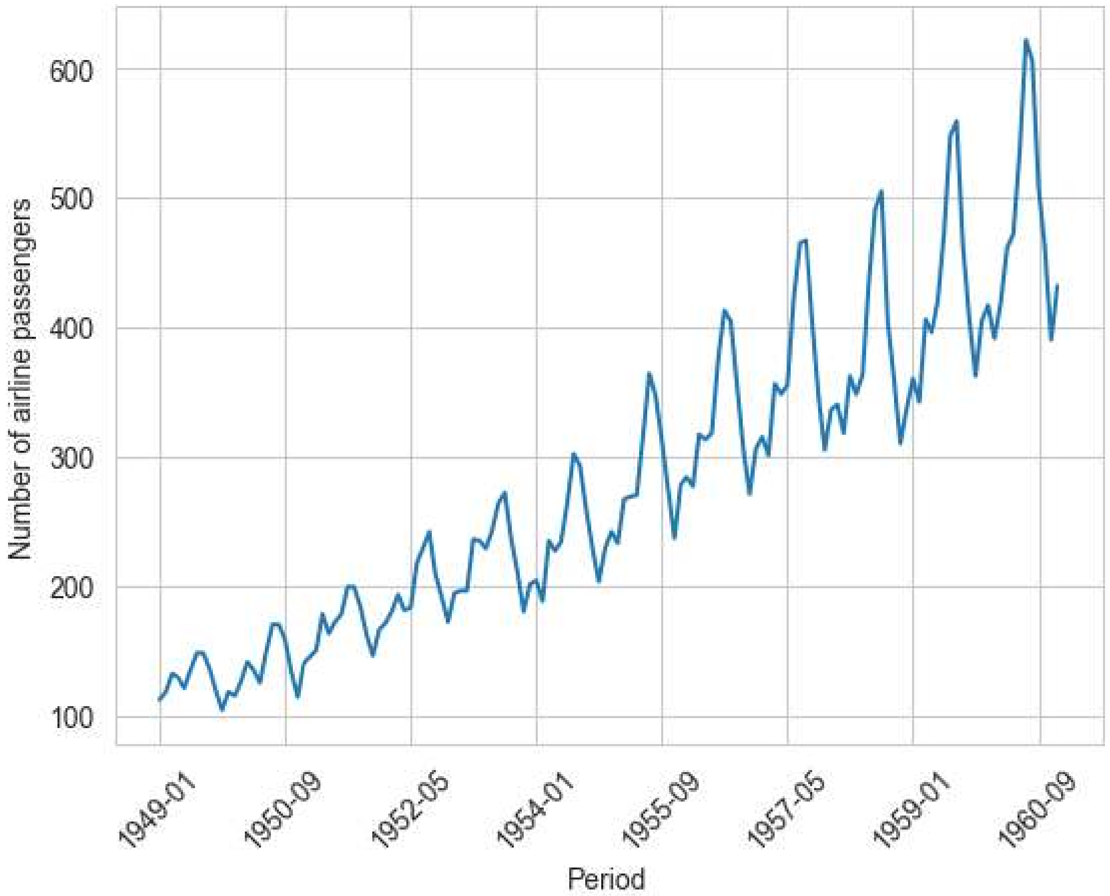
# Example 1

- Airline passengers from 1949-1961



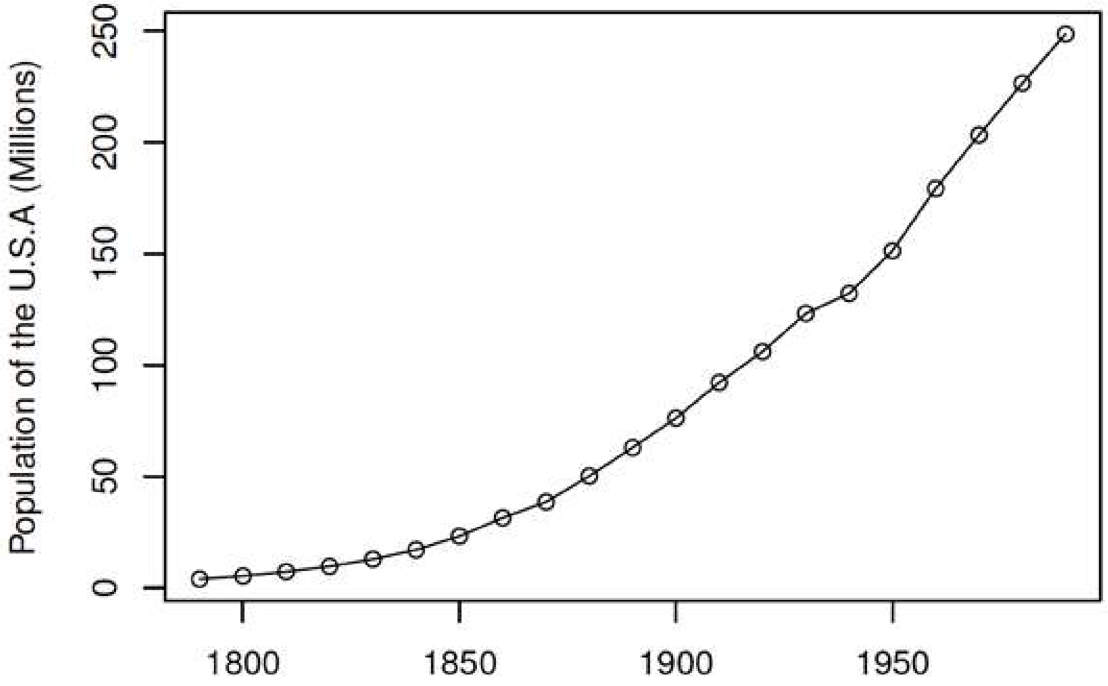
# Example 1

- Airline passengers from 1949-1961
- Upward trend
- Seasonality on a 12 month interval
- Increasing variability



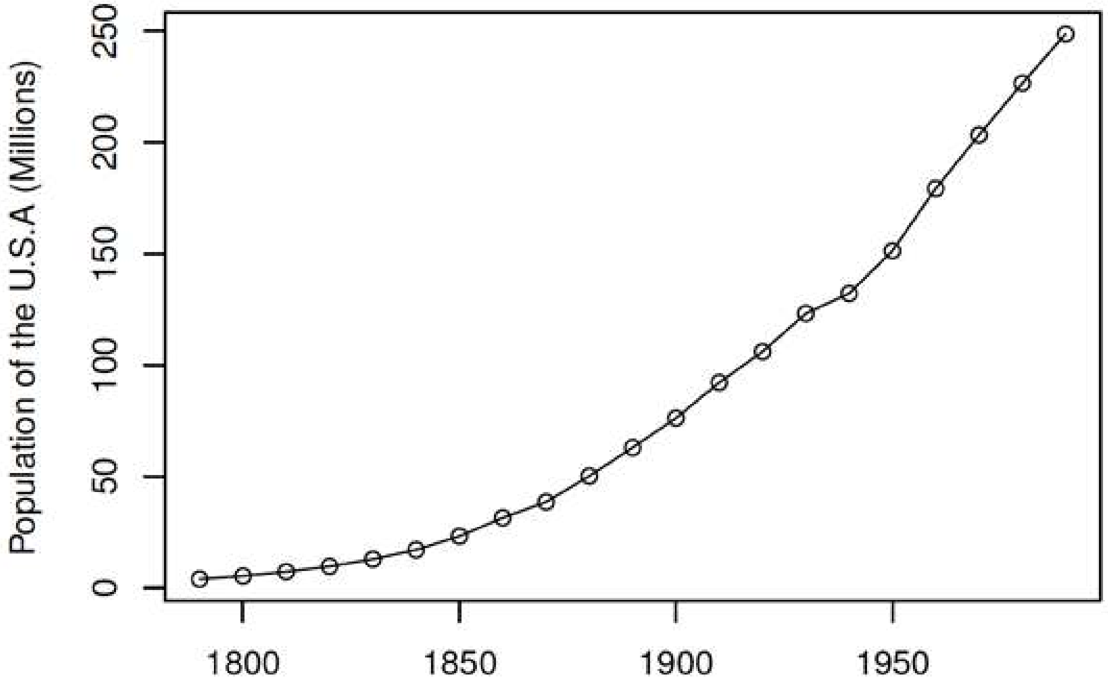
# Example 2

- U.S.A. population at ten year intervals from 1790-1990



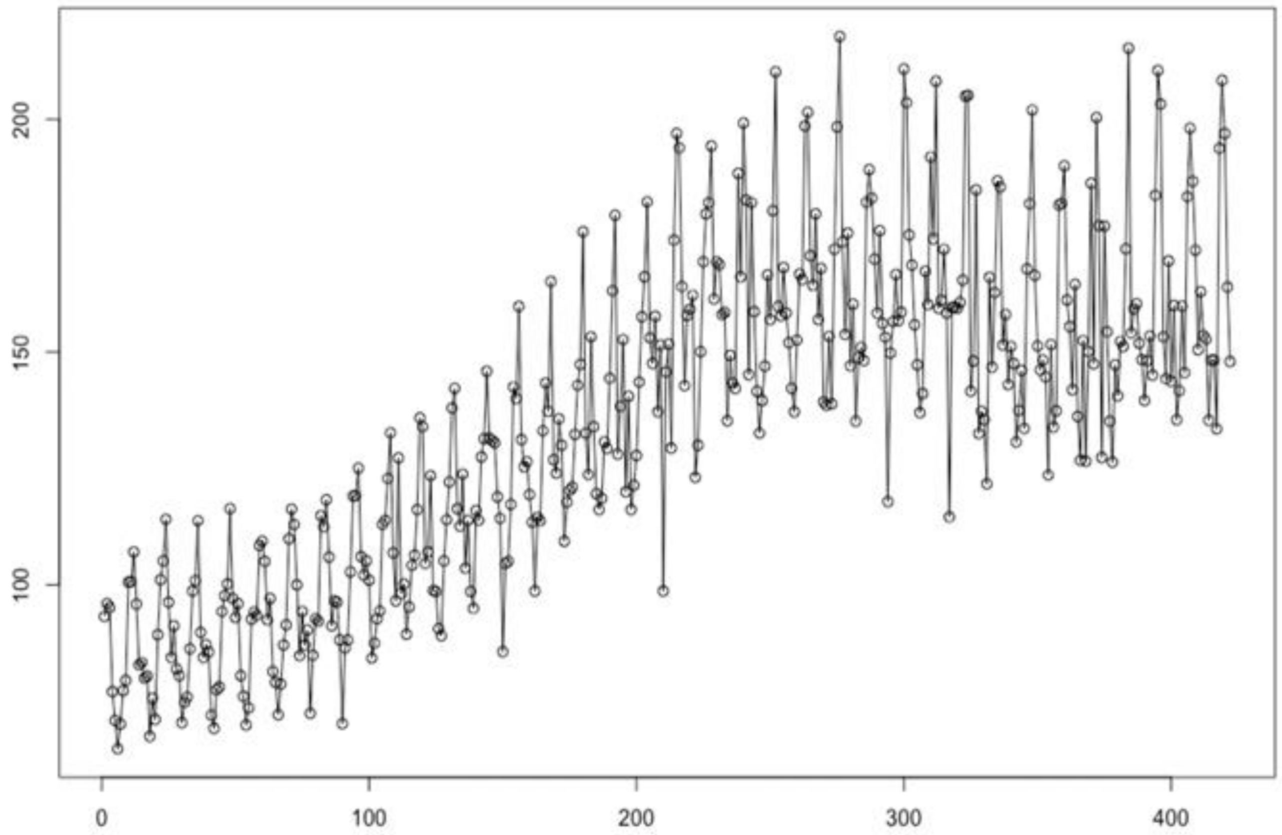
# Example 2

- U.S.A. population at ten year intervals from 1790-1990
- Upward trend
- Slight change in shape/structure
- Nonlinear behavior



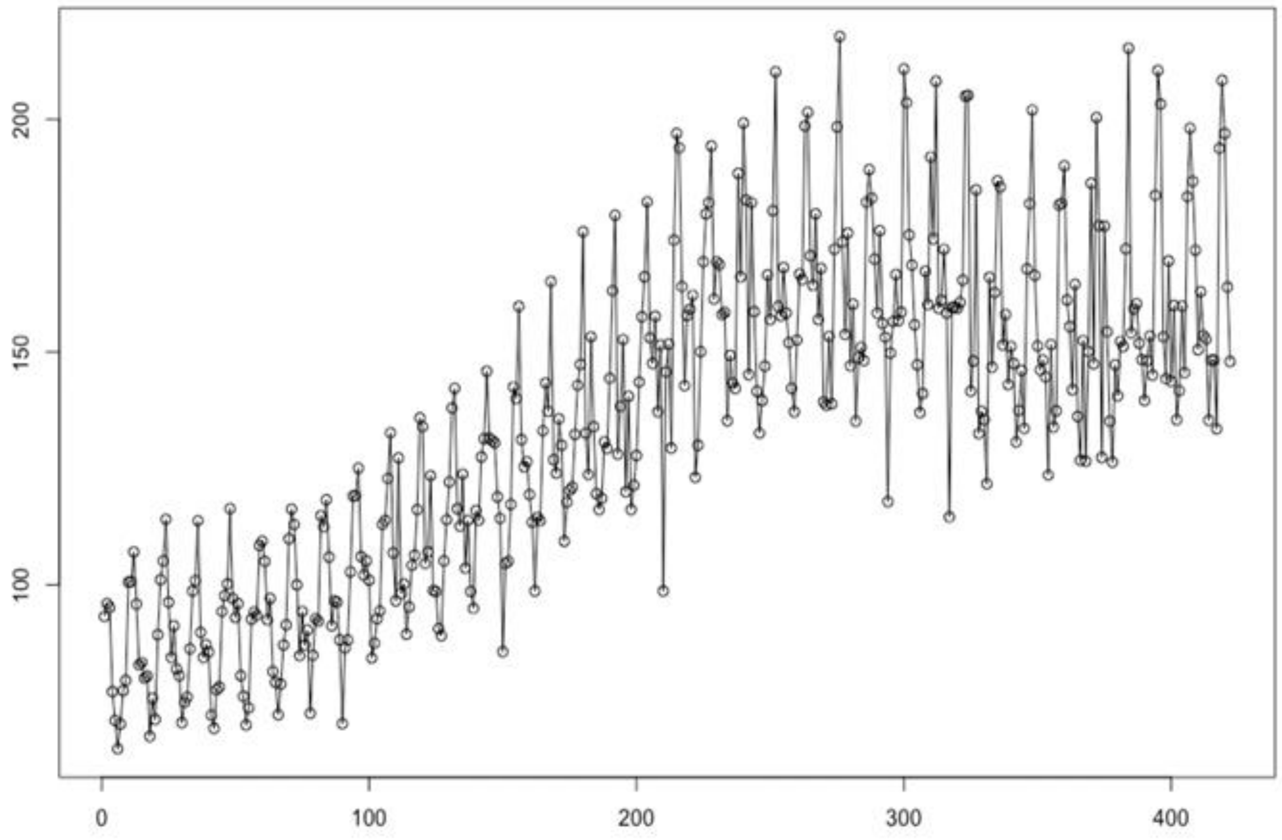
# Example 3

- Monthly Beer Production in Australia



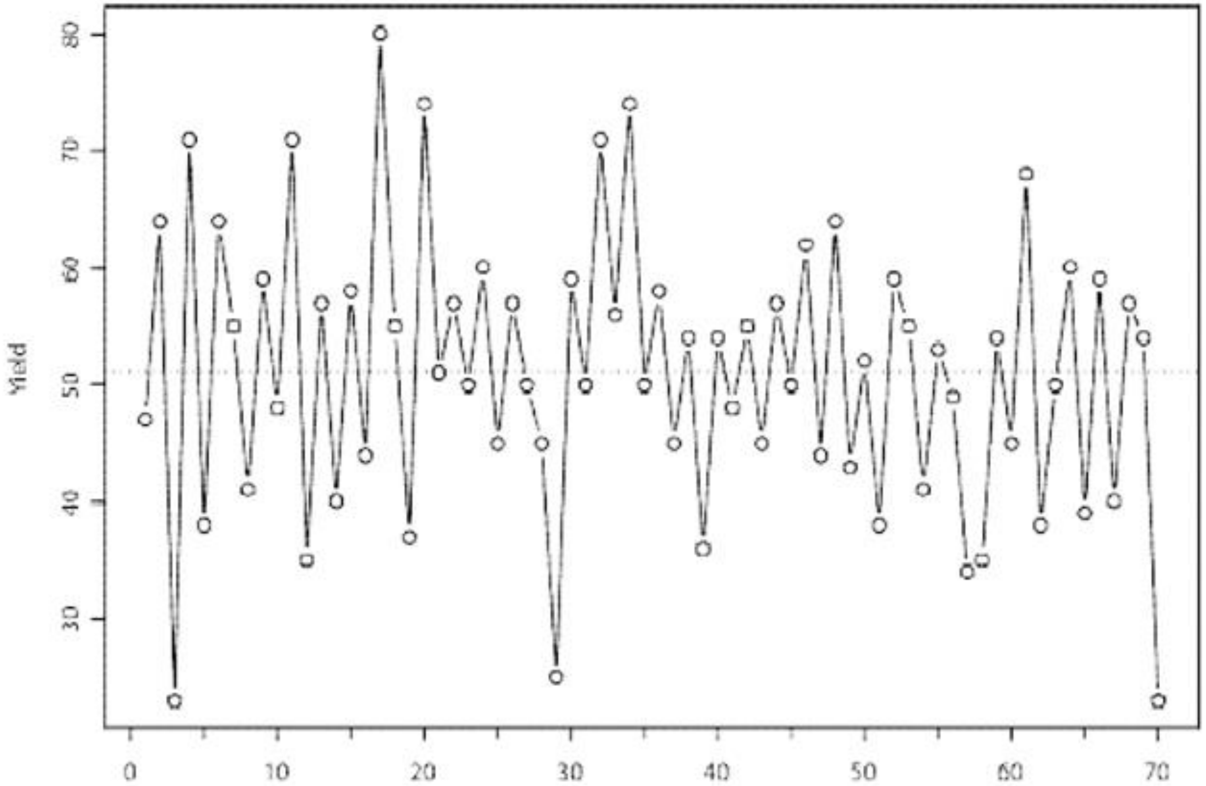
# Example 3

- Monthly Beer Production in Australia
- No trend in last 100 months
- No clear seasonality



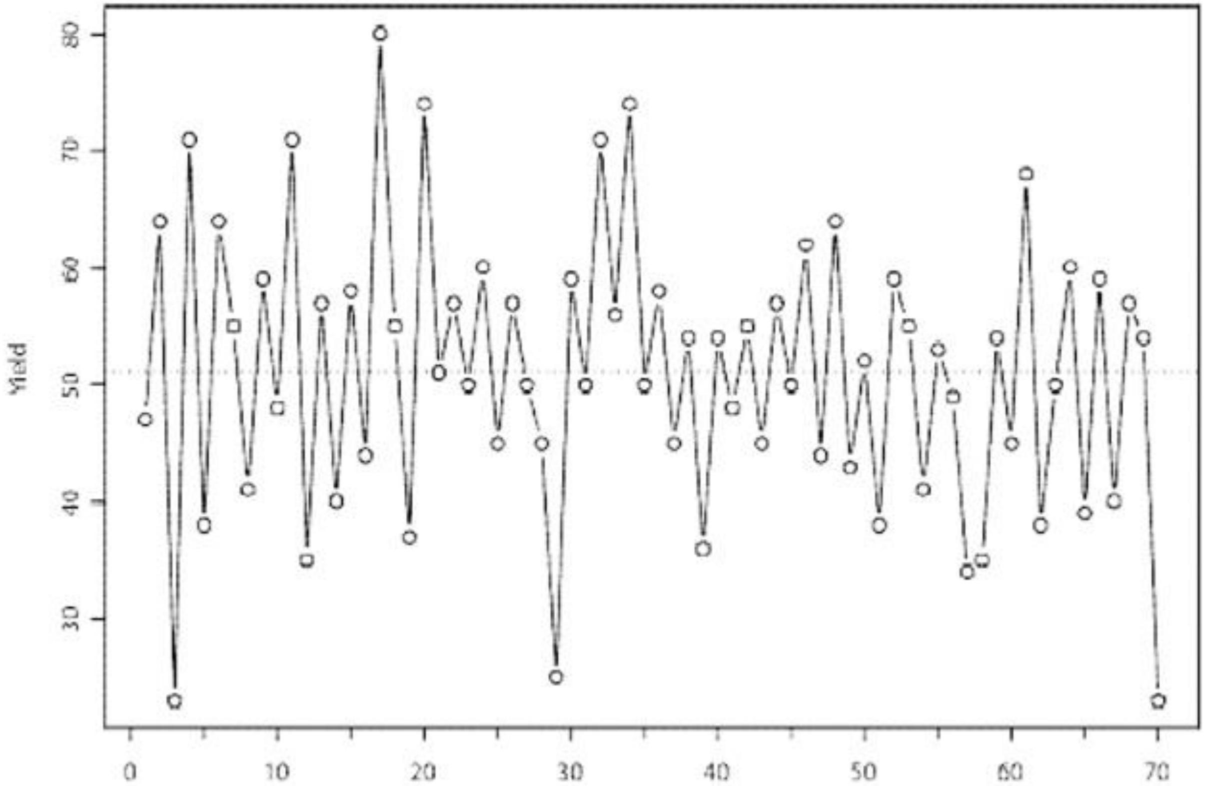
# Example 4

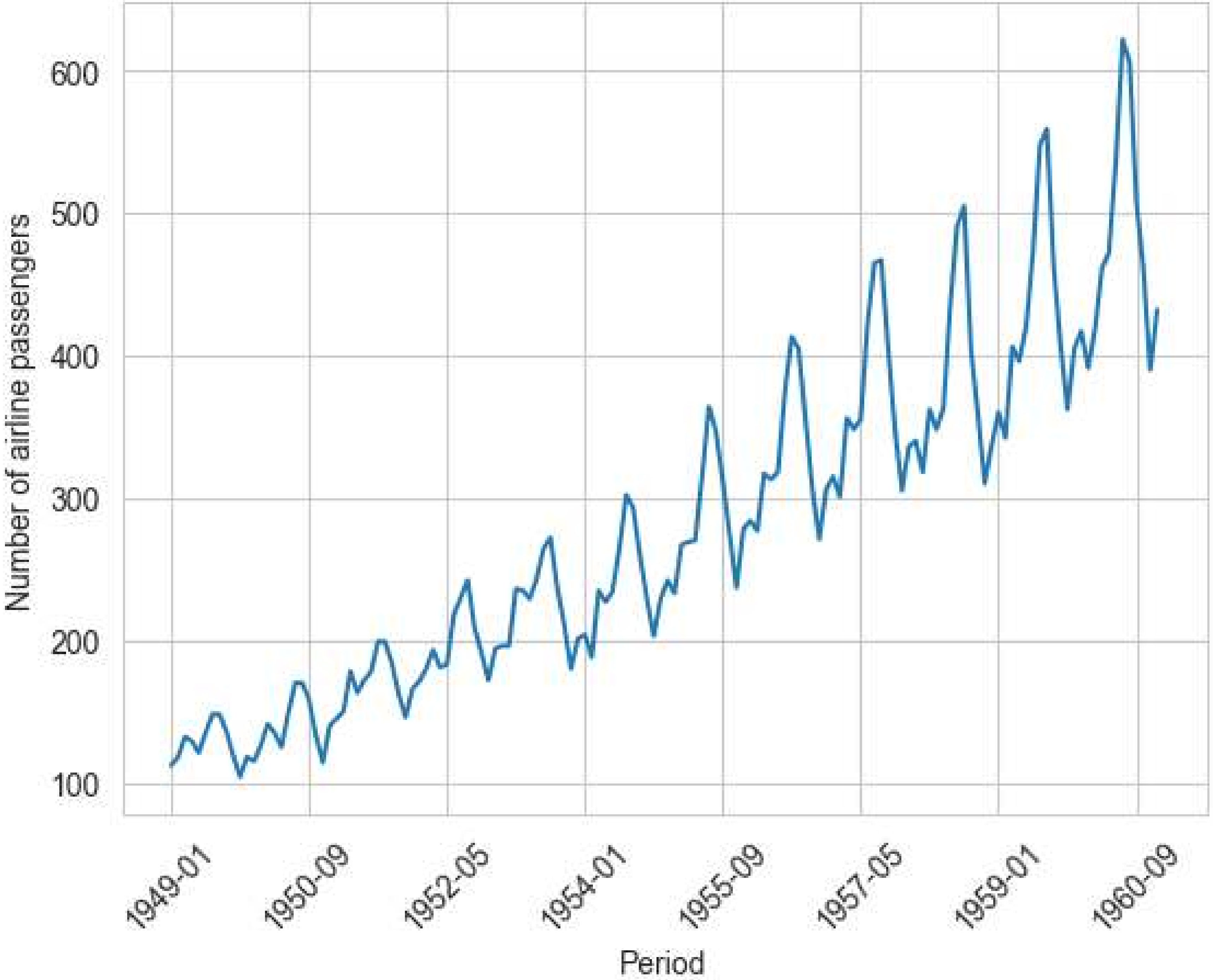
- Yield from a controlled chemical batch process



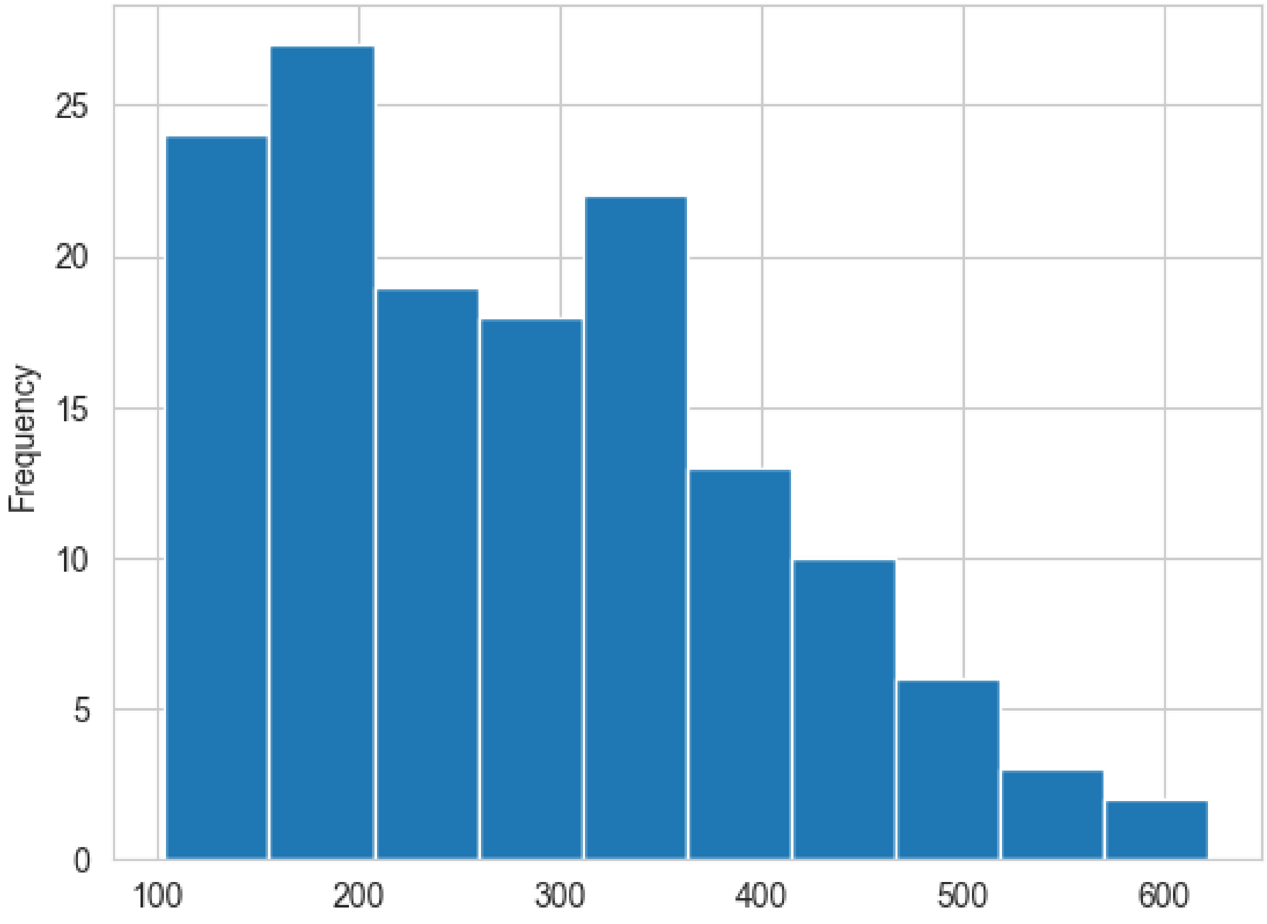
# Example 4

- Yield from a controlled chemical batch process
- Negative dependence: successive observations tend to lie on opposite sides of the mean.

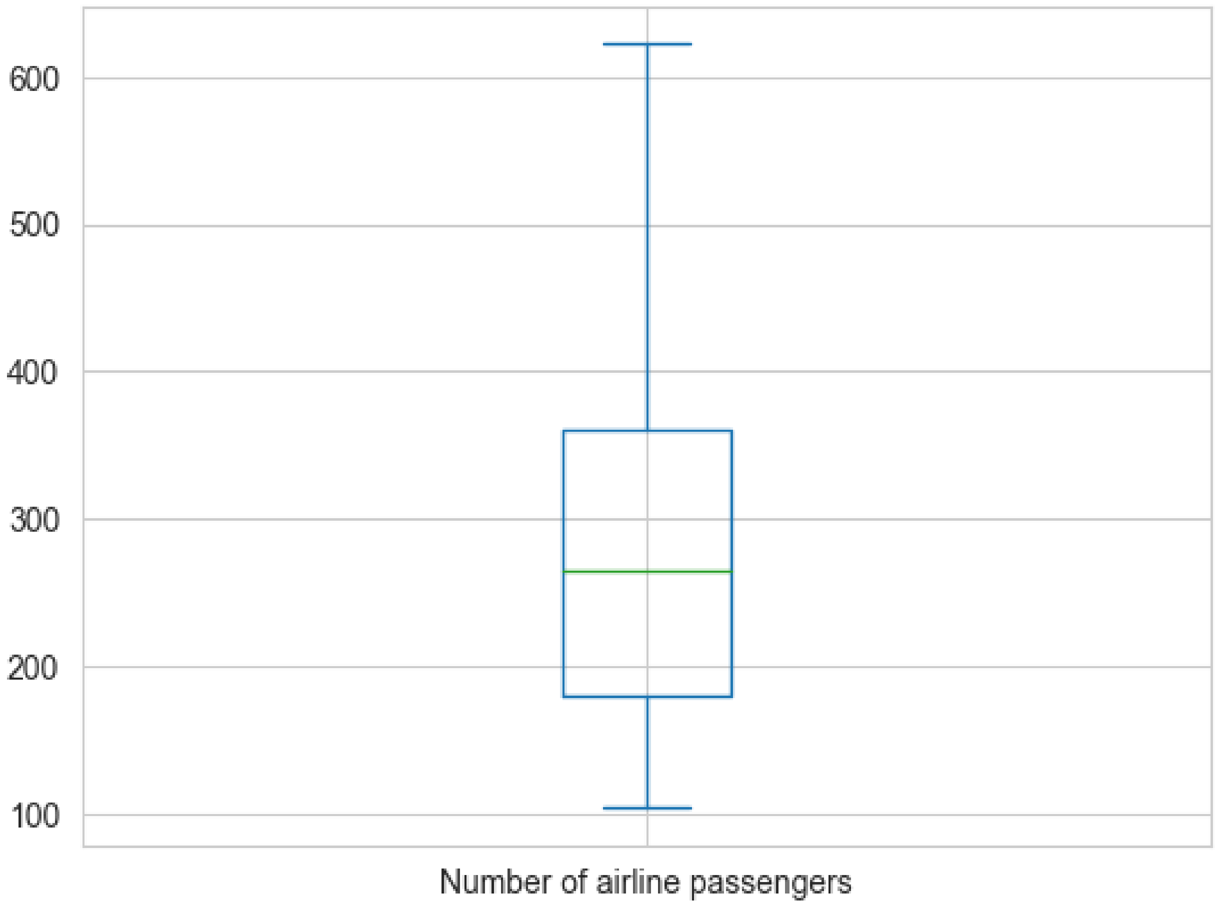




600
500
400
300
200
100
Number of airline passengers
1949-01
1950-09
1952-05
1954-01
1955-09
1957-05
1959-01
1960-09
Period




TS Histogram



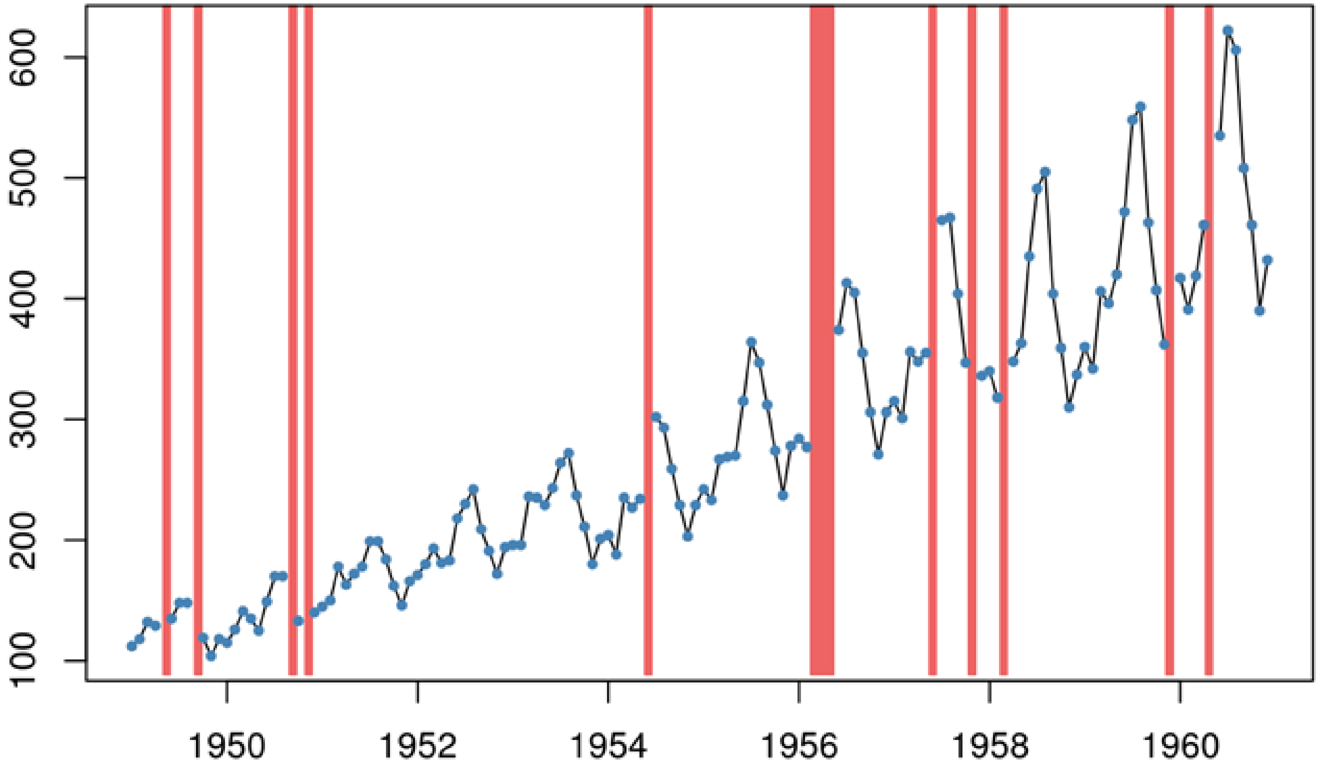

TS BoxPlot

Time in not considered in these plots!!!

# Missing Values

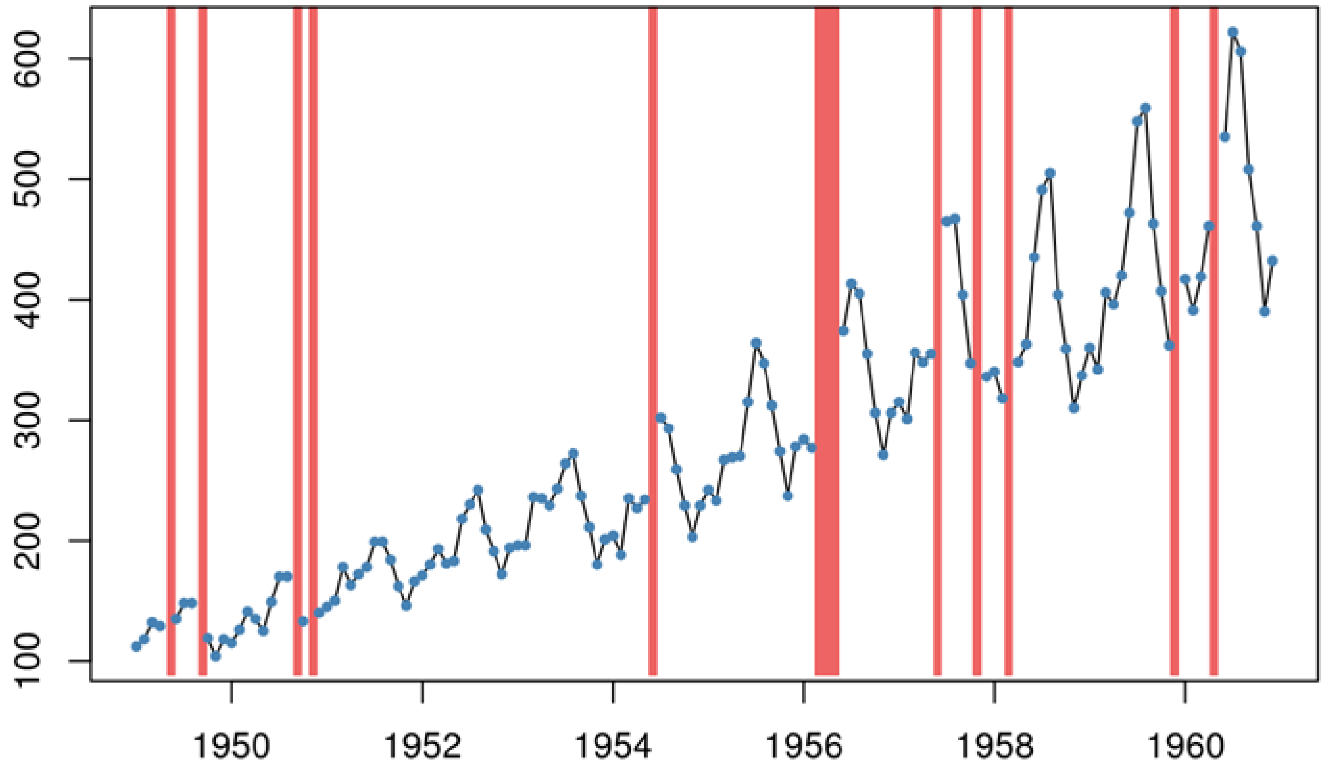
# Missing Values

- Individual values for a single time stamp can be missing
- Contiguous values for sequential time stamps can be missing

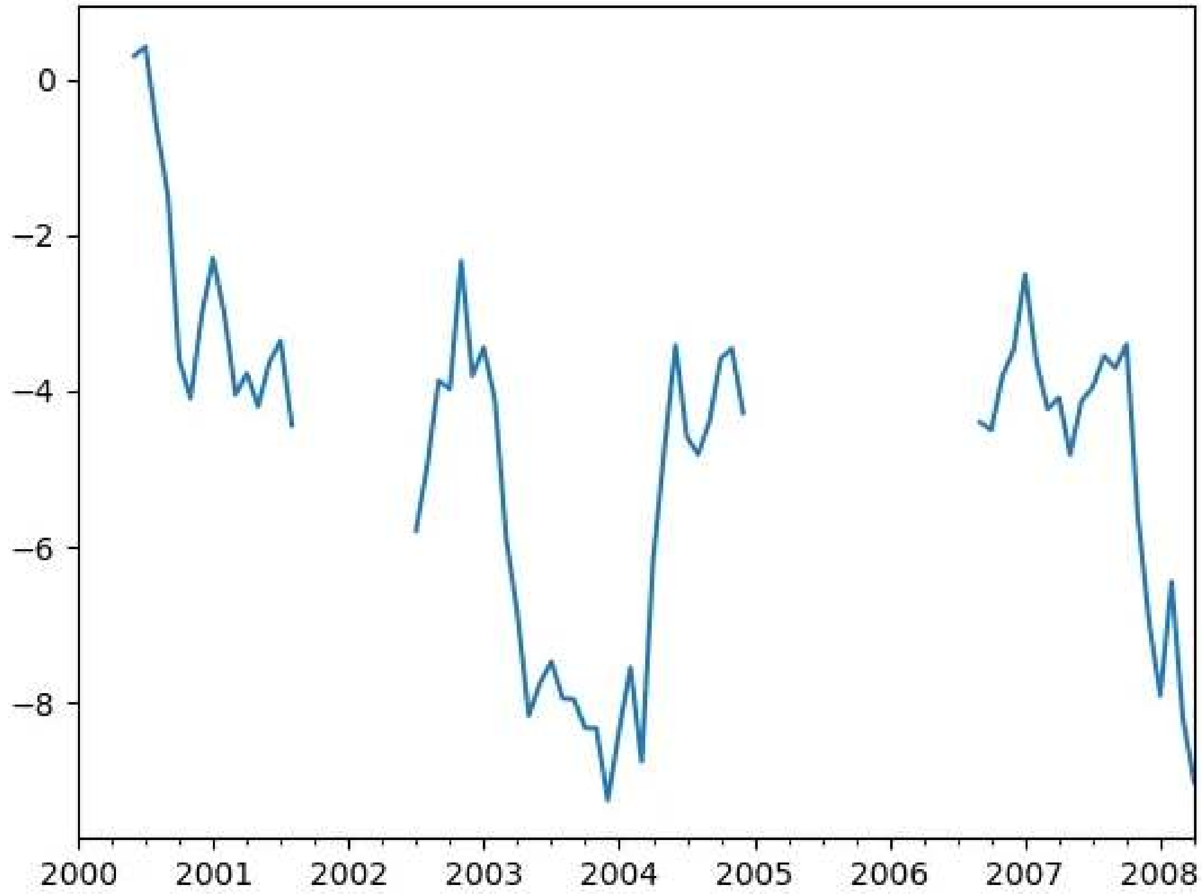


# Missing Values Imputation Methods

- Fill with constant value
- Linear interpolation
- Forecasting
- Random

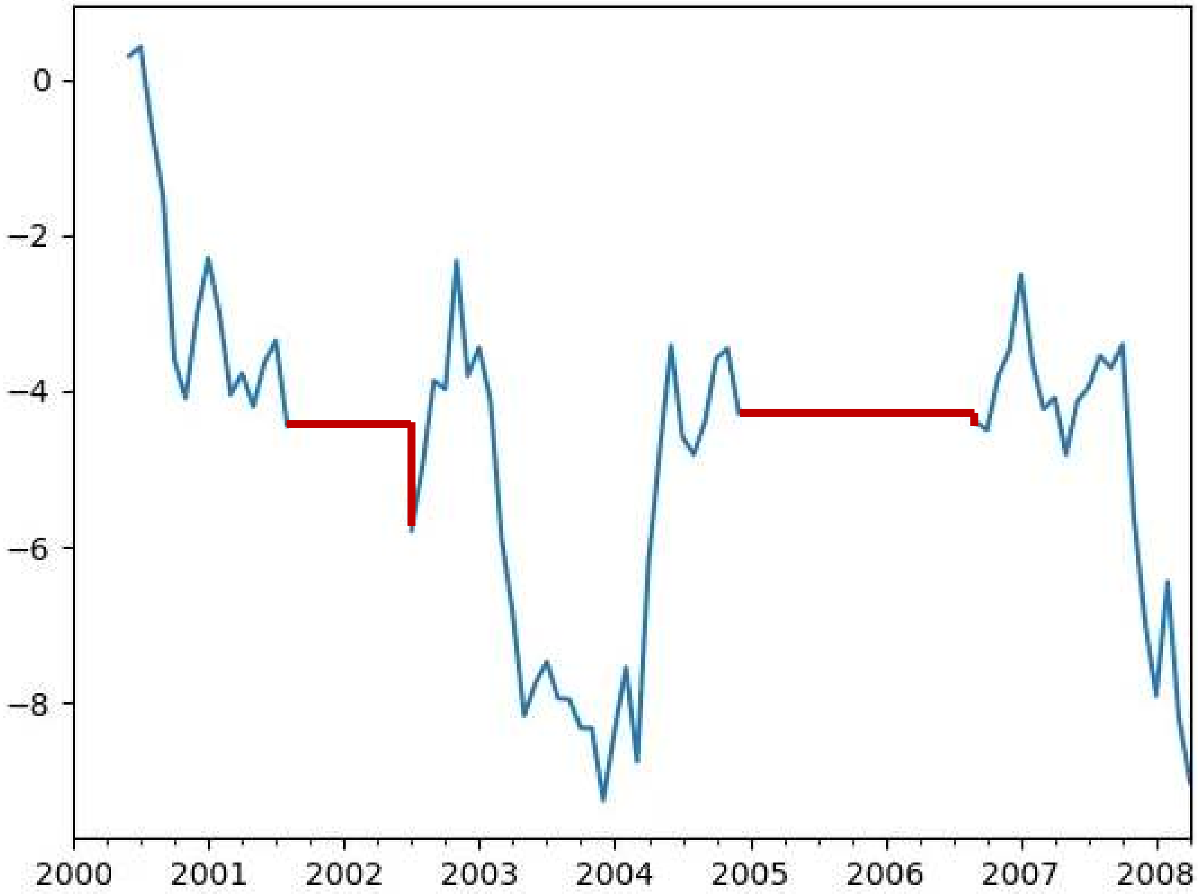


# Missing Values Imputation



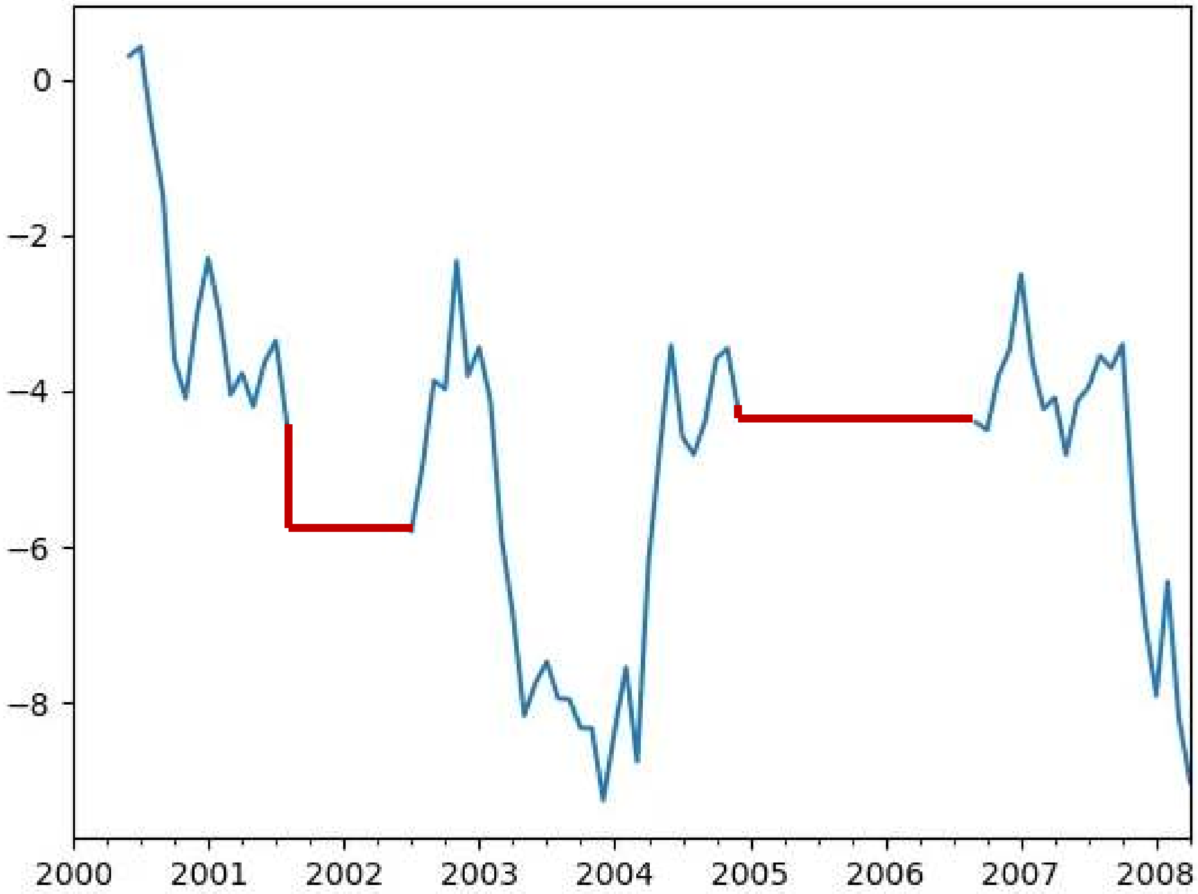
# Missing Values Imputation: ForwardFilling

- Fill with constant value: **last** value (pad)



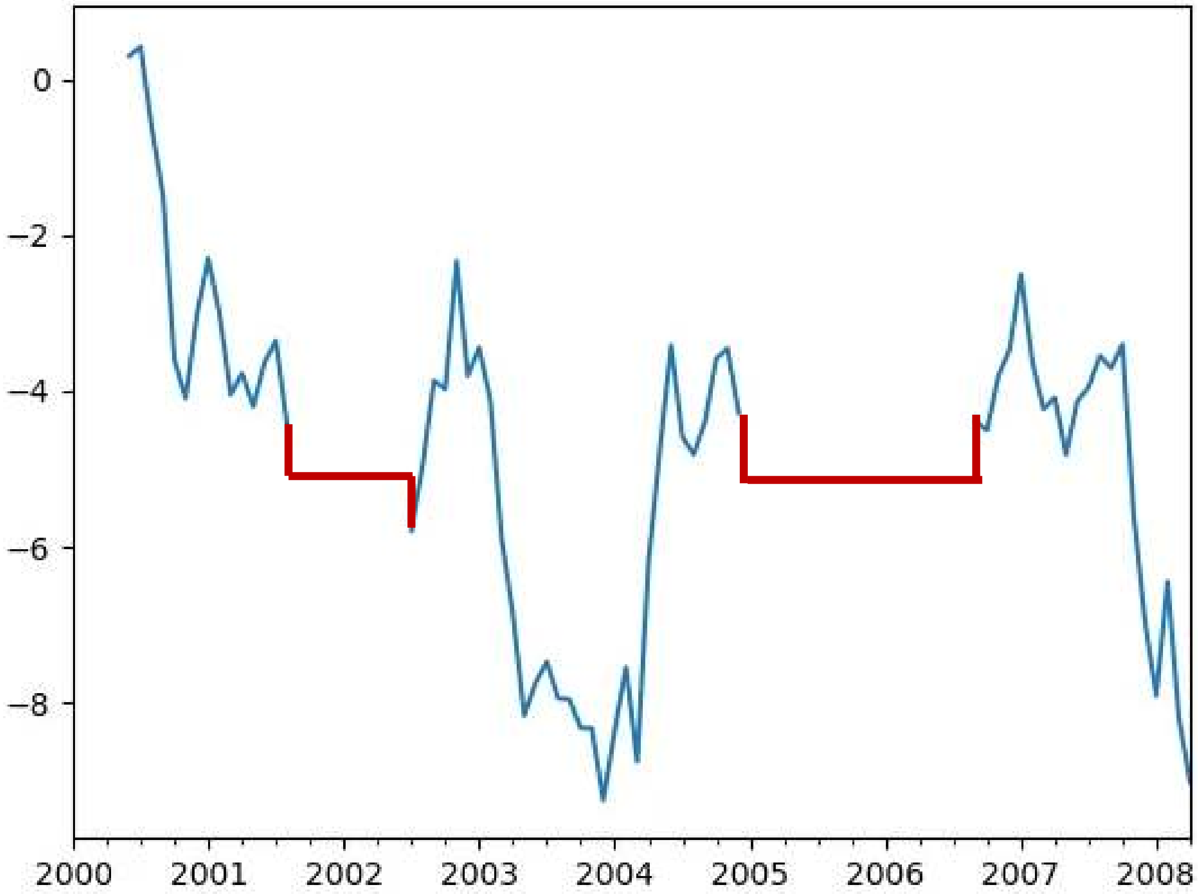
# Missing Values Imputation: BackFilling

- Fill with constant value: **next** value (backfill)



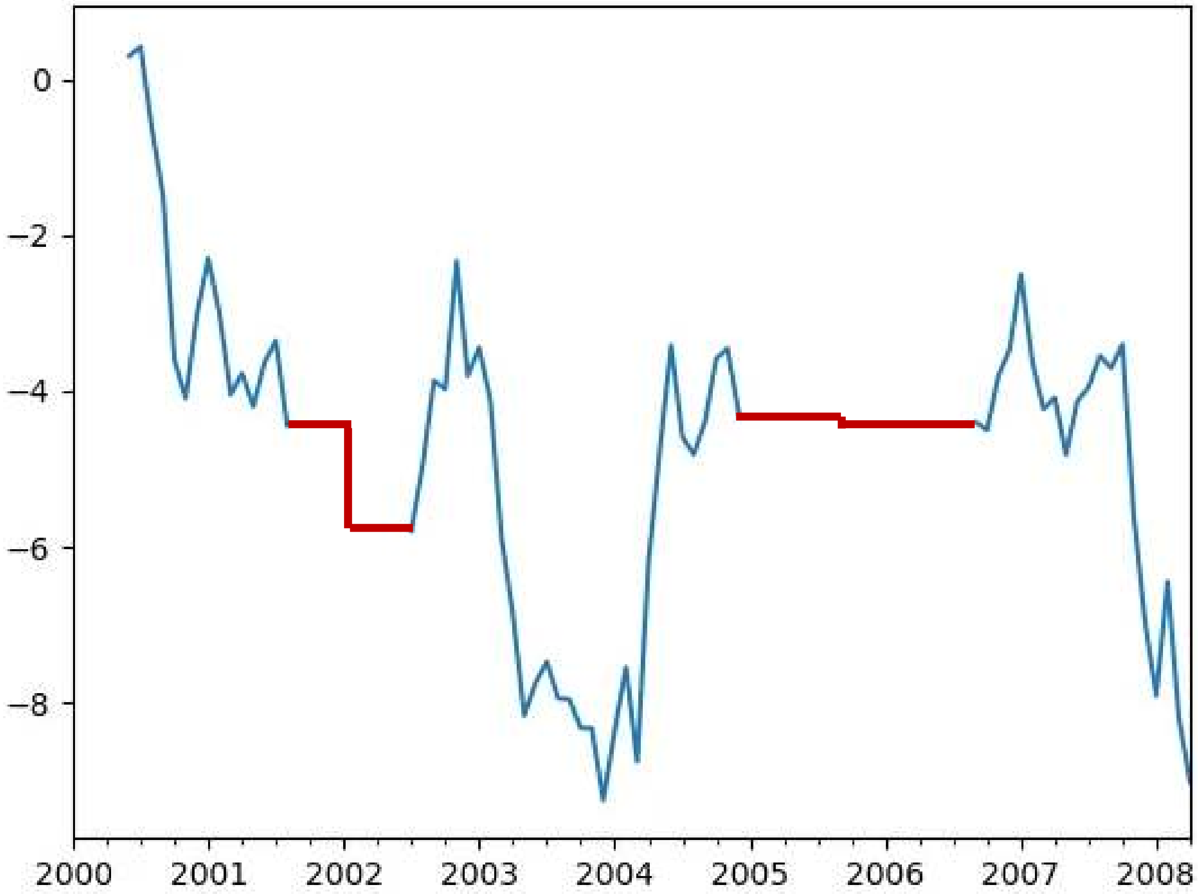
# Missing Values Imputation: Mean

- Fill with constant value: **mean/median** value



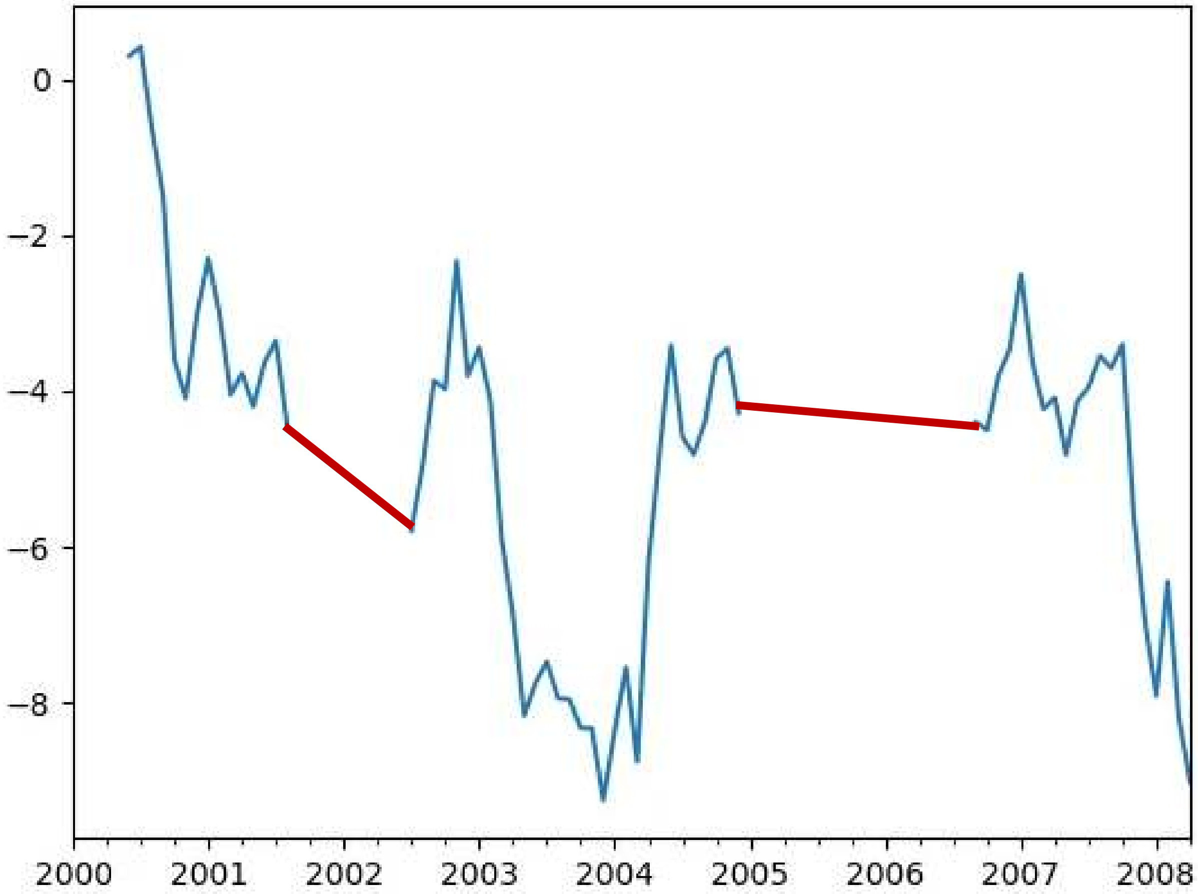
# Missing Values Imputation: Nearest Value

- Fill with constant value: **nearest** value



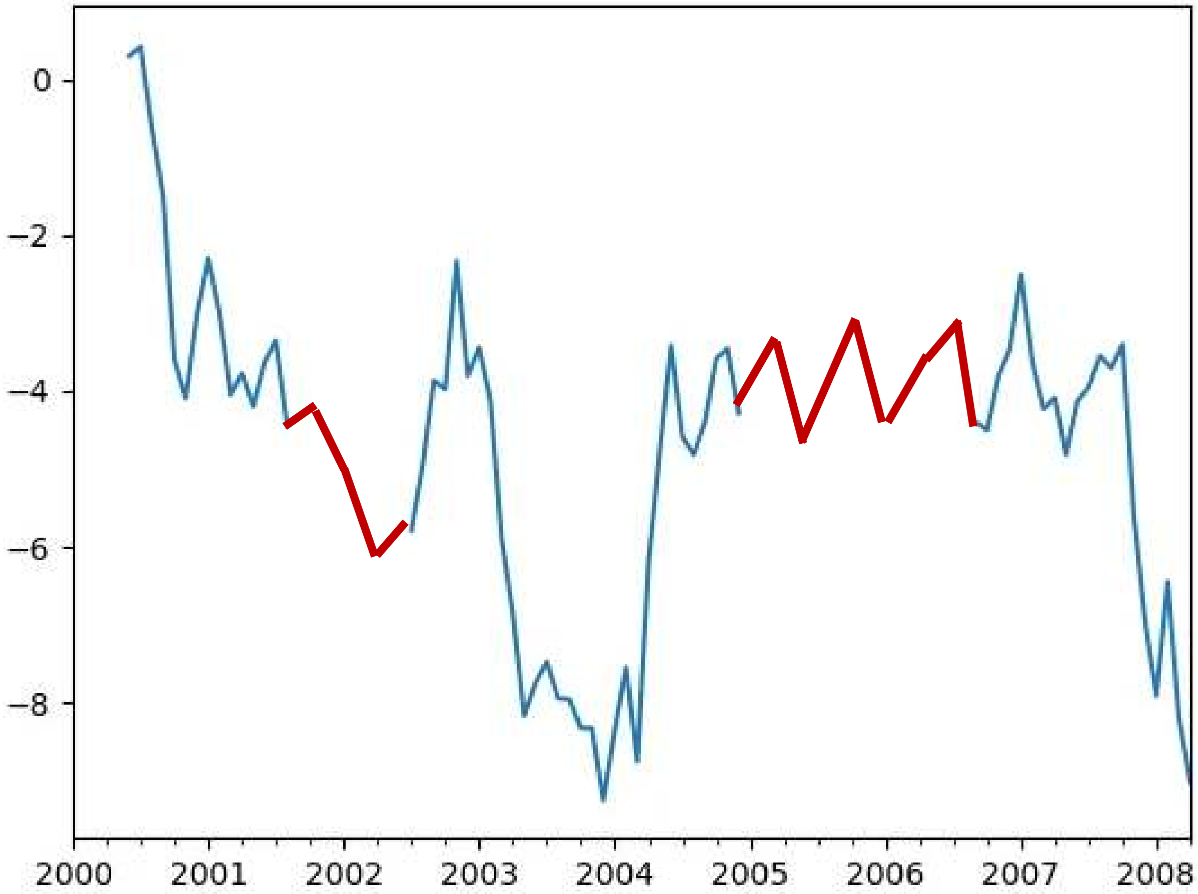
# Missing Values Imputation: Interpolation

- Interpolate the last and first not missing values to get the missing ones.



# Missing Values Imputation: Forecasting

- Interpolate using a forecasting or regressive model (see next lectures)



# Anomalies

# Anomalies and Outliers in Time Series

- An outlier (or anomaly) is a value or an observation that is distant from other observations, **a data point that differs significantly from other data points**.
- Outlier: “An observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism.” [Hawkins 1980]
- Sometimes it makes sense to formally distinguish **two classes of outlie**rs: Extreme values and mistakes.
- **Mistakes can be considered as missing values and removed**.
- **Extreme values might be considered in the analysis**.

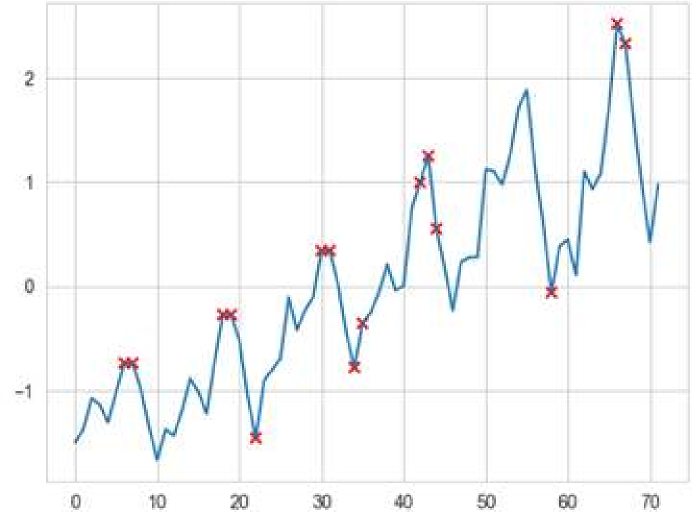
# Outlier Detection Methods

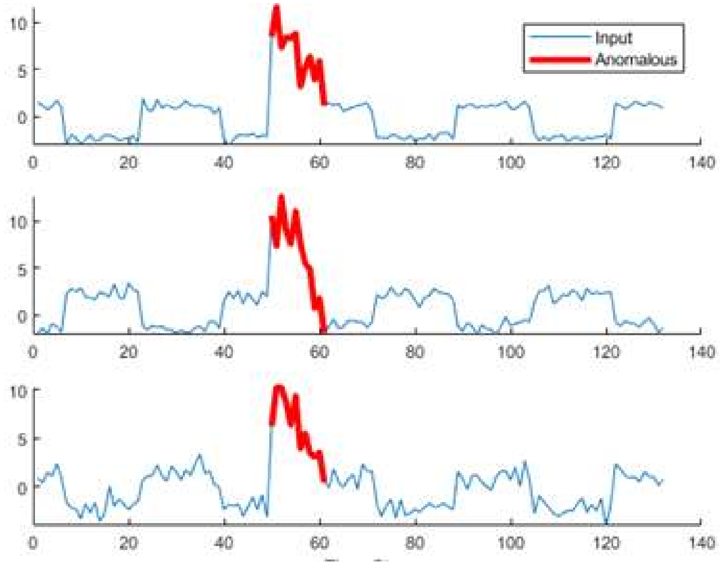
- Outlier detection methods **create a model of the normal patterns in the data**, and then compute an outlier score of a given data point based on the deviations from these patterns.
- Different models make different assumptions about the “normal” behavior.
- The outlier score of a data point is then computed by **evaluating the quality of the fit between the data points and the model**.
- In practice, the choice of the model is often dictated by the analyst‘s understanding of the kinds of deviations relevant to an application.

# Outlier Types in Time Series

Outlier detection methods may differ depending on the type of outliers:

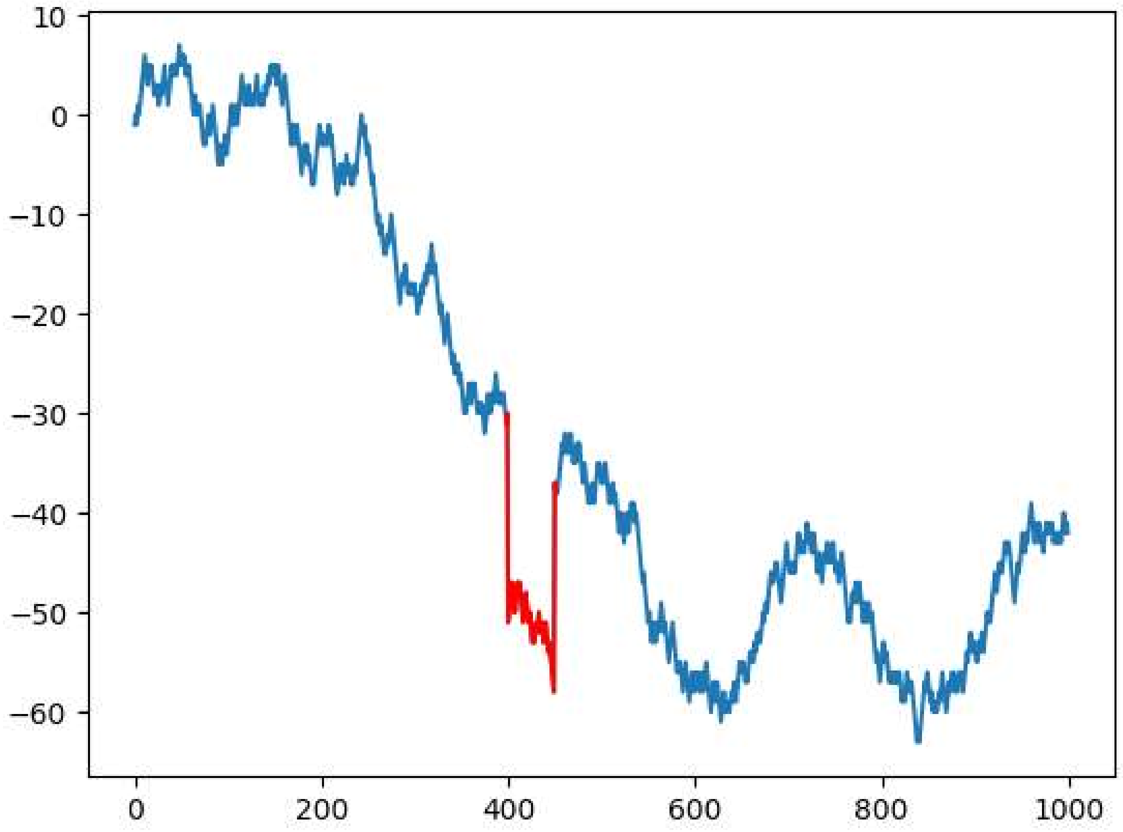
- **Point Outlier**: A value that behaves unusually in a specific time instant when compared either to the other values in the time series (global outlier) or to its neighboring points (local outlier).
- **Subsequences**: Consecutive points in time whose joint behavior is unusual, although each observation individually is not necessarily a point outlier
- **Instance**: Entire time series can also be outliers, but they can only be detected when the input data is a dataset of time series.





# Examples of Anomalies in Time Series

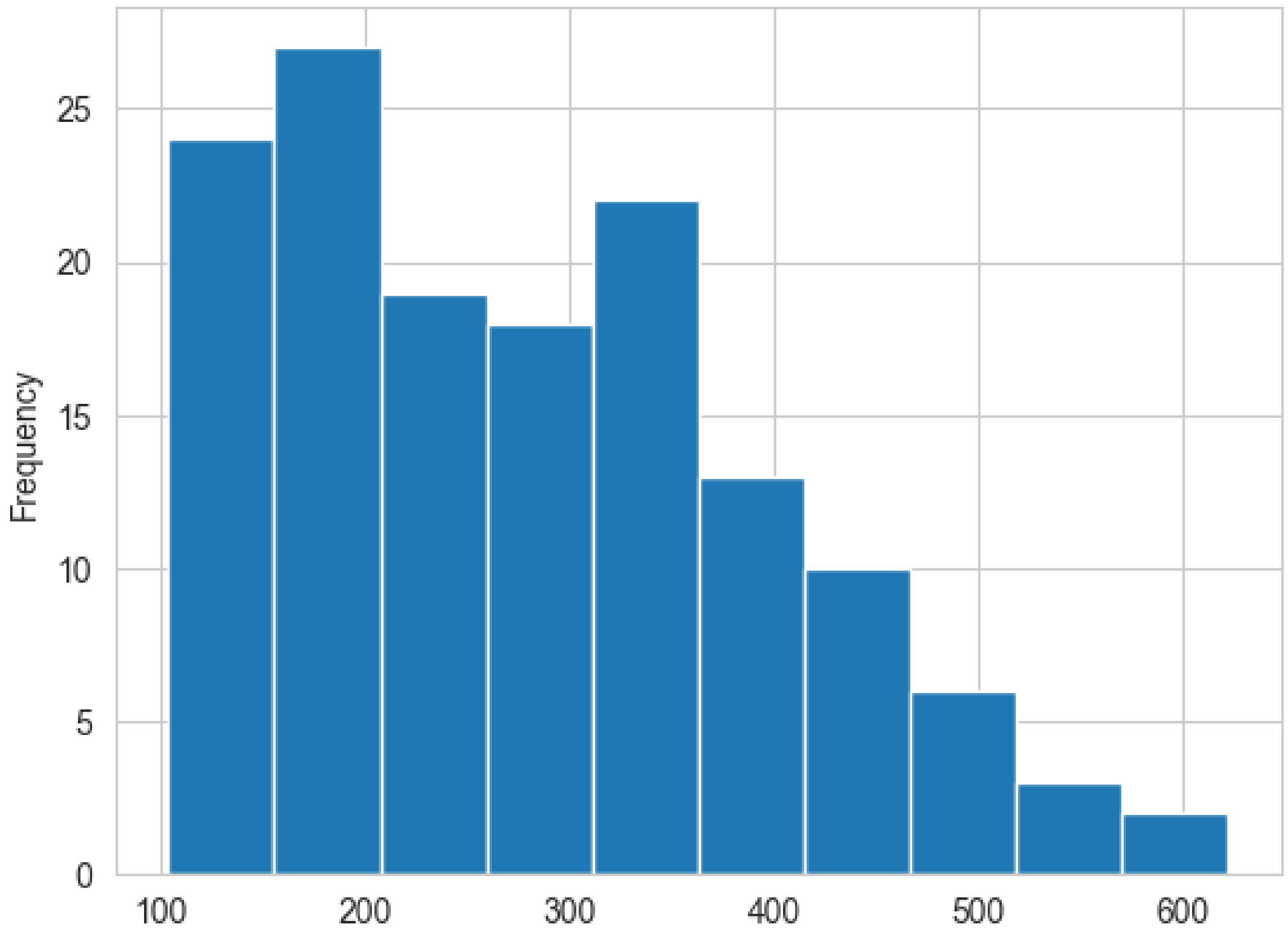
- **Level shifts:** when the signal moves to zero/default value for a short period of time.
- **Unexpected growth/decrease** in a short period of time that looks like a spike.



# Outlier Detection Method

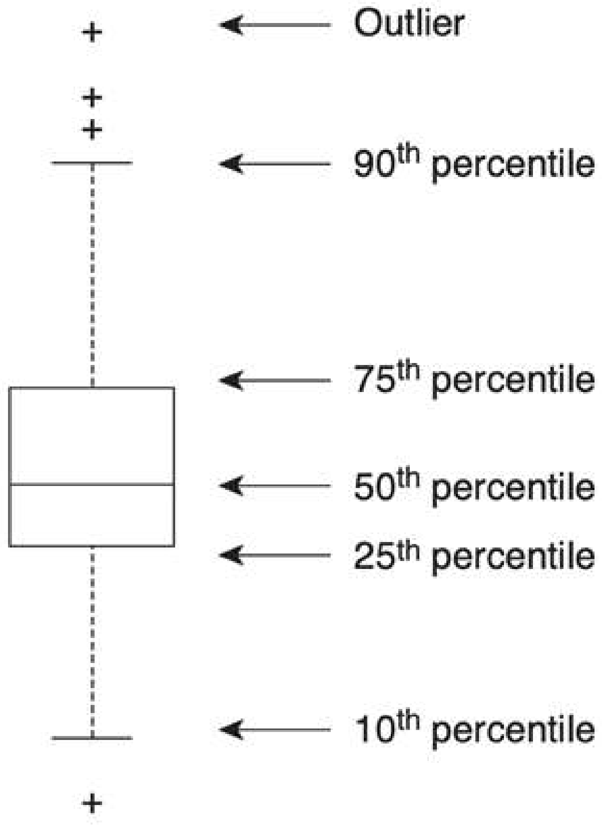
- An outlier is a point that significantly deviates from its expected value.
- Given a univariate time series *x*, a point at time *t* can be declared an outlier if the distance to its expected value is higher than a predefined threshold.
- **Estimation method**: if $\bar{x}_t$ is obtained using previous and subsequent observations to $x_t$ (past, current, and future data).
- **Prediction method**: if $\bar{x}_t$ is obtained relying only on previous observations to $x_t$ (past data).

# Histogram



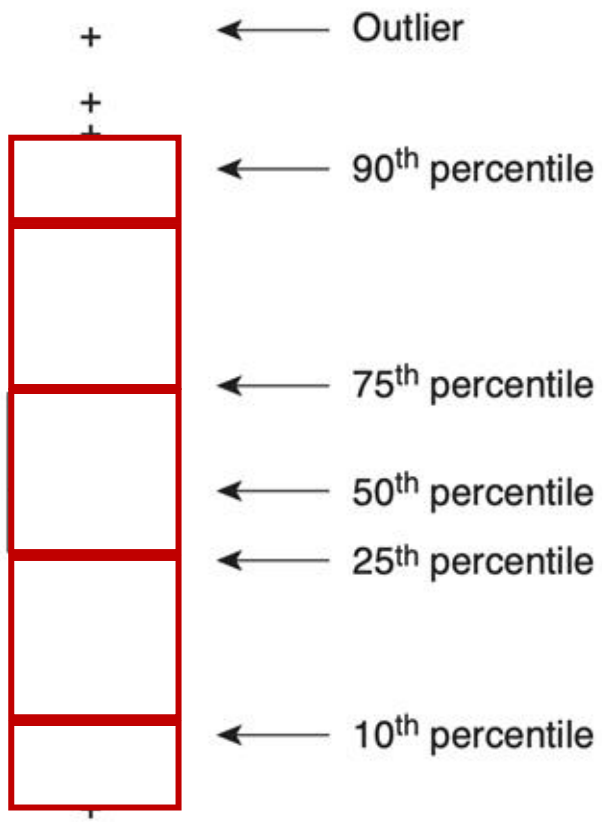
# Boxplot

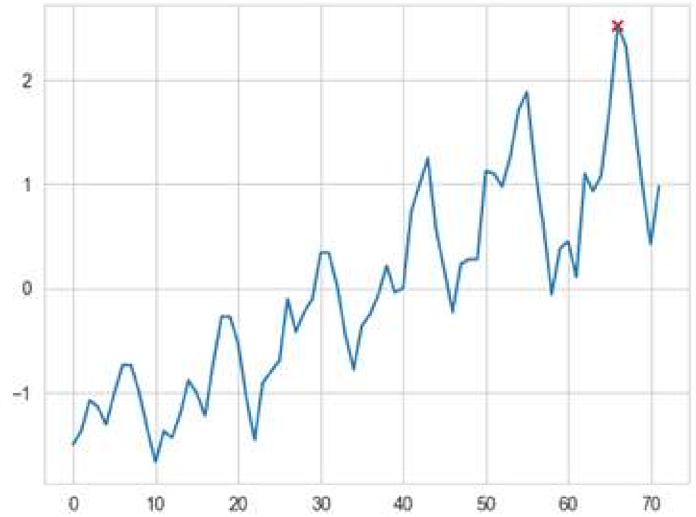
- Data represented with a **box**
- The ends of the box are at the
  - **Q1: 1$^{st}$ quartiles** (25%-quantile or 25$^{th}$ percentile)
  - **Q3: 3$^{rd}$ quartiles** (75%-quantile or 75$^{th}$ percentile)
- **Median**: value in the middle is the **Q2**: **2$^{nd}$ quartile** (50%-qua,, 50$^{th}$ perc.)
- The height of the box is **Interquartile range** (**IQR**): Q3 - Q1
- **Whiskers**: two lines outside the box extended from:
  - 1$^{st}$, or 5$^{th}$, or 10$^{th}$ percentile, or Q1 – *k* IQR (with k = 1.5)
  - 99$^{th}$, or 95$^{th}$, or 90$^{th}$ percentile, or Q3 + *k* IQR (with k = 1.5)
- **Outliers (fliers)**: are points beyond whiskers
- In general, p%-quantile (0 < p < 100): Is the value *x* s.t. p% of the values are smaller and 100-p% are larger.



# Inter-Quartile Range Filter

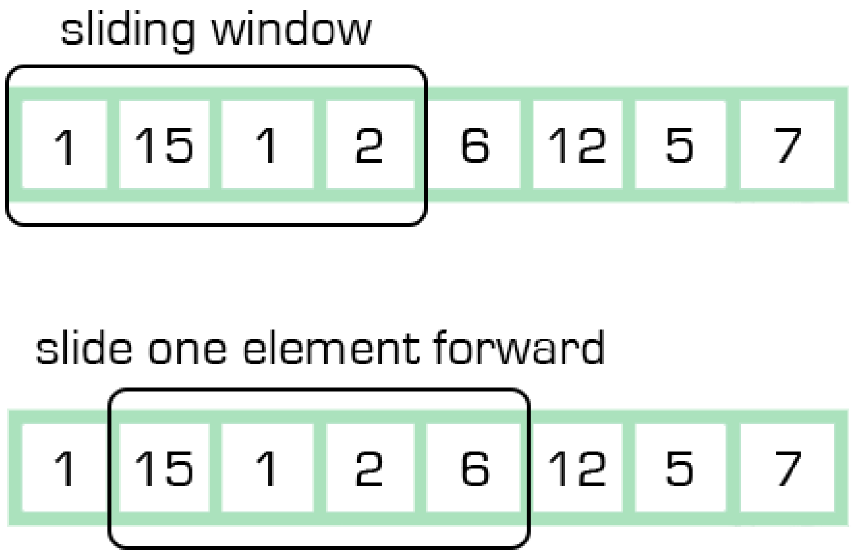
- The IQR criteria means that all observations
- above Q3 + 1.5 $*$ IQR or
- below Q1 − 1.5 $*$ IQR
- where Q3 and Q1 correspond to third ($q_{0.75}$) and first ($q_{0.25}$) quartile respectively, and $IQR$ is the difference between the third and first quartile) are considered as potential outliers.



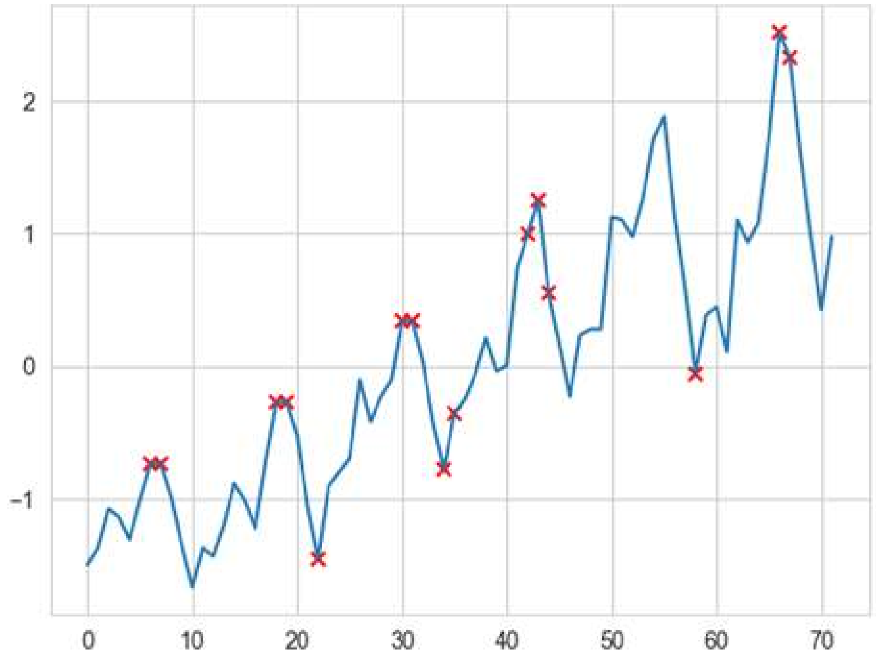
# Sliding Window

- A sliding window is a **fixed-size segment that moves across a time series**, letting us analyze or use local patterns one portion at a time.
- Given a time series channel $\mathbf{x}$ and a window length L, a sliding window is a contiguous subsequence of length L of the form $(x_t, x_{t+1}, \dots, x_{t+L-1})$, where the starting index t is shifted forward along the series.

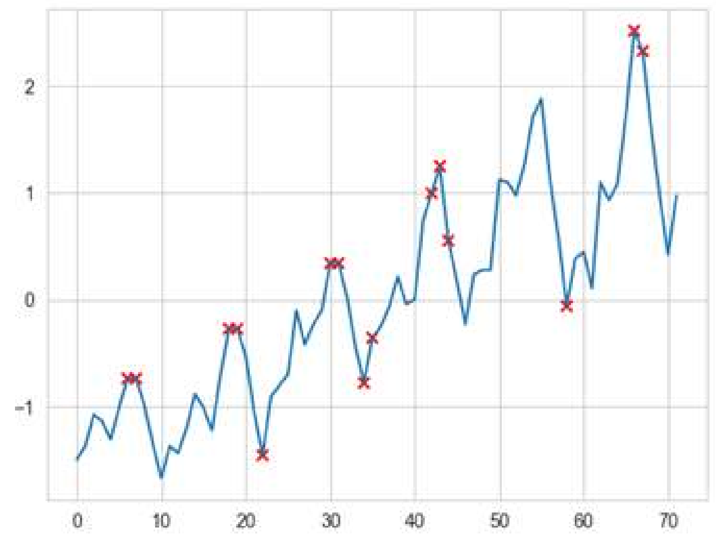


# Hampel Filter

- Consider as outliers the values outside the interval ($I$) formed by the median, plus or minus 3 Median Absolute Deviations (MAD), using a **sliding window**:
- $I$=[$median_X - 3 * MAD; median_X + 3 * MAD$]
- where $MAD$ is the median absolute deviation and is defined as the median of the absolute deviations from the data's median, i.e.,
- $MAD = median(|x_i - median_X|)$



# Grubbs' Test




- Detect outliers in univariate data
- Assume data comes from normal distribution
- Detects one outlier at a time, remove the outlier, and repeat (two-sided test with alpha/2N)
  - $H_0$: There is no outlier in data (all observations come from the same normal population.)
  - $H_A$: There is at least one outlier
- Grubbs' test statistic: $G = \frac{\max|X - \overline{X}|}{s}$

mean

std dev

- Reject null hypothesis $H_0$ of no outliers if: $G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t^2_{\alpha/(2N),N-2}}{N-2+t^2_{\alpha/(2N),N-2}}}$

upper critical value of t-distribution

alpha significance 2-sided

Sample size

degrees of freedom

# Handling Outliers in Time Series

Once that anomalies have been identified

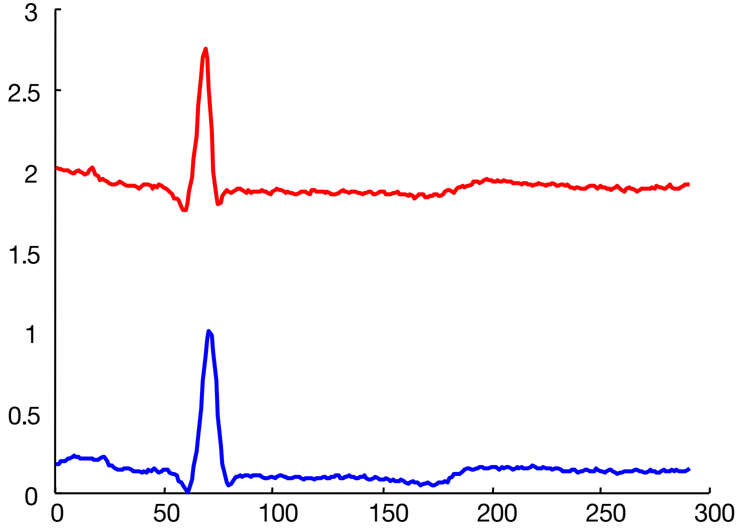
a) they can be removed and treated like missing values, i.e., replaced according to different strategies.
b) they can be maintained in the analysis.
c) they can be distinguished between “mistakes” and “extreme values” and treated according the most preferrable strategy.

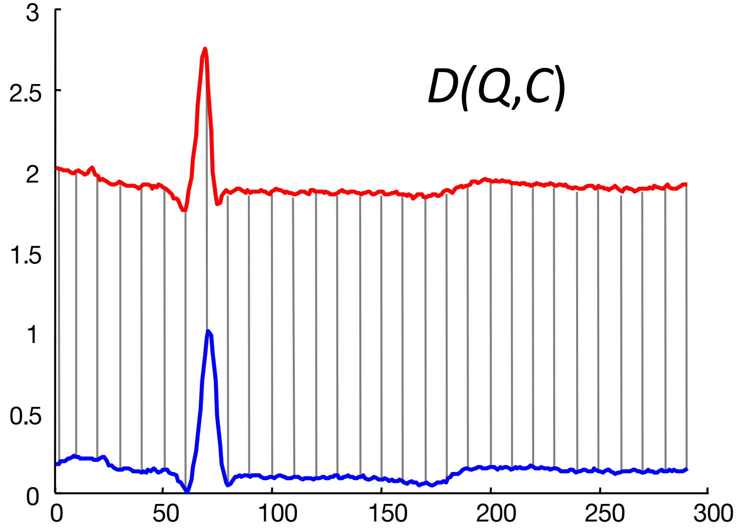
# Normalizations

# Problem with Time Series Distortions

- A common tool of TSA consists in **calculating distances among time series**.
- **Many ML-tools require that** the input data is represented in the same range of values or that **values are comparable**.
- **Distance calculations** and ML models are **very sensitive to “distortions”** in the data.
- These distortions are dangerous and should be removed.
- Most common distortions:
  - Offset Translation
  - Amplitude Scaling
  - Linear Trend
  - Presence of Noise
- These distortions can be removed by using the appropriate normalizations.

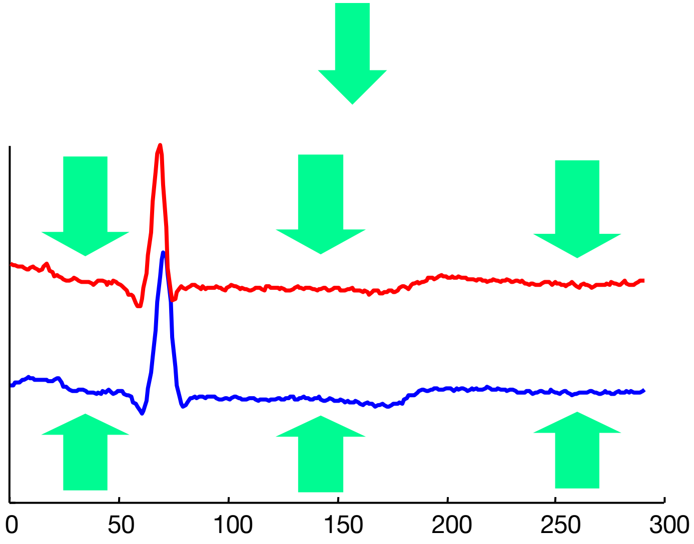
# Offset Translation: Mean Removal

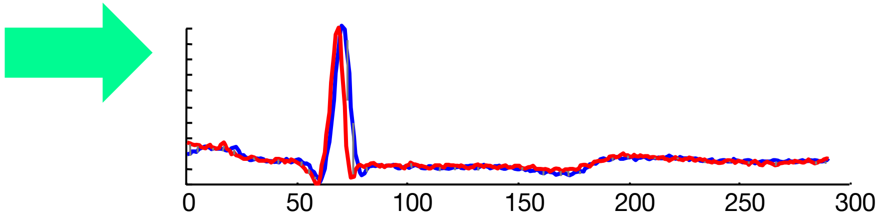


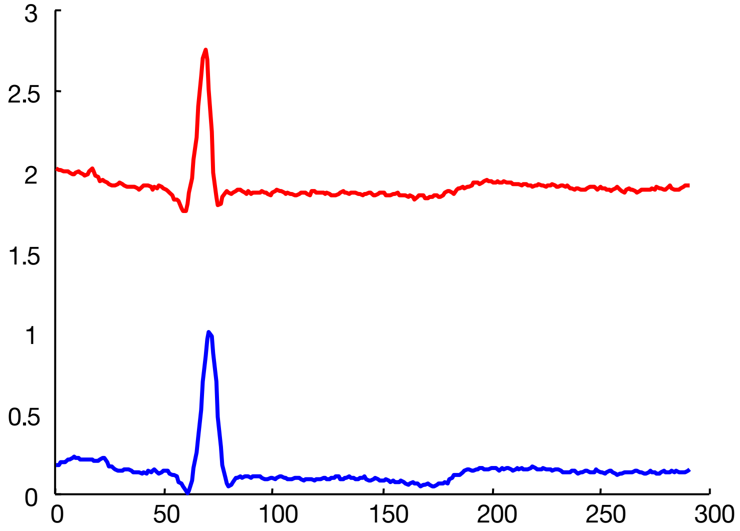

*Q = Q - mean(Q)*

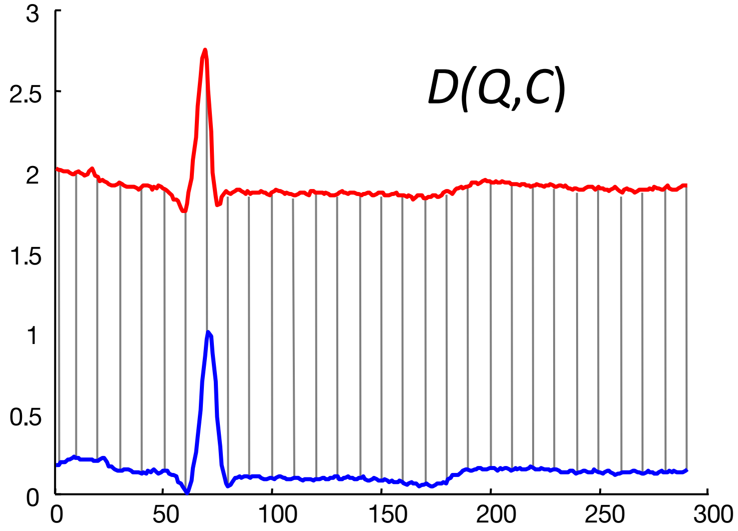
*C = C - mean(C)*

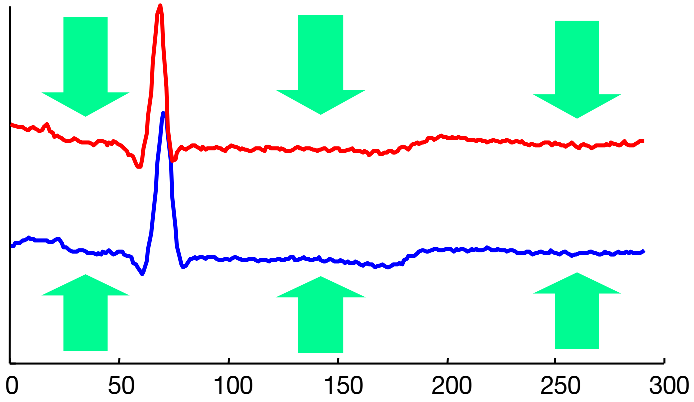


# Offset Translation: Min-Max Normalization

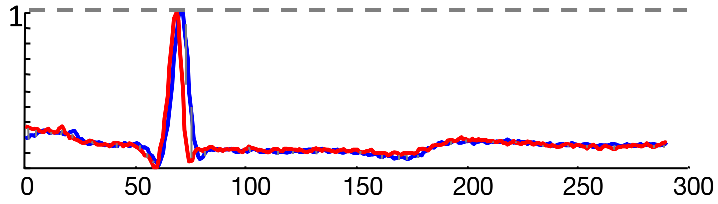


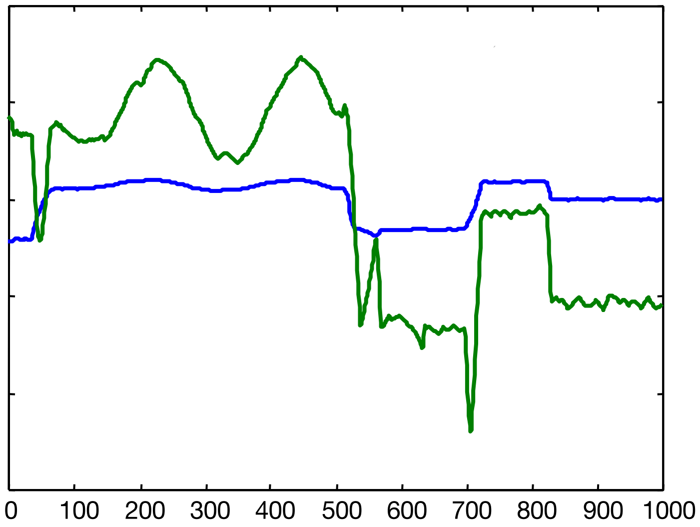



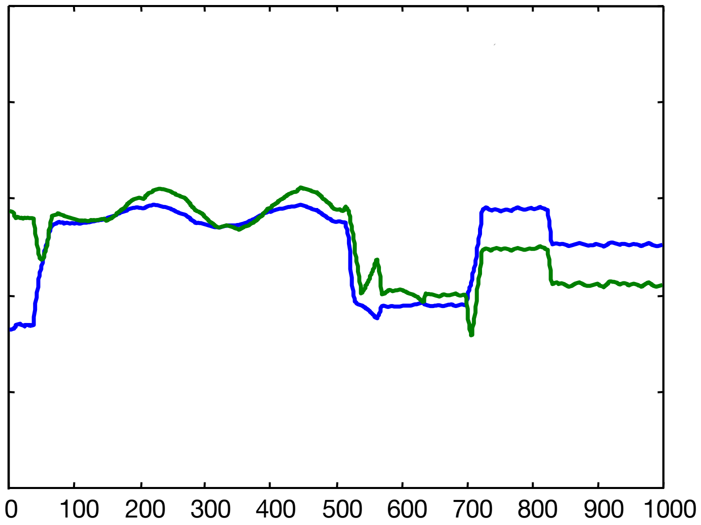

*Q = (Q – min(Q)) / (max(Q) – min(Q))*

*C = (C – min(C)) / (max(C) – min(C))*



# Amplitude Scaling: Z-Score Normalization



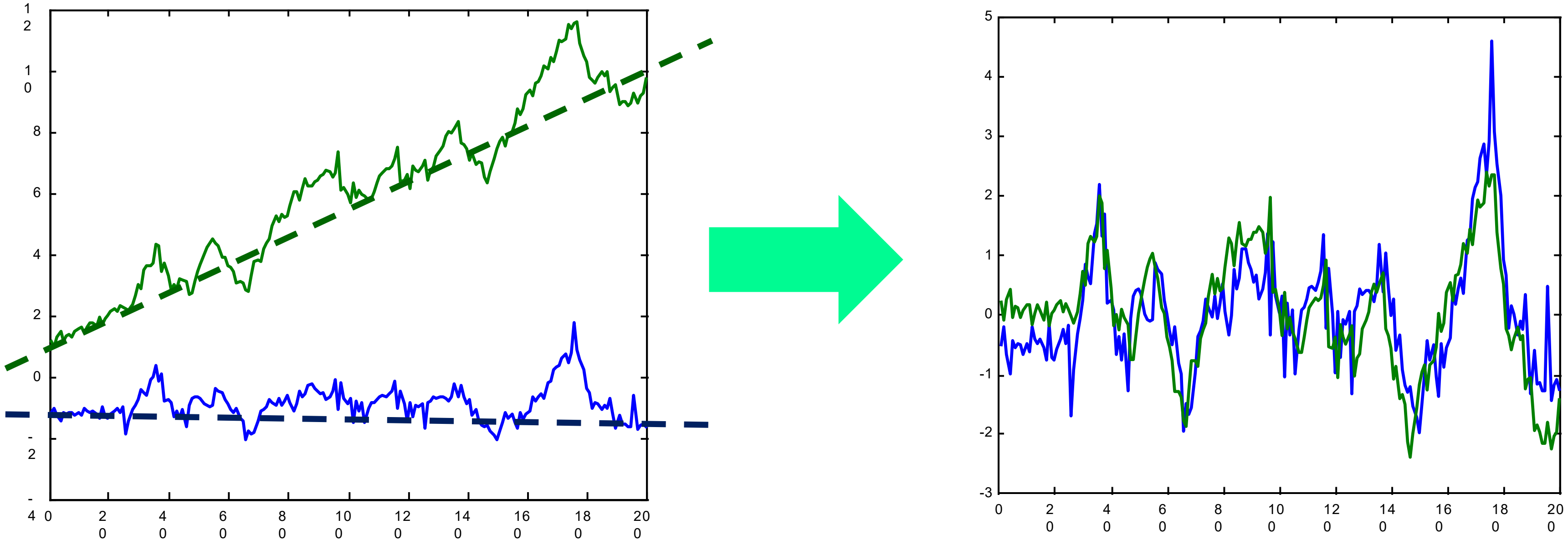




*Q = (Q - mean(Q)) / std(Q)*

*C = (C - mean(C)) / std(C)*

# Linear Trend: Detrending

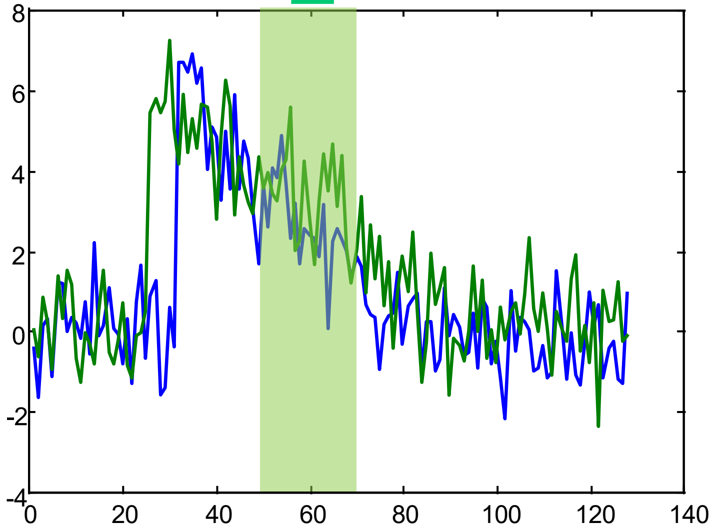
- Fit the best fitting straight line to the time series, then subtract that line from the time series.

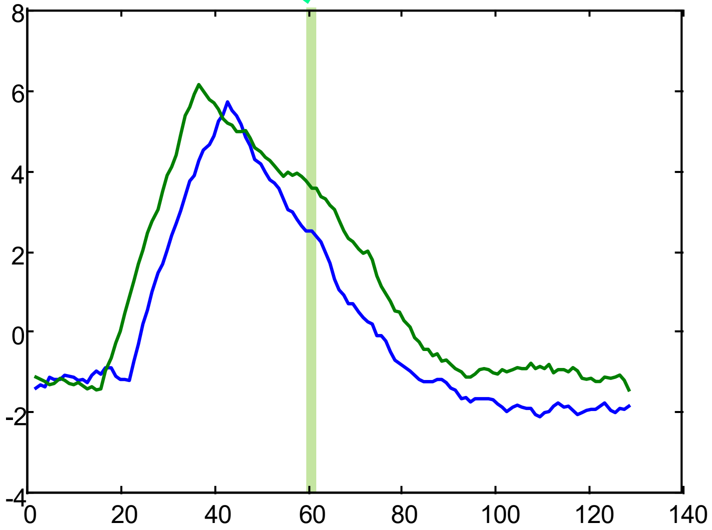


Removed linear trend,
offset translation,
amplitude scaling

# Noise Removal: Mean Smoothing

- The intuition behind removing noise is to average each datapoints value with its neighbors.




*Q = smooth(Q)*

*C = smooth(C)*

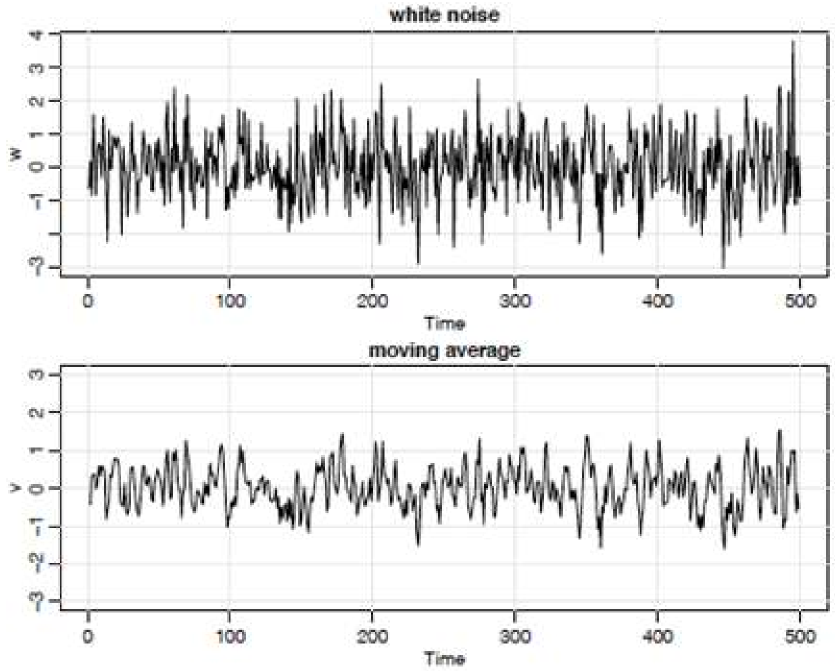
*D(Q,C)*

# Moving Average

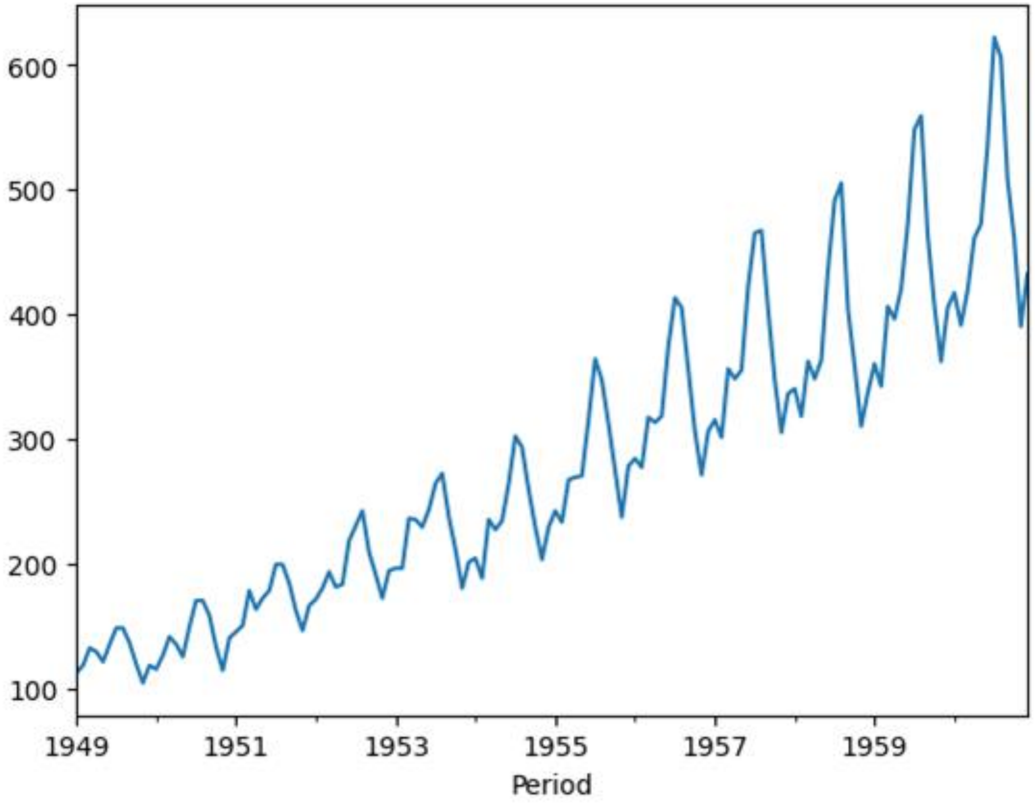
w=3

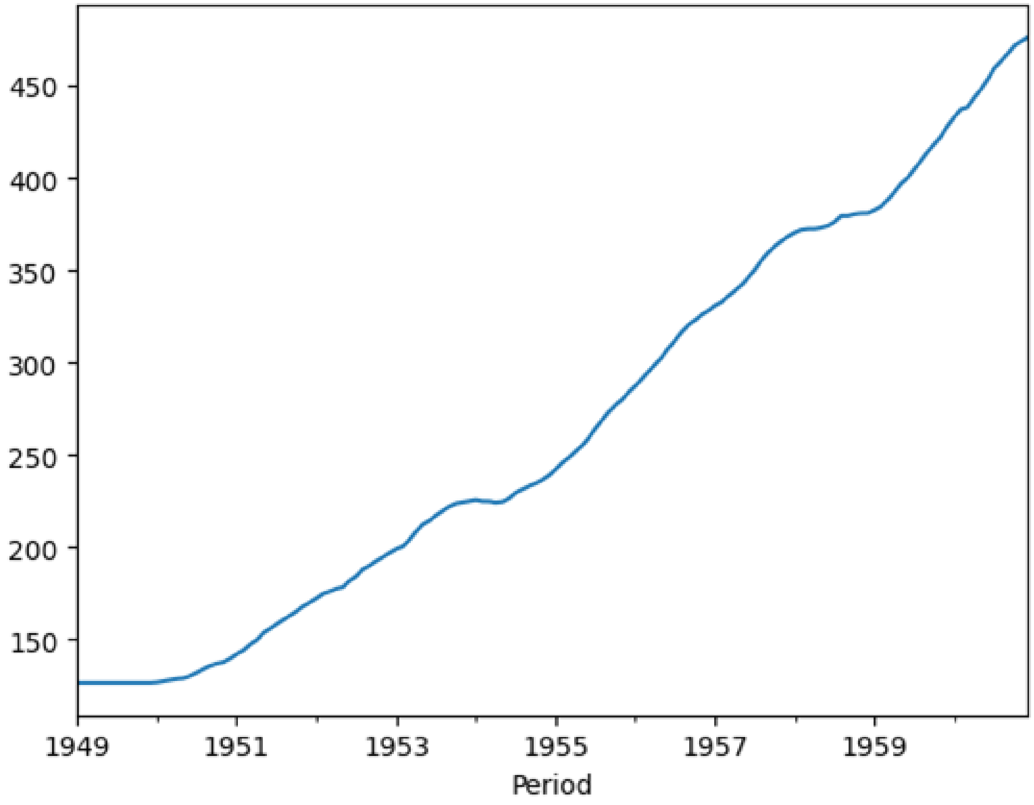
| time | value | ma |
|---|---|---|
| t1 | 20 | - |
| t2 | 24 | 22.0 |
| t3 | 22 | 24.0 |
| t4 | 26 | 24.3 |
| t5 | 25 | - |

- Noise can be removed by a **moving average** (MA) that smooths the TS.
- Given a window of length *w* and a TS *t*, the MA is applied as follows
- $t_i = \frac{1}{w}\sum_{j=i-w/2}^{w/2} t_j$ for $i = 1, \dots, n$
- For example, if w=3 we have
- $t_i = \frac{1}{3}\left(t_{i-1} + t_i + t_{i+1}\right)$



# Moving Average Example



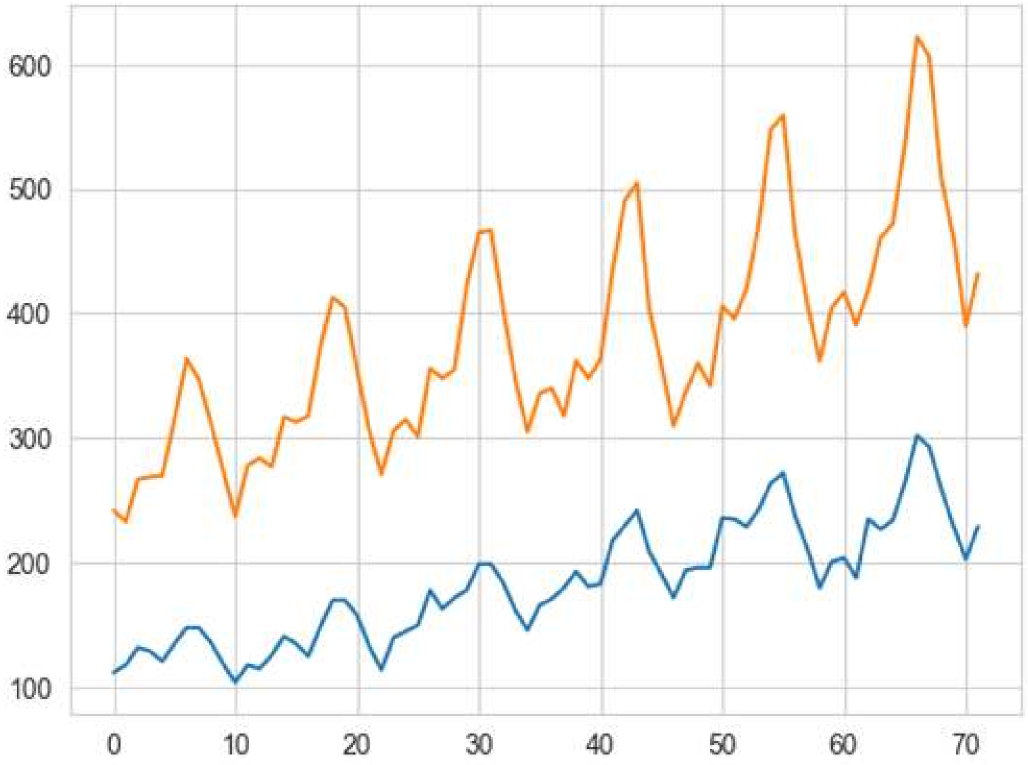



# Log Transformation

- We apply the logarithm to each value of the TS.

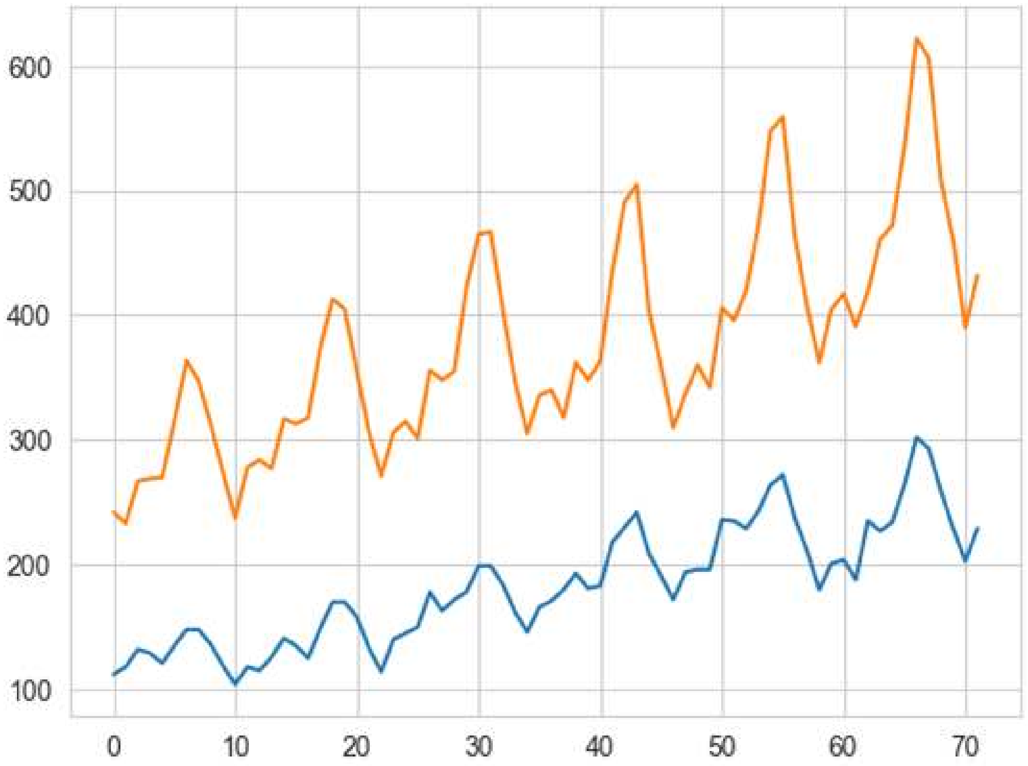
***Log***
*Q = log(Q)*
*C = log(C)*

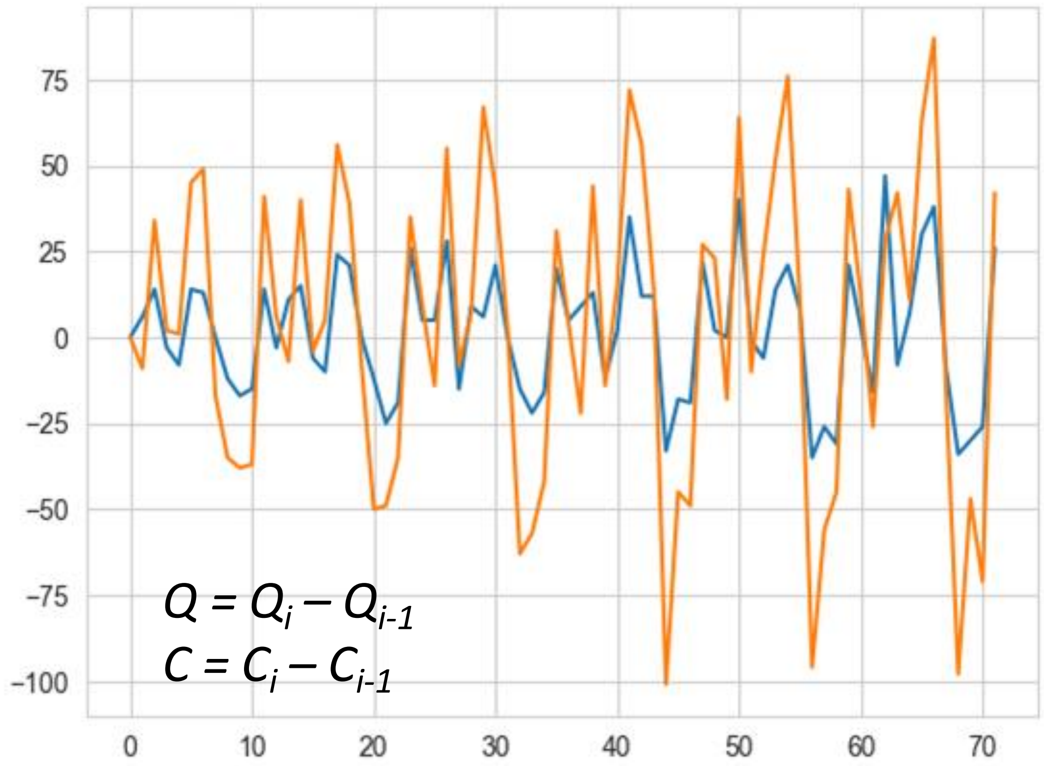
***Log1p***
*Q = log(Q + 1)*
*C = log(C + 1)*




# Differencing Transformation

- **Differencing**: we take the difference of the observation at a particular instant with that at the previous instant.



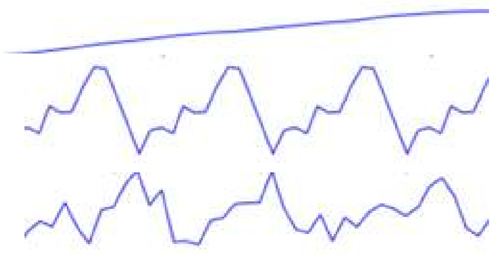
# Time Series Components

# Assumption

- Various Data Mining, Machine Learning and TSA models assume that variables are Independent and Identically Distributed (IID)
  - In times series values are (usually) not independent
  - Trend and seasonality might be present
  - Variance may change significantly (heteroskedasticity)
- A first goal in TSA is to reduce the time series to a simpler case:
  - Eliminate trend
  - Eliminate seasonality
  - Eliminate heteroskedasticity
- Then we model the remainder as dependent but identically distributed variables.

# Time Series Components

- A given TS consists of three systematic components including level, trend, seasonality, and one non-systematic component called noise.
  - **Level**: The average value in the series.
  - **Trend**: The increasing or decreasing value in the series.
  - **Seasonality**: The repeating short-term cycle in the series.
  - **Noise**: The random variation in the series.



- A **systematic** component has consistency or recurrence and can be described and modeled.
- A **non-systematic** component cannot be directly modeled.

# Time Series Components

- A TS can be modeled as an aggregate or combination of these four components.
- All series have a level and noise. The trend and seasonality components are optional.
- Level can be omitted if Offset Translation with Mean Removal is applied

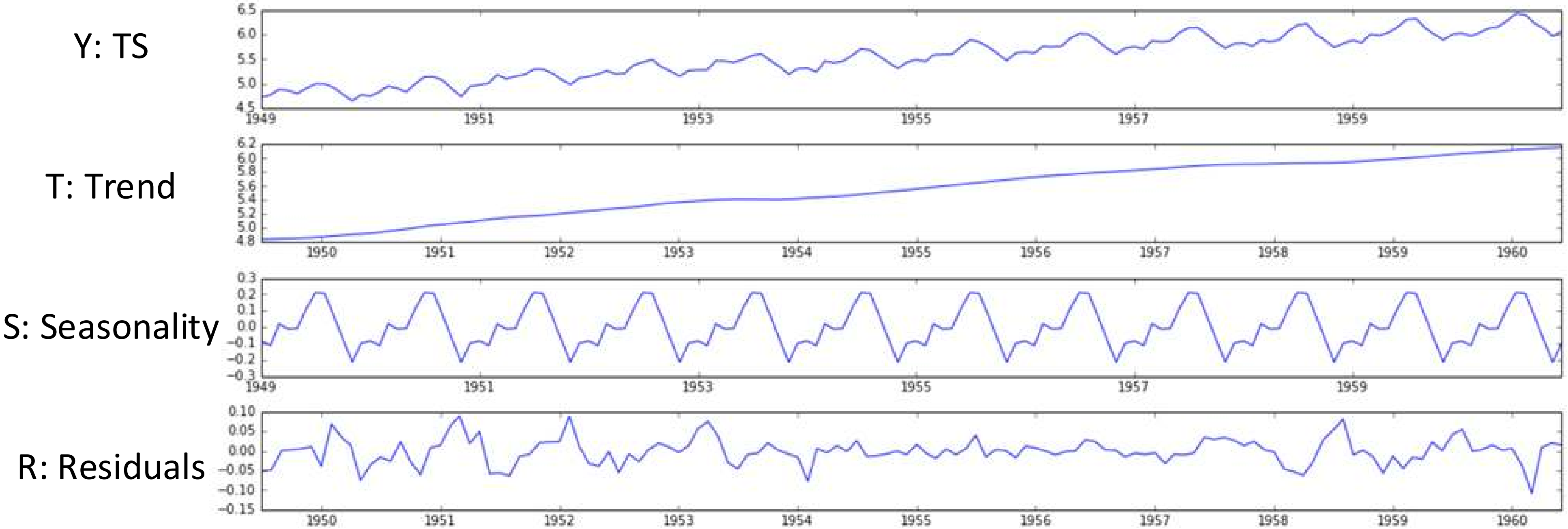
- **Additive Model**: Y = Level + Trend + Seasonality + Noise/Residuals
  - Changes over time are consistently made by the same amount
  - A linear trend is a straight line.
  - A linear seasonality has the same frequency (width of cycles) and amplitude (height of cycles).

- **Multiplicative Model**: Y = Level * Trend * Seasonality * Noise/Residuals
  - A multiplicative model is nonlinear, such as quadratic or exponential.
  - Changes increase or decrease over time.
  - A nonlinear trend is a curved line.
  - A non-linear seasonality has an increasing/decreasing frequency and/or amplitude over time.

# Time Series Components

Components meaning

- Trend: upward or downward pattern that might be extrapolated into the future.
- Seasonality: periodic behavior with a known period (hourly, weekly, monthly...).
- Residuals: containing anything else, i.e., noise.

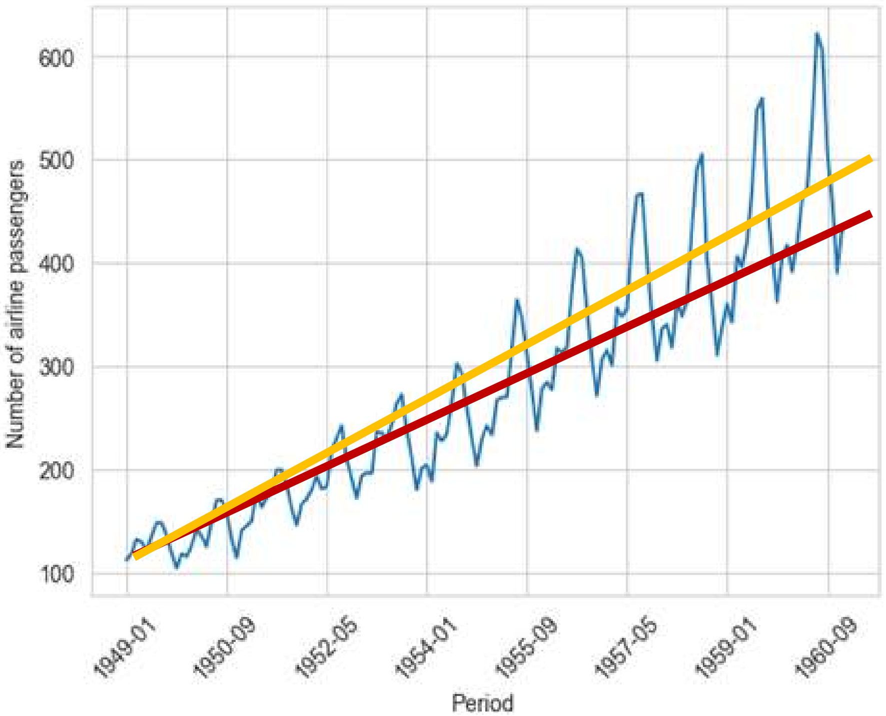



**Additive Model**

$$Y = T + S + R$$

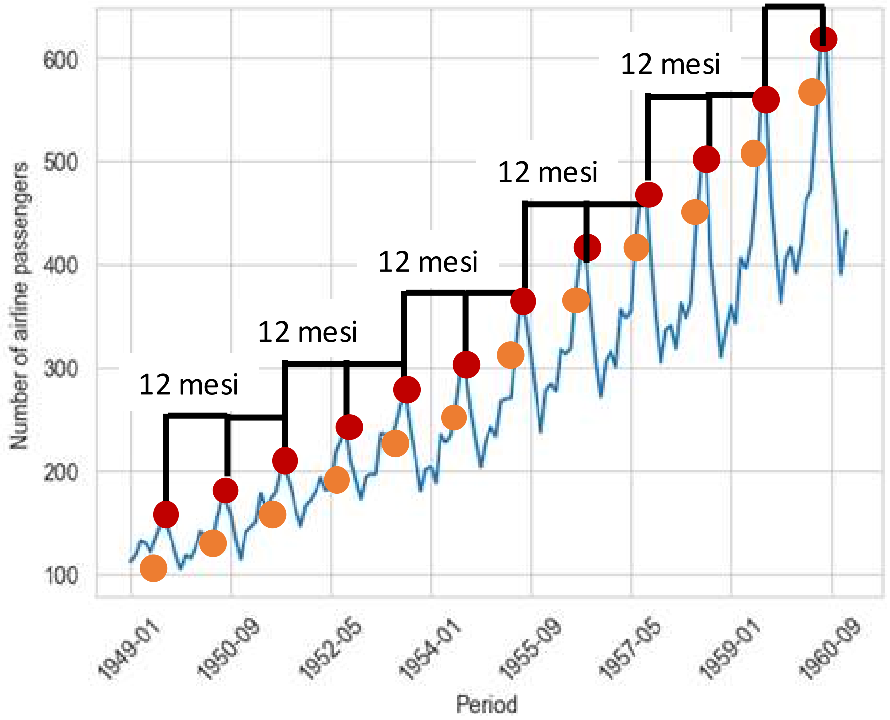
# Time Series Decomposition: Trend

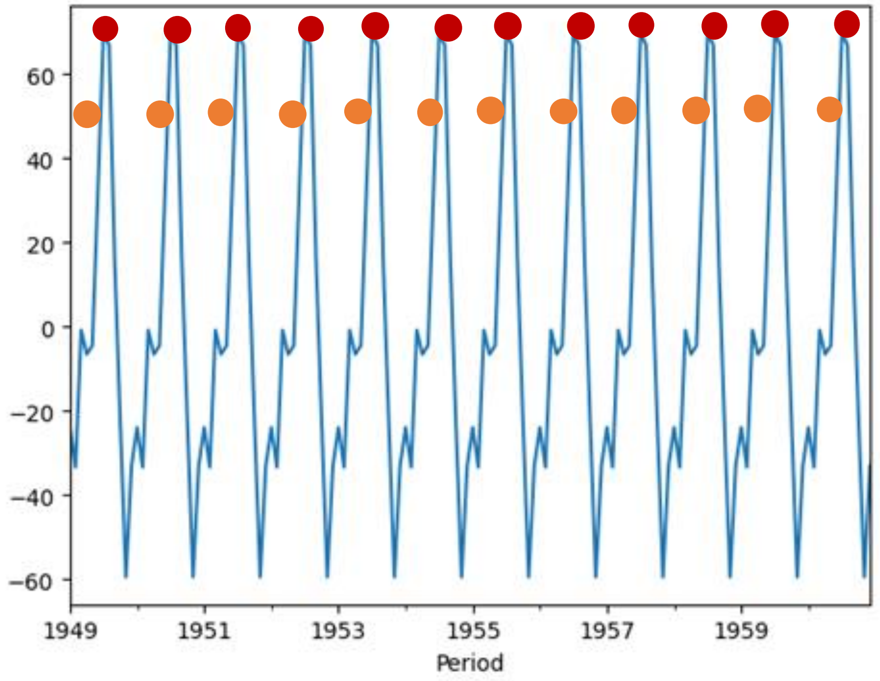
- Trend as a straight line between the first and the last point of the TS.
- Given a window *w*, trend as the moving average along TS with size *w*.
- Trend as the best fitting straight line obtained with a linear regression.



# Time Series Decomposition: Seasonality

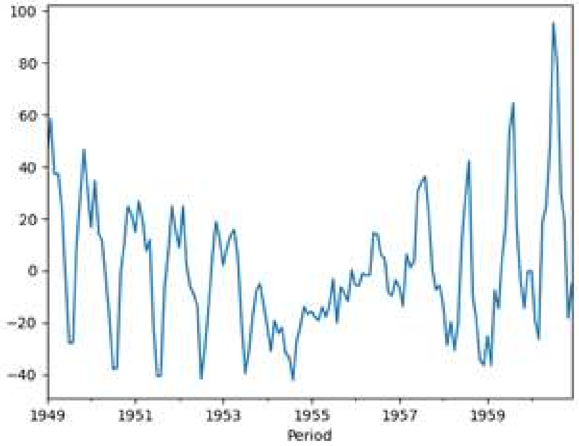
- Given the estimated number of periods *p,* and a detrended time series D = Y – T considering TS modeled with additive models, seasonality can be calculated as the mean value of for each time stamps among the various periods *p* repeated *p* times.



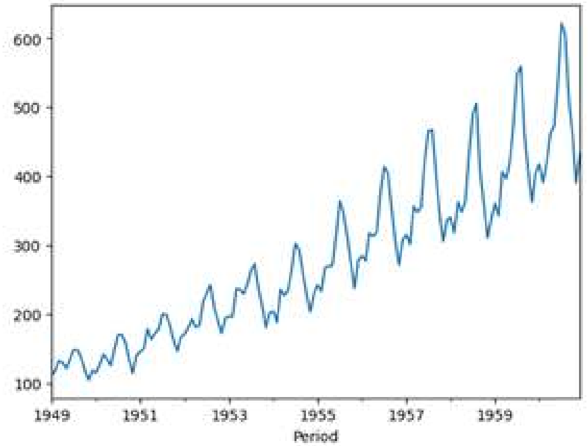



# Time Series Decomposition: Residuals

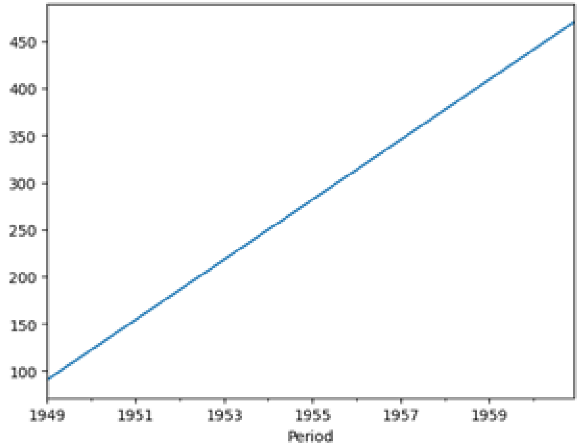
- Given the time series Y, its trend T, its seasonality S, the residuals R are obtained as R = Y – T – S.



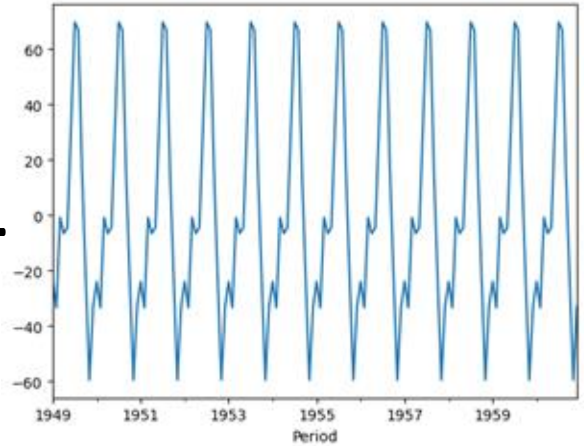

=




–




–



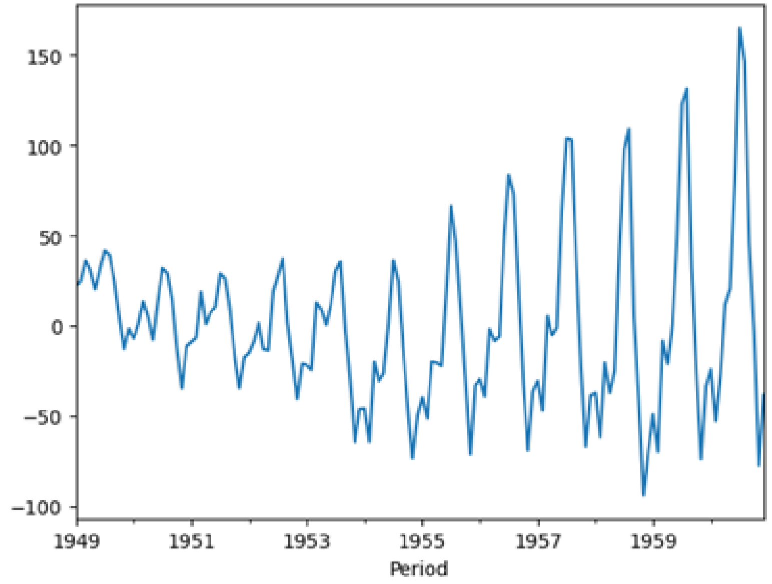
# Component-based Normalizations

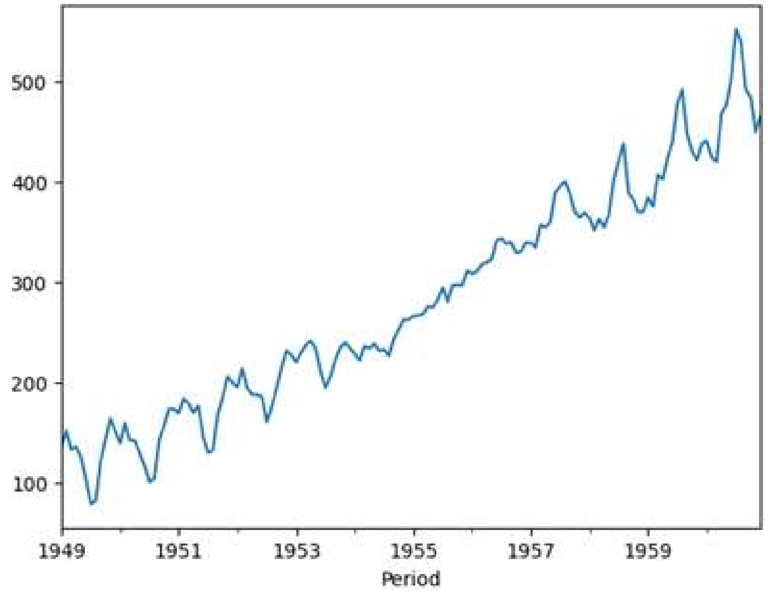
Detrend

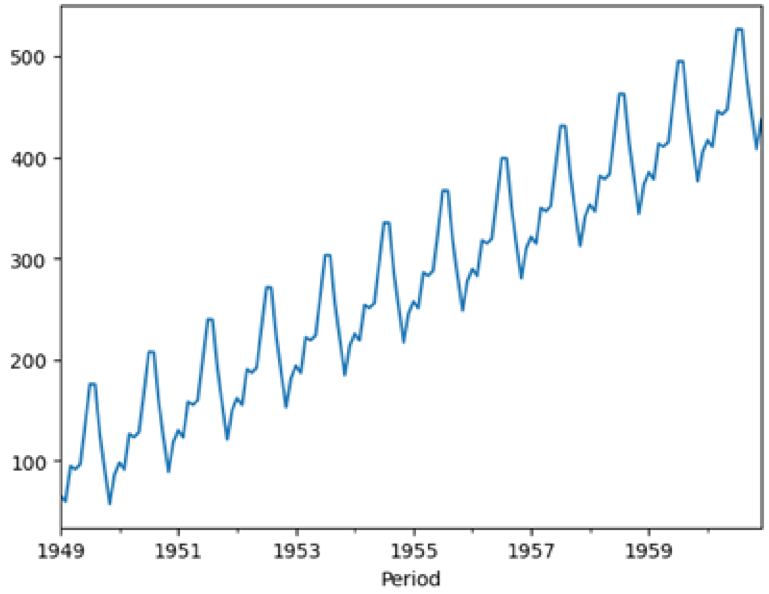
- Y’ = Y – T

Deseasonalize

- Y’ = Y – S



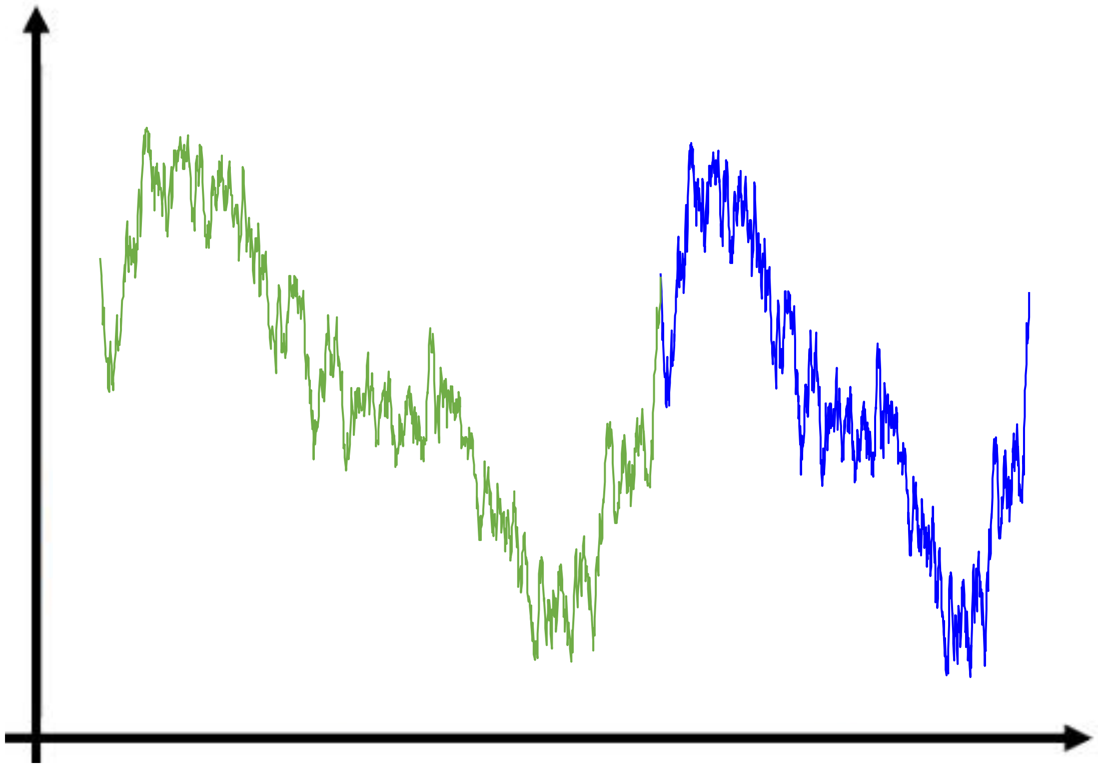



# Stationarity

# Why Do We Care About Stationarity?

- If your TS is not stationary, you cannot build reliable statistical TS predictive models.



# Elementary Statistics (population)

- **Expectation**: The expectation $E(x)$ of a variable $x$ is its mean average value in the population. We denote the expectation of $x$ by $\mu$ such that $E(x)=\mu$
- **Variance**: The variance of a variable is the expectation of the squared deviations of the variable from the mean, denoted by $\sigma^2(x)=E[(x-\mu)^2]$
- **Standard Deviation**: The standard deviation of a variable $x$, $\sigma(x)$, is the square root of the variance of $x$.
- **Covariance**: Covariance tells us how linearly related are two variables. Given two variables $x$ and $y$ with respective expectations $\mu_x$ and $\mu_y$, the covariance is $\sigma(x,y)=E[(x-\mu_x)(y-\mu_y)]$.

# Elementary Statistics (sample)

- **Sample Means:** In statistical situations, we estimate the covariance from a sample of population using $\bar{x}$ and $\bar{y}$.
- **Sample Covariance:**

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

- **Correlation:** Correlation is a dimensionless measure of how two variables vary together. It is a "covariance" of two variables normalized by their respective spreads.

$$Cor(x,y) = \frac{Cov(x,y)}{\sigma(x)\sigma(y)}$$

# Probabilistic Model

- A complete probabilistic TS model can be specified as the **joint distribution function** of the sequence $x_1, x_2, \ldots, x_n$ of $n$ random variables as the probability that the values of the series are jointly less than $n$ constants $c_1, c_2, \ldots, c_n$.
  - $F(c_1, c_2, \ldots, c_n) = P(x_1 \leq c_1, x_2 \leq c_2, \ldots, x_n \leq c_n)$

- Although the joint distribution function describes the data completely, it is an unwieldy tool for analyzing TS data.

# Stationary Time Series

- A **strictly stationary** TS is one for which the probabilistic behavior of every collection of values $\{x_1, x_2, ..., x_n\}$ is identical to that of the time shifted set $\{x_{1+h}, x_{2+h}, ..., x_{n+h}\}$
  - $P(x_1 \leq c_1, x_2 \leq c_2, \ldots, x_n \leq c_n) = P(x_{1+h} \leq c_1, x_{2+h} \leq c_2, \ldots, x_{n+h} \leq c_n)$
  - for all time points n > 0, for all numbers $c_1, c_2, ..., c_n$, for all time shifts $h$.

- If a TS is strictly stationary, then all of the distribution functions for subsets of variables must agree with their counterparts in the shifted set for all values of the shift parameter $h$.
- In other words, shifting the time axis does not affect the distribution.

# Weakly Stationary Time Series

- The lack of independence between two subsequent values $x_t$ and $x_{t+h}$ can be assessed numerically using **covariance** and **correlation**.
- A **weakly stationary** TS $x_t$ is one for which
  - the mean $\mu_t = \mu$ is independent of $t$, i.e., it is constant.
  - the autocovariance $\gamma(x_t, x_{t+h})$ is independent of $t$ for each $h$.
- We typically use the term stationary to refer to weakly stationary.

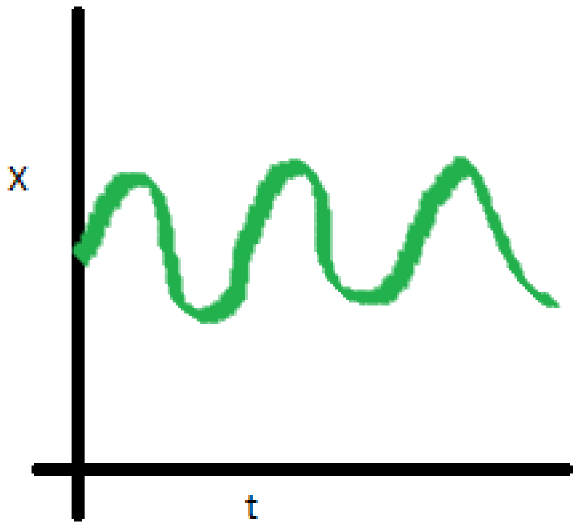
- In short, a TS with a certain trend or with a certain seasonality is not stationary.

# Stationary Criteria

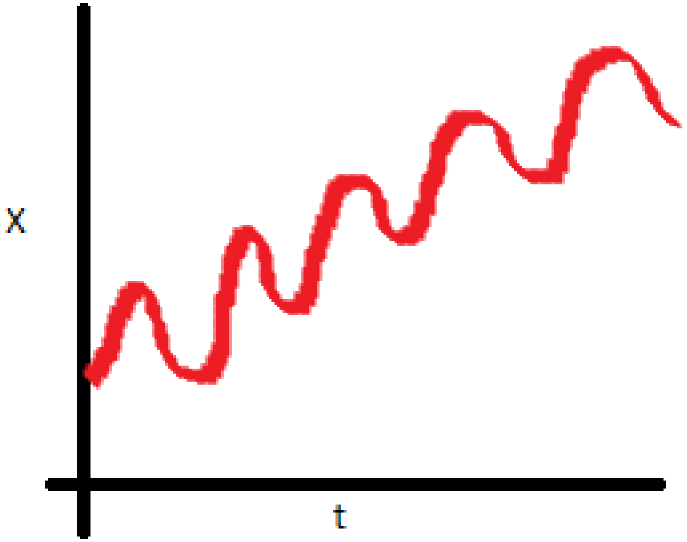
- There are three basic criterion for a TS to be stationary:
- The **mean of the series should not be a function of time** but should be a constant. We consider a time series as stationary in the mean if $\mu(t)$=$\mu$, a constant.
- The variance of the series should not be a function of time. This property is called: homoscedasticity. Hence, a time series is stationary in the variance if $\sigma^2(t)$=$\sigma^2$, a constant.
- The covariance of $x_t$ and $x_{t+h}$ should not be a function of time.
- In short, a weakly stationary time series has statistical properties that are stable over time.

# Stationary Time Series

- The mean of the series should not be a function of time, rather should be a constant.



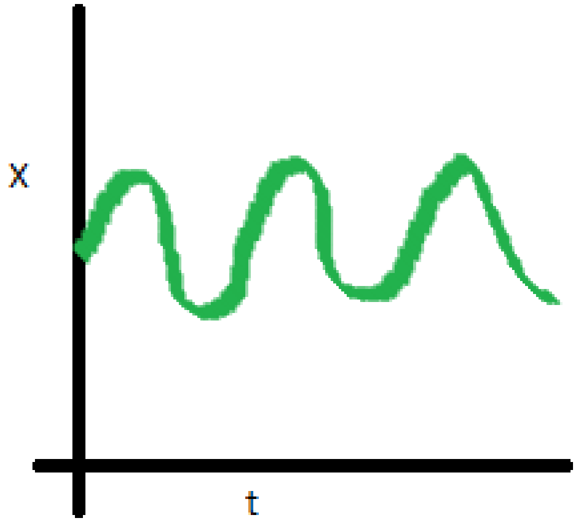

Stationary series



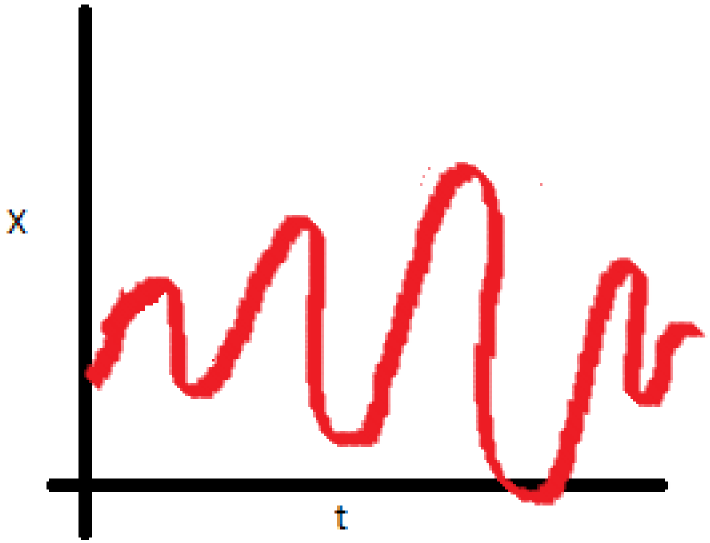

Non-Stationary series

# Stationary Time Series

- The variance of the series should not a be a function of time, rather should be a constant.



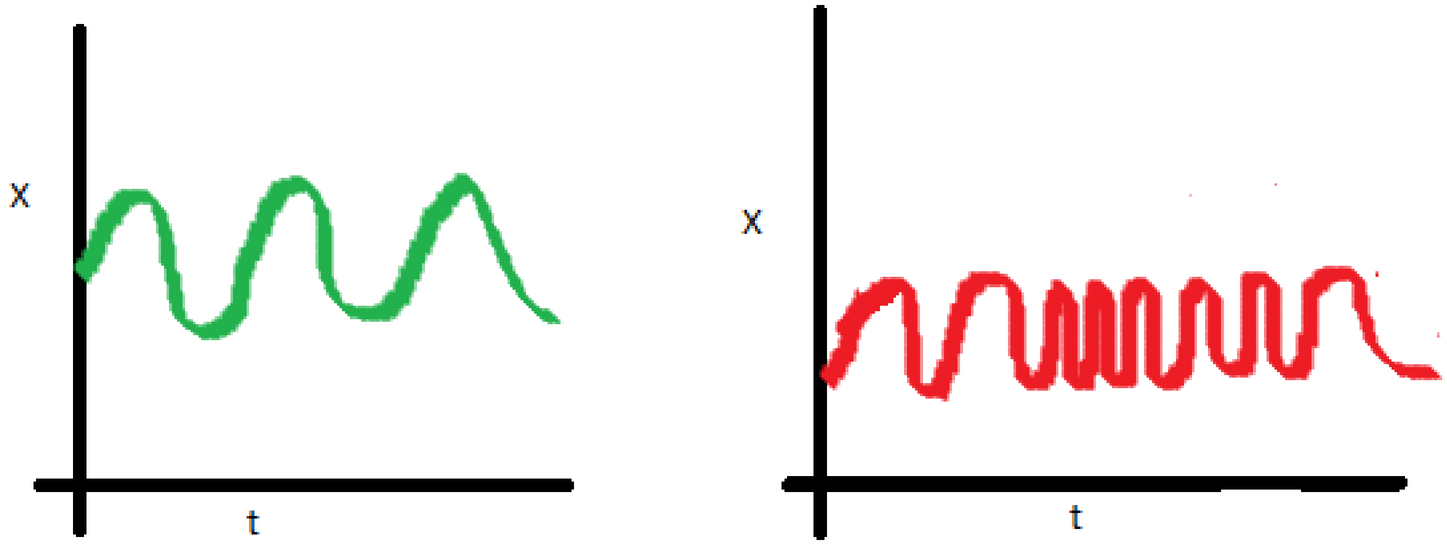

Stationary series




Non-Stationary series

# Stationary Time Series

- The covariance of the $t_{th}$ point and the $(t+h)_{th}$ point should not be a function of time.




Stationary series

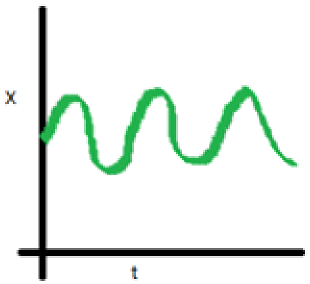
Non-Stationary series

# Make Stationarity a Non-Stationary Time Series

- There are various techniques to bring stationarity to a non-stationary time series.
- **Detrending**: removing the trend component from the TS by using
  - Log
  - Removing the smoothed mean
  - Linear regression
- **Differencing**: consider as TS the differences of the values instead of actual values
- **Decomposition**: remove all the schematic components and consider the residuals

# Stationarity Test

- The Augmented Dickey-Fuller (ADF) test of stationarity results comprise of an ADF statistic and some Critical Values for different confidence levels.
- If the **ADF statistic is less than the Critical Value**, we can reject the null hypothesis and say that the series is stationary and does not have any time-dependent structure.
- ADF null hypothesis: a unit root is present in an autoregressive model.
- A unit root is a feature of some stochastic processes (such as random walks) that can cause problems in statistical inference involving TS models.



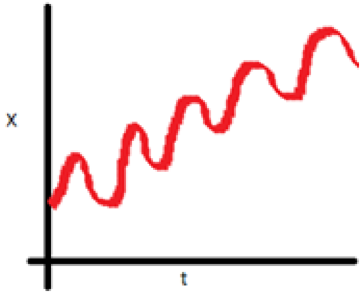

Stationary series

ADF Statistic: -4.808291
p-value: 0.000052
Critical Values:
5%: -2.870
1%: -3.449
10%: -2.571

Test NOT Passed
Null Hypotesis rejected

ADF Statistic: 0.815369
p-value: 0.991880
Critical Values:
5%: -2.884
1%: -3.482
10%: -2.579

Test
Passed




Non-Stationary series

# Autocorrelation

# The Sample Autocorrelation Function

• In practical problems, we do not start with a model, but with observed data.

• To assess the degree of dependence in the data and to select a model, one of the important tools we use is the sample Autocorrelation Function (Sample ACF).

• Let $x_1, x_2, \ldots, x_n$ be the observations of a time series.

• Sample $mean\,\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$

• Sample Autocovariance $\gamma(h) = \frac{1}{n}\sum_{t=1}^{n-h}\left(x_{t+|h|} - \bar{x}\right)(x_t - \bar{x})$

• It measures the linear dependence between lagged TS starting at two different time points on the same TS.

• Very smooth TS exhibit autocovariance that stay large even when h is large, whereas choppy TS tend to have autocovariance that are nearly zero for large separations. If $\gamma(h) = 0$ they are not linearly related

• Notice that if a TS is stationary autocovariance is not a function of time.

# The Sample Autocorrelation Function

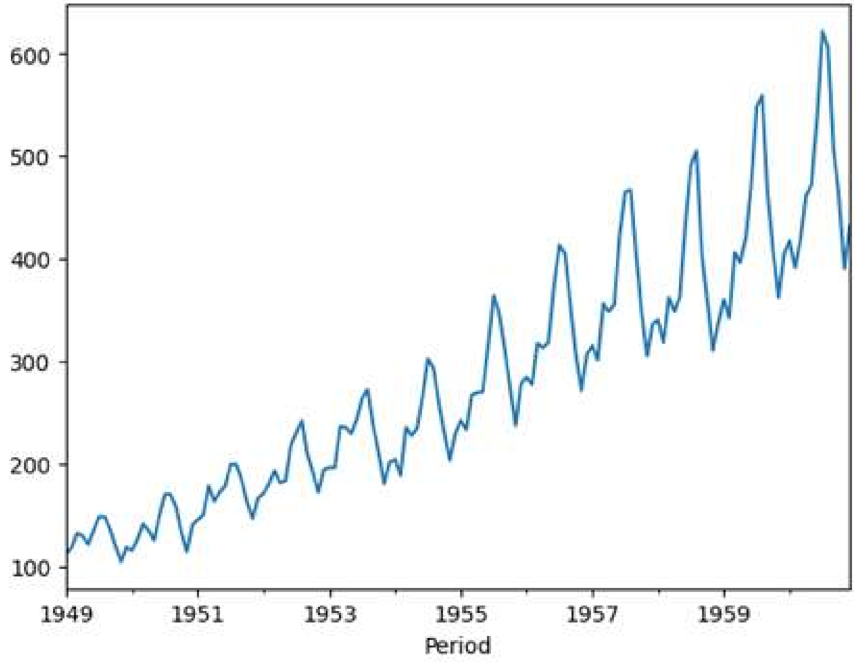
- Sample Mean: $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$
- Sample Autocovariance at lag h: $\gamma(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}\left(x_{t+|h|} - \bar{x}\right)(x_t - \bar{x})$
- At lag 0, the autocovariance equals the sample variance: $\gamma(0) = \frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2$
- Sample Autocorrelation Function (ACF): $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$, $\rho(h) \in [-1,1]$
- At lag 0, the autocorrelation is always equal to 1:
- The ACF measures the linear relationship between a time series and its lagged values.
- Each lag has its own autocorrelation coefficient: $h = 0,1,2,3, \ldots$

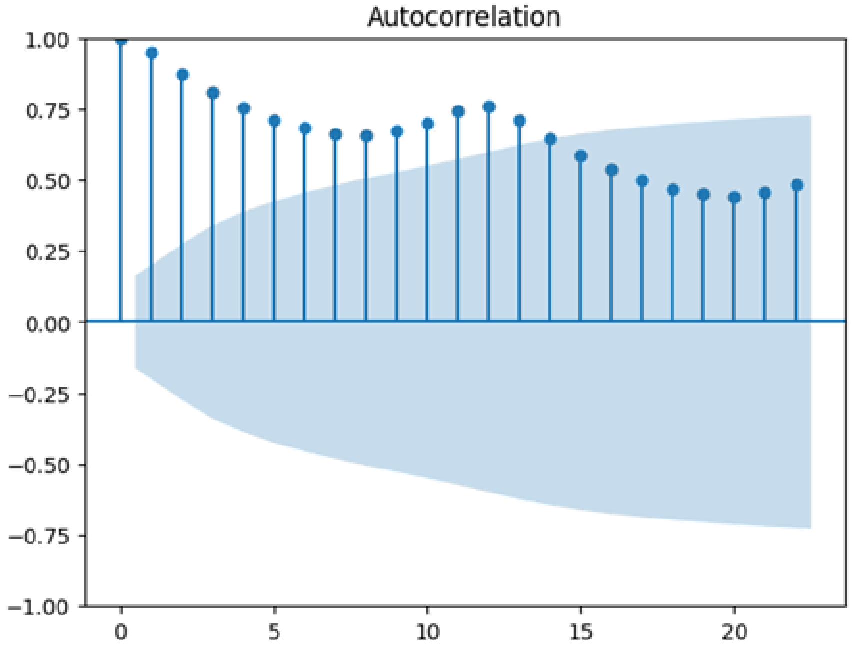
# Autocorrelation - Remarks

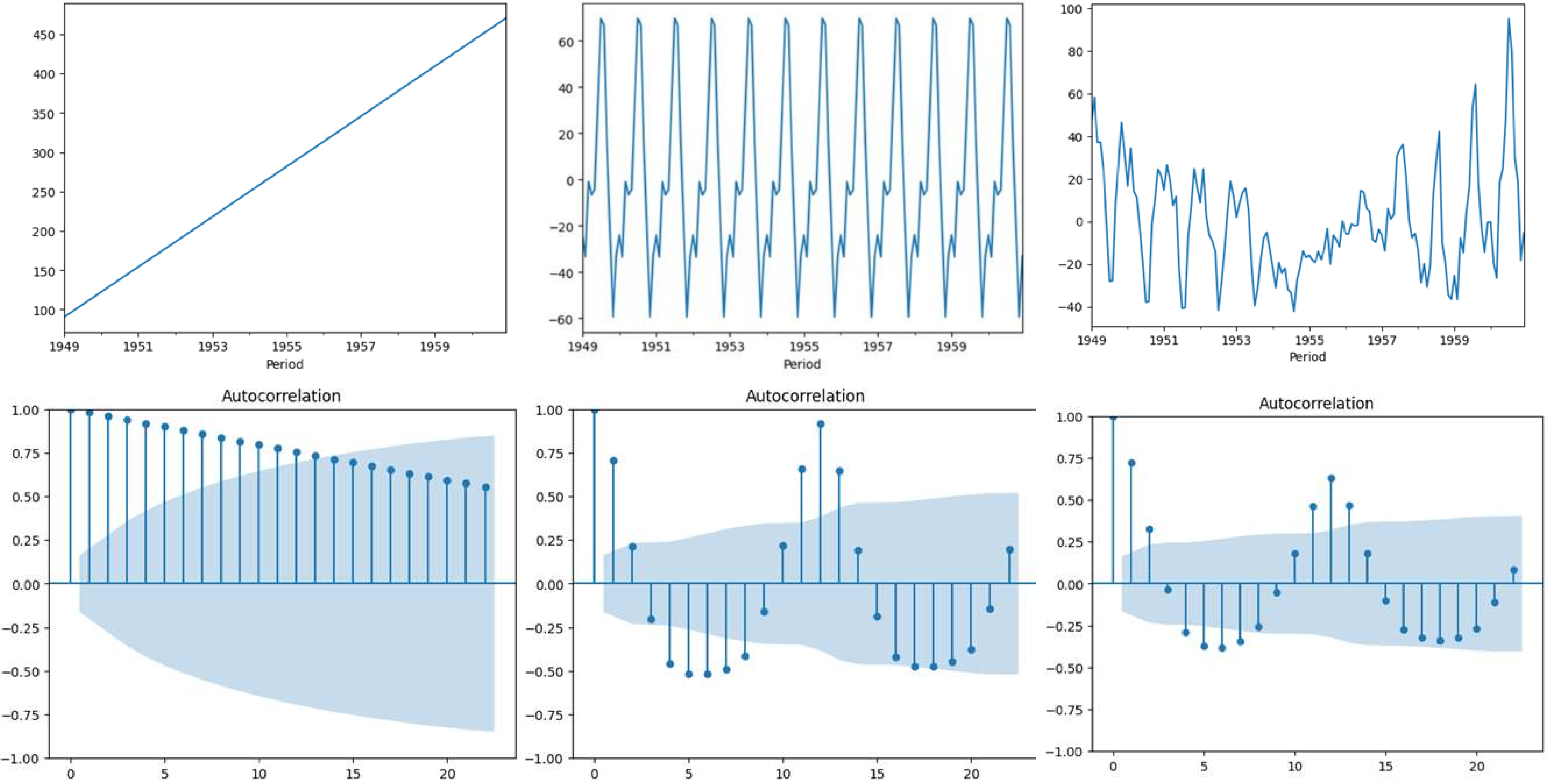
- Autocorrelations or lagged correlations are used to know whether a TS is dependent on its past, i.e., how sequential observations in a TS affect each other.
- ACF can be computed for any data set and is not restricted to stationary time series.
- For data containing a trend ACF will display slow decay as $h$ increases.
- For data containing a deterministic periodic component ACF will exhibit similar behavior with the same periodicity.

# ACF Plot

- An ACF plot is a correlogramn, i.e., a plot of the autocorrelation functions for sequential values of lag $h = 0, 1, ..., n$
- It shows the correlation structure in each lag.
- Confidence intervals are drawn as a cone.

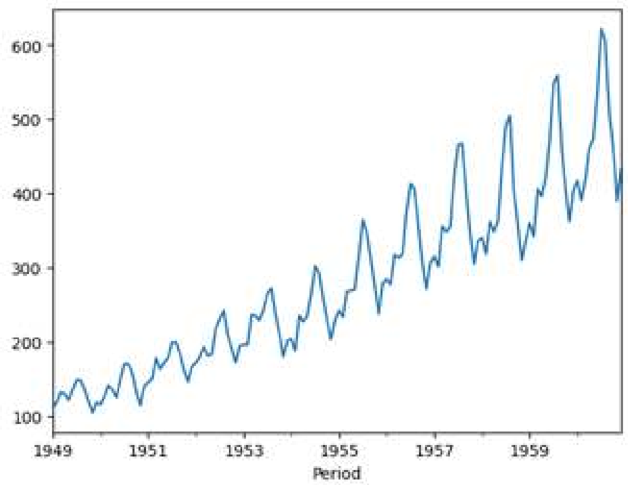


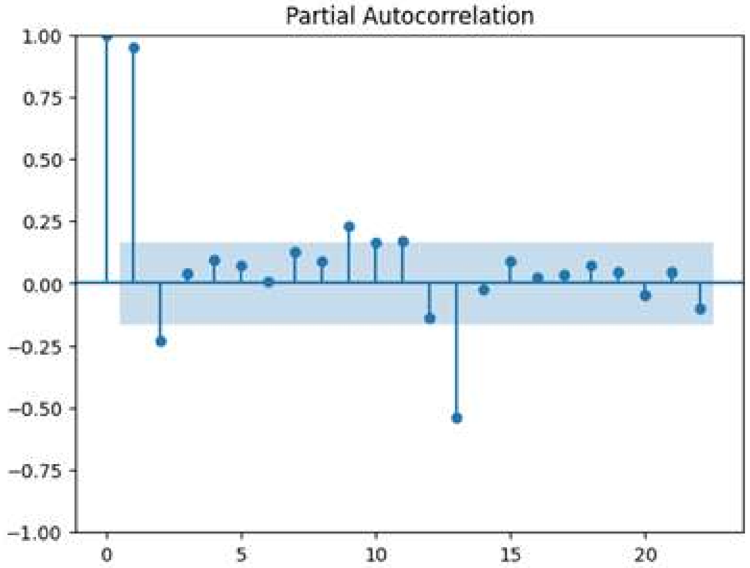



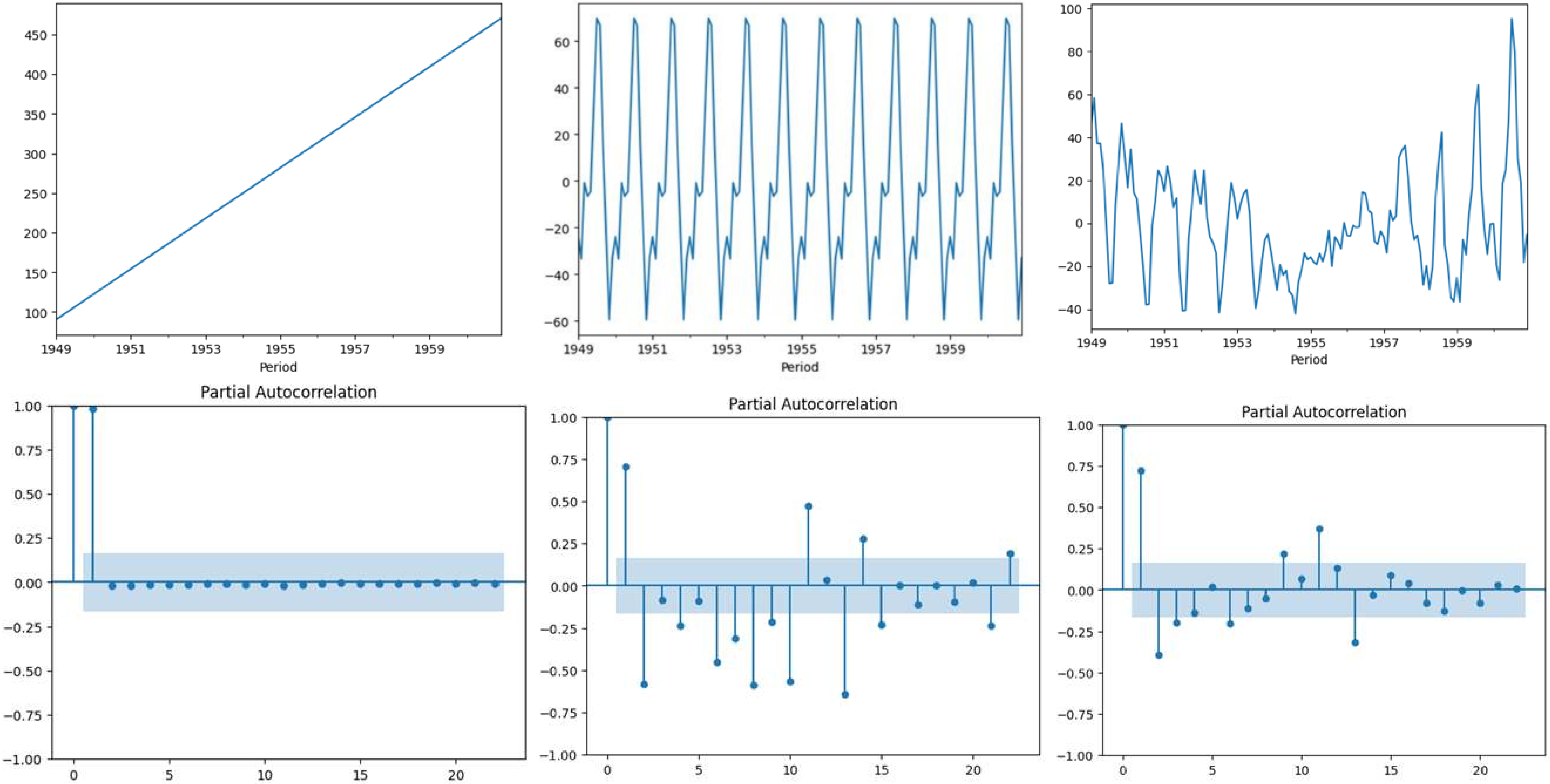

Period
Period
Period
Autocorrelation
Autocorrelation
Autocorrelation

# PACF plot

- Partial autocorrelation measures the direct relationship between a time series value and an earlier value, after removing the effects of the values in between.
- For example, the partial autocorrelation at lag $h$ measures the relationship between $x_t$ and $x_{t-h}$ after removing the effects of $x_{t-1}, x_{t-2}, \ldots, x_{t-h+1}$
- Ordinary autocorrelation includes both direct and indirect relationships. For example, $x_t$ may be related to $x_{t-3}$ because $x_{t-3}$ affects $x_{t-2}$, which affects $x_{t-1}$, which affects $x_t$.
- So the autocorrelation at lag 3 includes this indirect chain: $x_{t-3} \rightarrow x_{t-2} \rightarrow x_{t-1} \rightarrow x_t$
- Partial autocorrelation removes these indirect effects. It asks: After we already account for the closer lags, does $x_{t-h}$ still have a relationship with $x_t$?



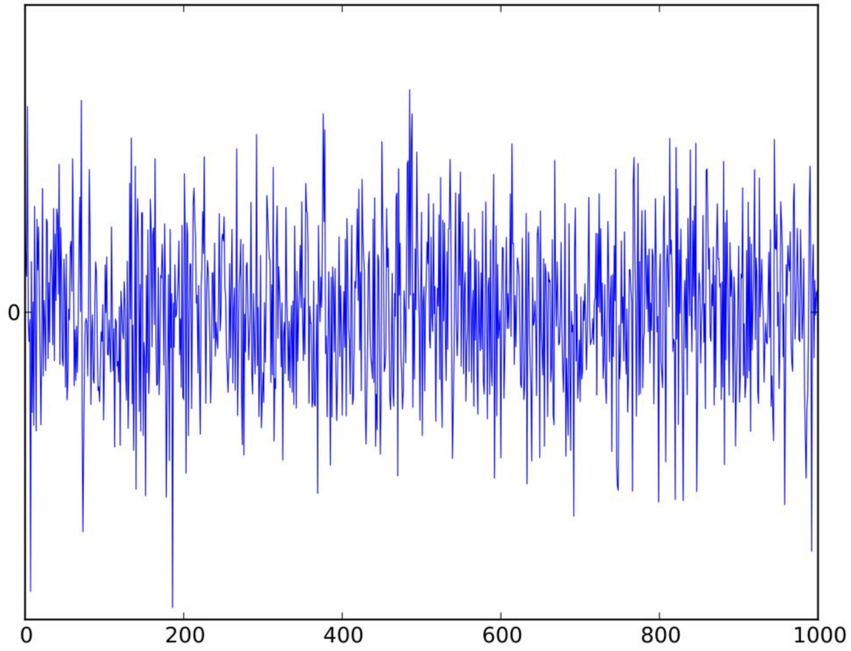

Period
Period
Period
Partial Autocorrelation
Partial Autocorrelation
Partial Autocorrelation

# ACF and PACF Summary

- **Autocorrelation Function (ACF)**: It is a measure of the correlation between the TS with a lagged version of itself.
  - For instance, at lag *k=5*, ACF would compare TS at time instant $t_1 \dots t_n$ with TS at instant $t_1$-5, …, $t_2$-5 ($t_1$-5 and $t_n$ being end points).
- **Partial Autocorrelation Function (PACF):** This measures the correlation between the TS with a lagged version of itself but after eliminating the variations already explained by the intervening comparisons.
  - For instance, at lag *k=5*, would compare the correlation but remove the effects already explained by lags 1 to 4.

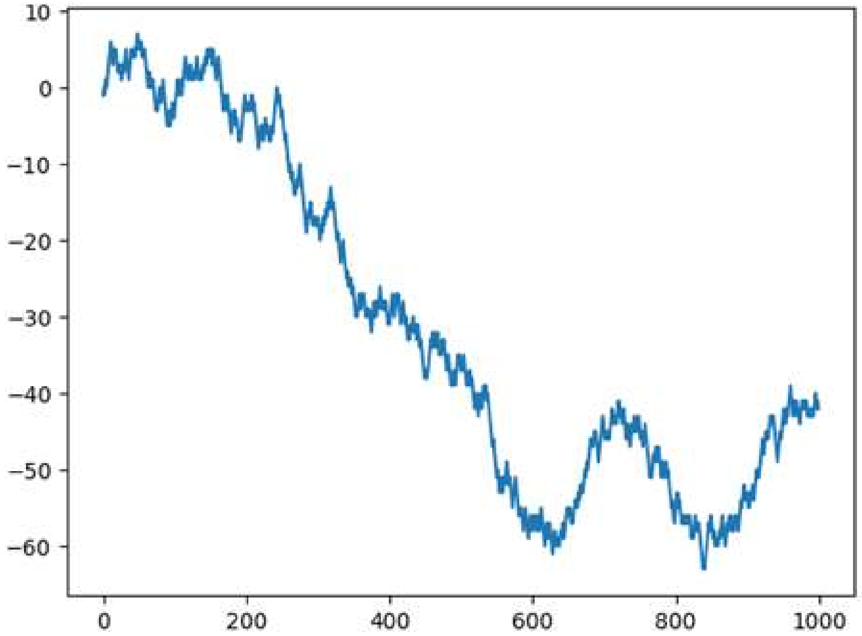
# White Noise

- White noise is a discrete signal whose samples are regarded as a sequence of serially uncorrelated random variables with a mean of zero and a finite variance
- In some contexts, it is also required that the samples be independent and have identical probability distribution (IID)
- if each sample has a normal distribution with zero mean, the signal is said to be additive white Gaussian noise.



# Random Walk

- Let $\{x_t\}$ with $t = 0, 1, 2, \ldots$ be a time series then $\{x_t\}$ is a **random walk** if $x_t = x_{t-1} + w_t$ where $w_t$ is a white noise series.
- Substituting $x_{t-1} = x_{t-2} + w_{t-1}$ in $x_t = x_{t-1} + w_t$ and for others $t-i$
- We get $x_t = w_t + w_{t-1} + w_{t-2} + w_{t-3} + \ldots$
- If the series is not infinite, we have $x_t = w_1 + w_2 + w_3 + \ldots + w_t$
- where each $w_i$ is IID noise.
- Random walk TS plays an important role as a building block for complex TS models

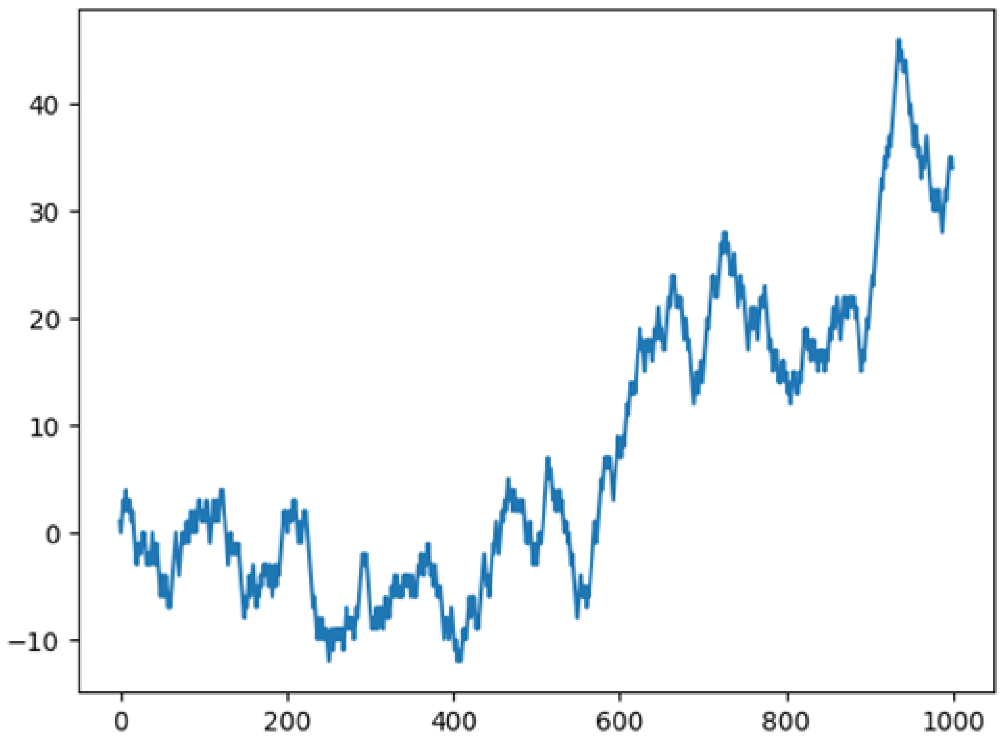


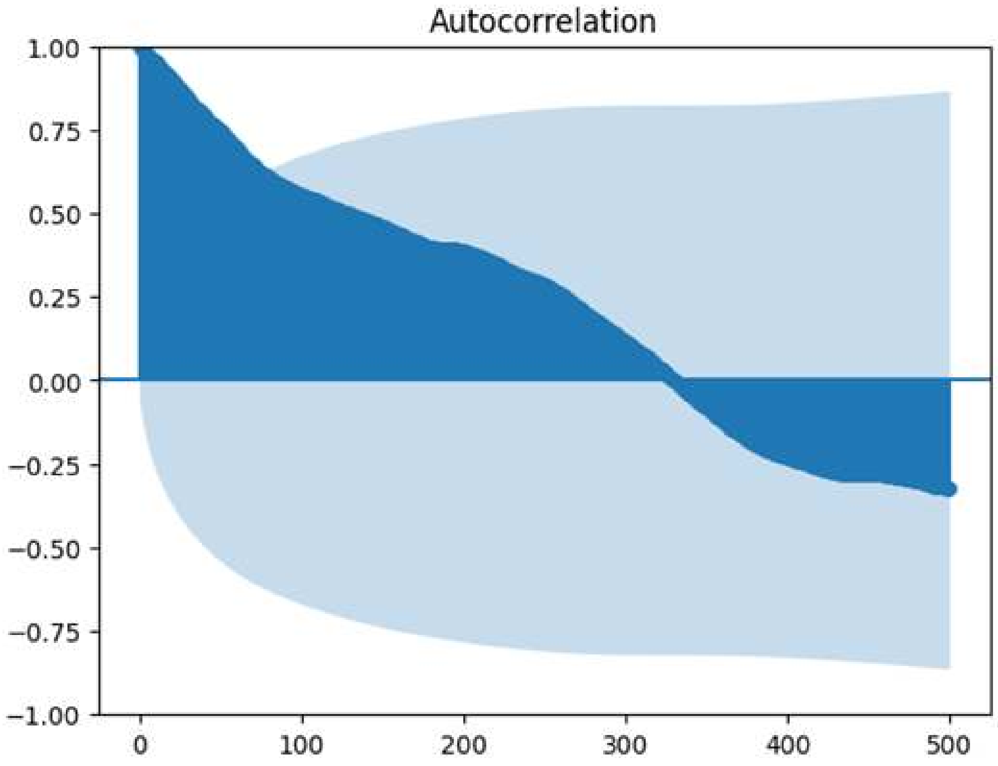
# Random Walk

- Random walk models are widely used for non-stationary data, e.g. financial and economic data.
- Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction.
- The forecasts for a random walk model are equal to the last observation, as future movements are unpredictable, and are equally likely to be up or down.

# Random Walk and Autocorrelation

- Given the way that the random walk is constructed, we expect a strong autocorrelation with the previous observation and a linear fall off from there with previous lag values.



# Random Walk and Stationarity

- Given the way that the random walk is constructed and the results of reviewing the autocorrelation, we know that the observations in a random walk are dependent on time.
- The current observation is a random step from the previous observation.
- Therefore, a random walk is a non-stationary TS.
- In fact, **all random walk processes are non-stationary**. Note that not all non-stationary time series are random walks.

ADF Statistic: -0.373987
p-value: 0.914352
Critical Values:
1%: -3.437
5%: -2.864
10%: -2.568

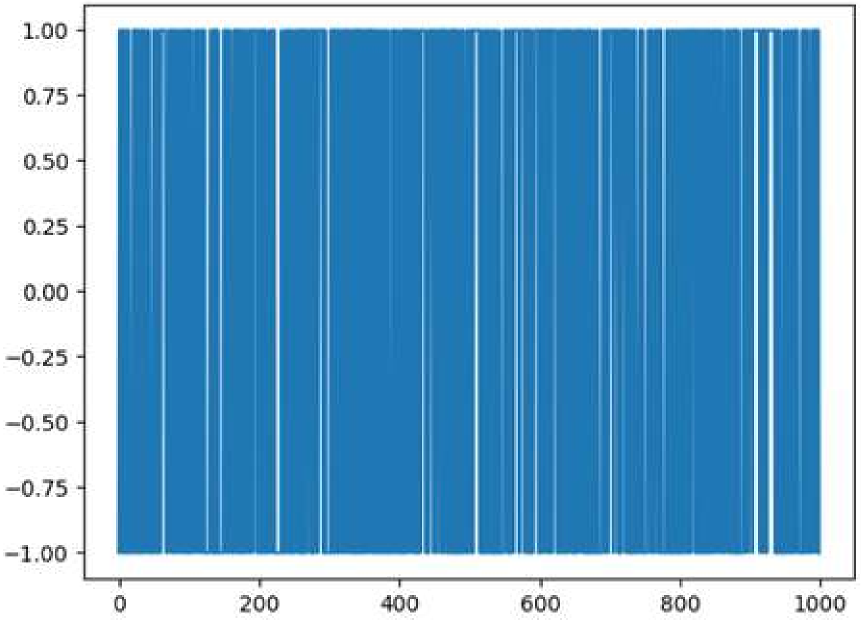
Test Passed

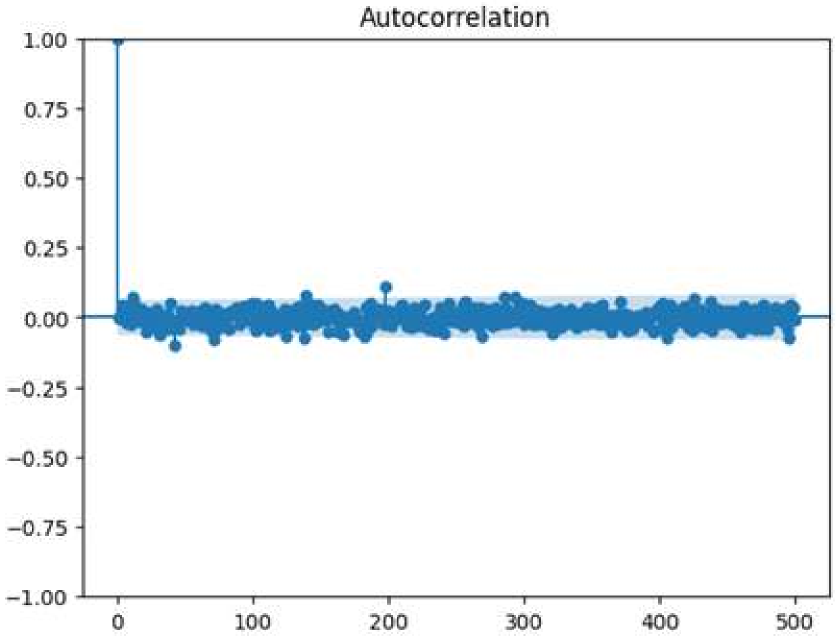
# White Noise and Autocorrelation

- Time series that show no autocorrelation are called **white noise**.
- In other words, a white noise is made of random values with a given mean and standard deviation but not autocorrelation.
- When the differenced series is white noise, i.e. $\varepsilon_t = x_t - x_{t-1}$, where $\varepsilon_t$ denotes white noise, than $x_t = x_{t-1} + \varepsilon_t$ is a **random walk model**

# White Noise Autocorrelation and Stationarity

- We can obtain white noise by **differencing a random walk**.
- All correlations are small, close to zero and below the 95% and 99% confidence levels
- The time series is stationary.




Test NOT Passed
Null Hypotesis rejected

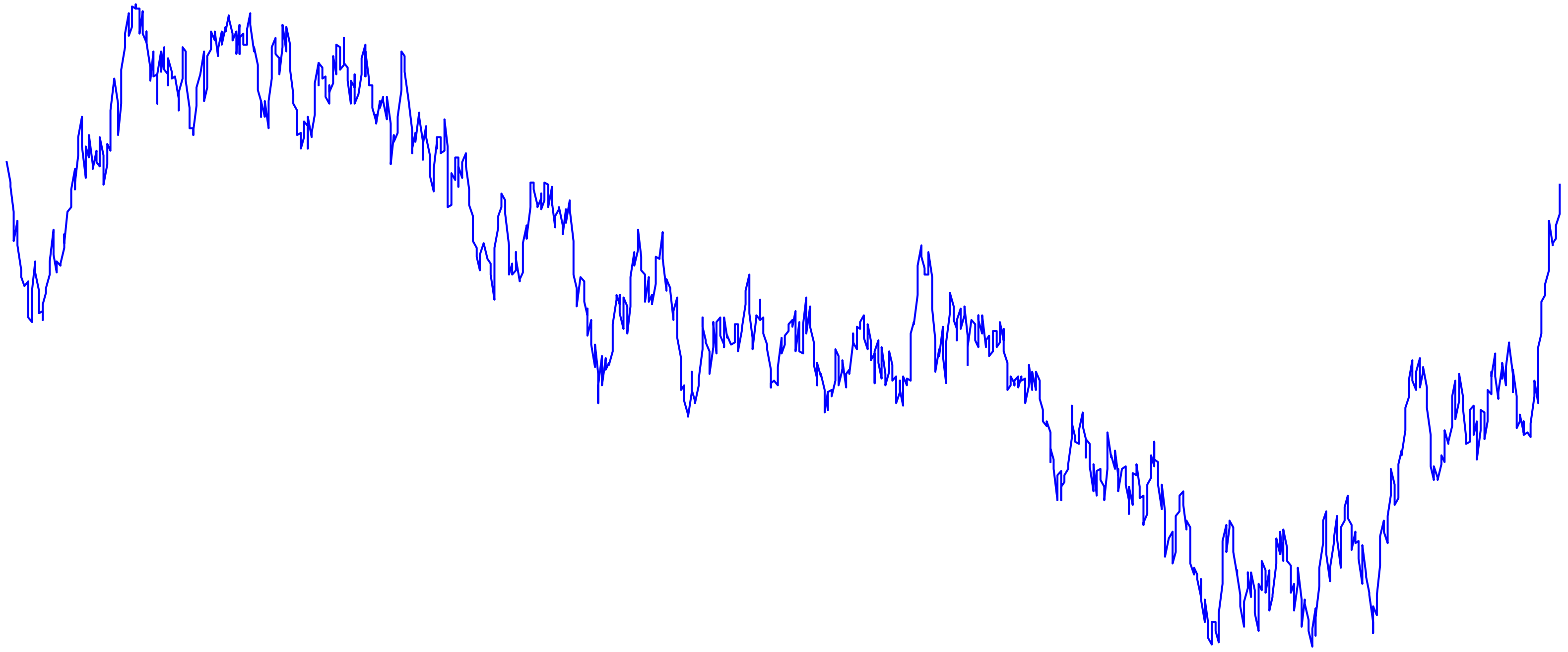
ADF Statistic: -31.659617
p-value: 0.000000
Critical Values:
1%: -3.437
5%: -2.864
10%: -2.568

# Time Series Preprocessing for Forecasting

1. Plot the Time Series to examine the main characteristics (trend, seasonality, ...)
2. Remove the trend and seasonal components to get stationary residuals/models
3. Choose a model to fit the residuals using sample statistics (sample autocorrelation function)
4. Forecasting will be given by forecasting the residuals to arrive at forecasts of the original series $X_t$

# Time Series Preprocessing Remarks

- Depending on the TSA task different normalizations might be required but not necessarily all of them!!!



# References


- Forecasting: Principles and Practic. Rob J Hyndman and George Athanasaopoulus. (https://otexts.com/fpp2/)
- Time Series Analysis and Its Applications. Robert H. Shumway and David S. Stoffer. 4th edition.(http://www.stat.ucla.edu/~frederic/415/S23/tsa4.pdf)
- Time Series Analysis in R (https://s-ai-f.github.io/Time-Series/)
- Mining Time Series Data. Chotirat Ann Ratanamahatana et al. 2010. (https://www.researchgate.net/publication/227001229_Mining_Time_Series_Data)