The Statistics of Time Series

Martin Burger
STATS PROGRAMMING TUTOR

www.r-tutorials.com

Time Series Statistics

Data visualization techniques with Matplotlib

Unique statistics of time series data

Stationarity: Constant statistical properties of the time series

- Augmented Dickey-Fuller test

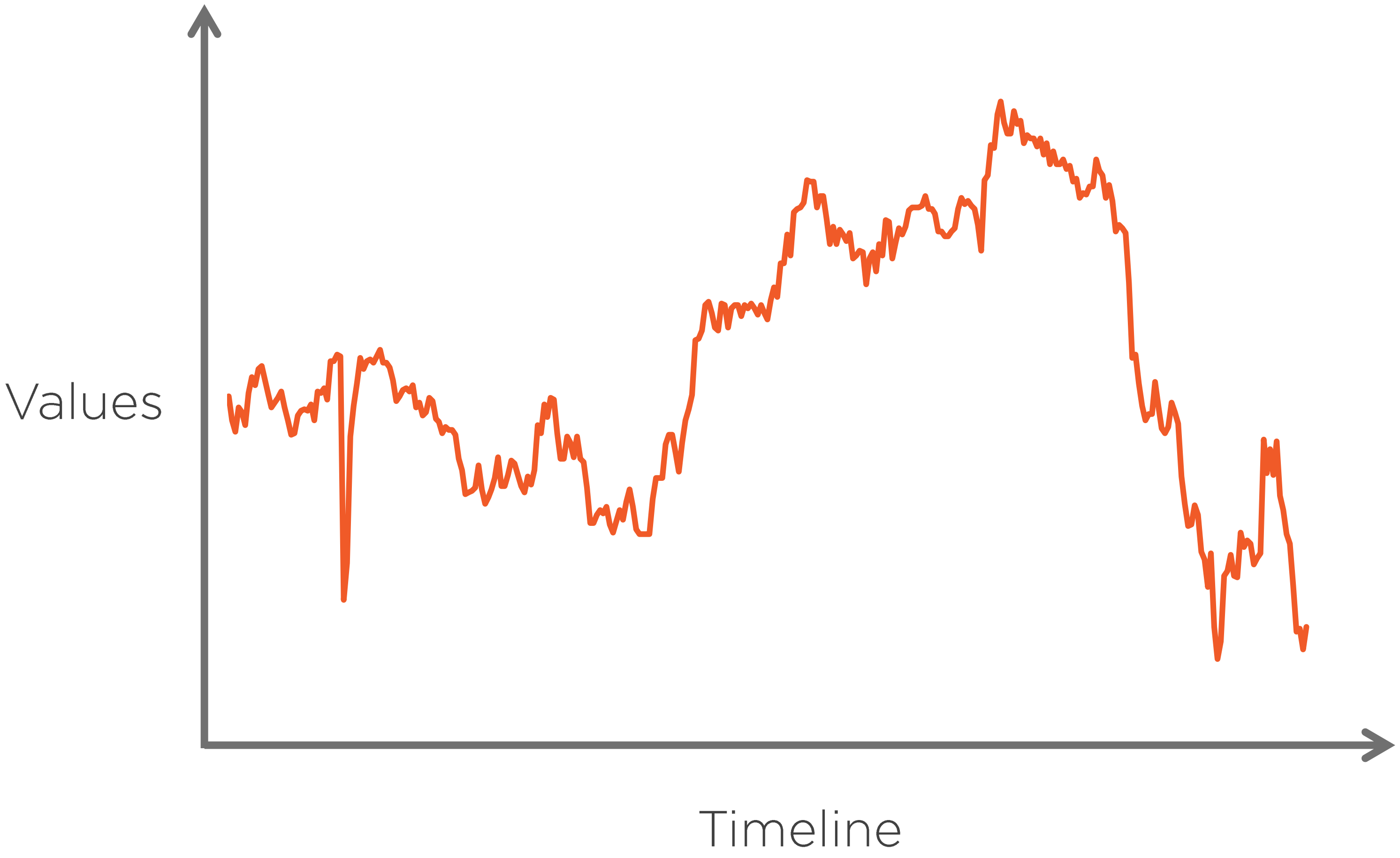
Autocorrelation: Correlation between the observations of a time series variable

- ACF and PACF plots

Moving averages

Visualizing Time Series with Matplotlib

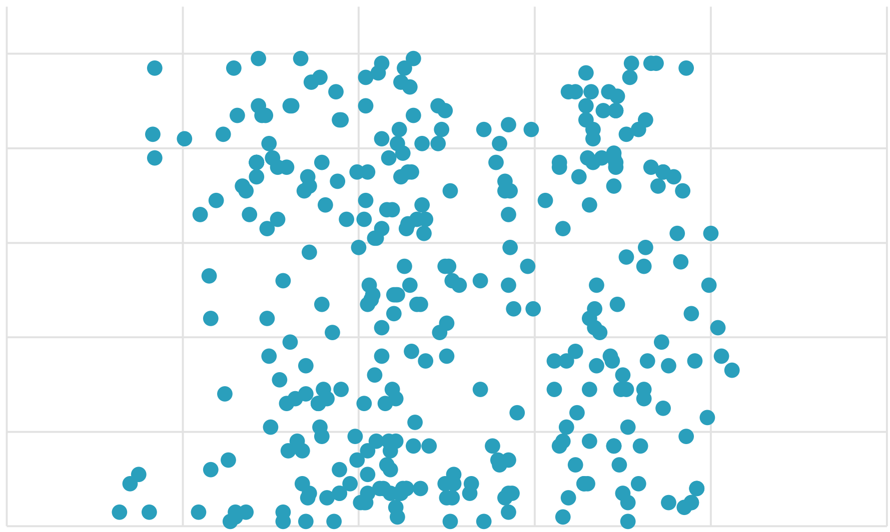
Time Series Visualization via Line Plot



The Significance of the Time Component

Time Series

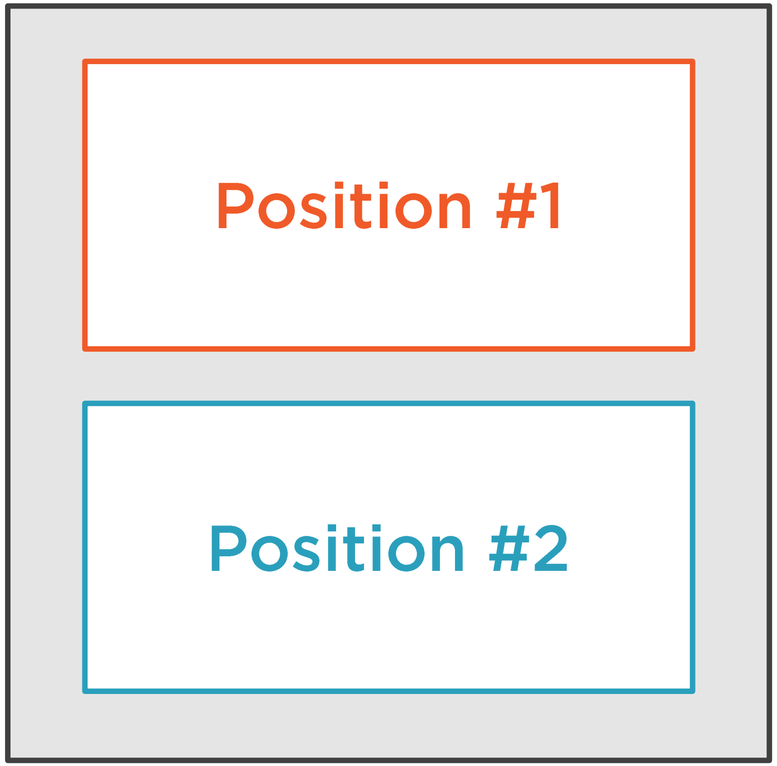
Data points indexed by a time stamp which results in a successive order



Unordered Data

The time point of an observation doesn't have significance

Positions in a Matplotlib Subplot Figure



Building a visualization matrix with `plt.subplot()`

Specifying the main structure and the actual position by integers

Example setup: 2, 1, 1

- First integer: Number of rows
- Second integer: Number of columns
- Third integer: Actual position

Stationarity in Time Series

Ensuring Consistent Statistical Properties

Variance

Mean

Autocorrelation

Transformation

Differencing

Trend and Non-stationarity

The mean changes over time as a result of trend

- Underestimated predictions

Ensuring trend-stationarity by taking the trend component out of the time series

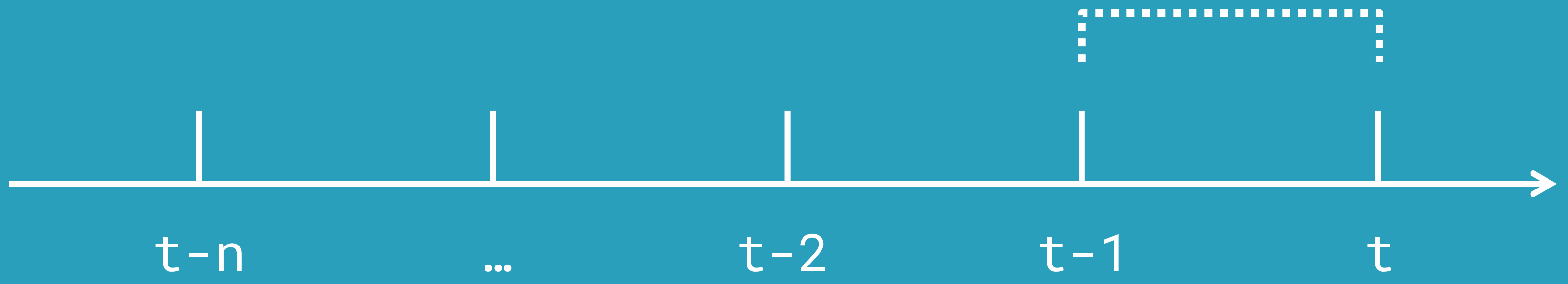
Ensuring difference-stationarity by differencing the dataset

Unit root tests for stationarity:

- If $Y_t - Y_{t-1}$ is stationary and random: Y is a random walk
- If $Y_t - Y_{t-1}$ is stationary but not random: Refined model is required

The First Order of Differencing (Y')

$$Y'_t = Y_t - Y_{t-1}$$



Time series Y

Visual Indications of Non-stationarity

The data has a clear trend and/or seasonality

Changes in variance and/or mean

Unit Root Tests

Statistical tests for non-stationarity

Augmented Dickey-Fuller test

- Removes autocorrelation and tests for non-stationarity
- Equal mean and variance throughout the time series

Null hypothesis: Non-stationarity

- Stationarity: P-value < 0.05

Complement the test statistic with data visualizations for a deeper understanding.

Autocorrelation

Autocorrelation

Describes the correlation between the values of an ordered series at different time points.

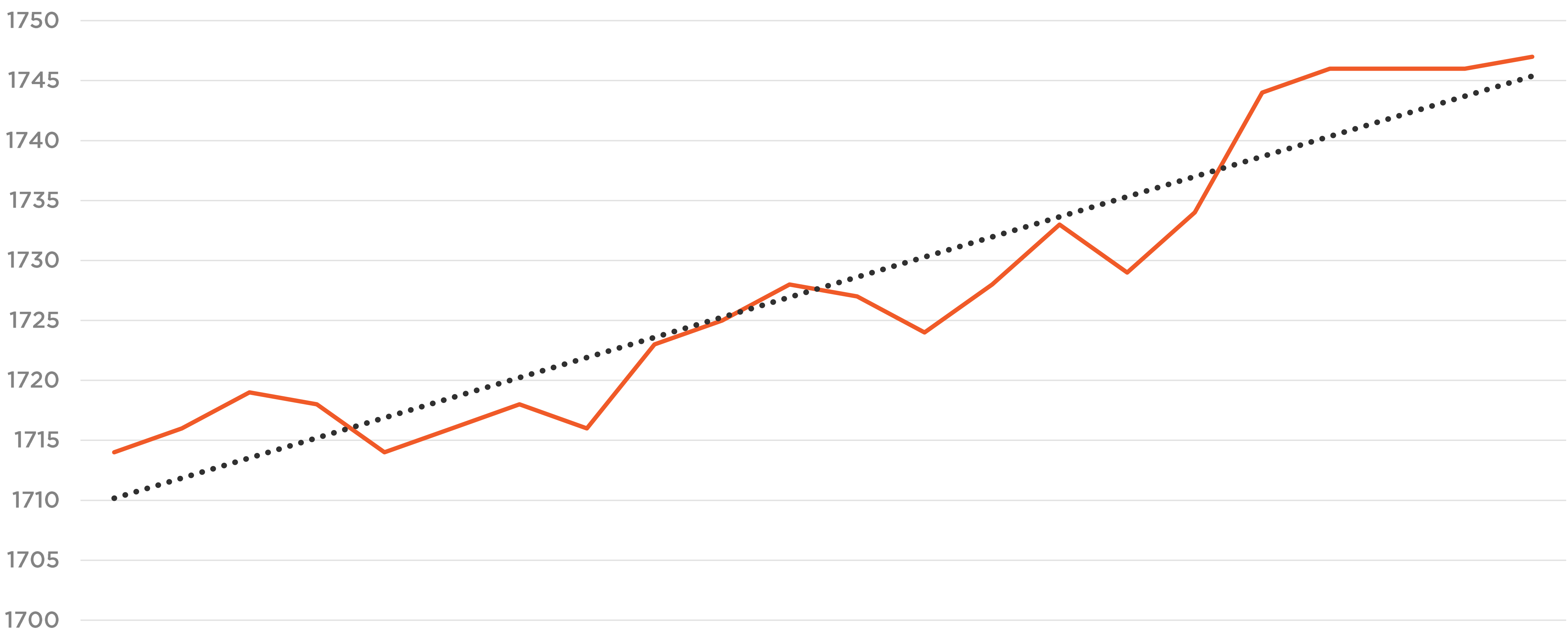
Autocorrelation in Time Series

The influence of previous observations on the recent one

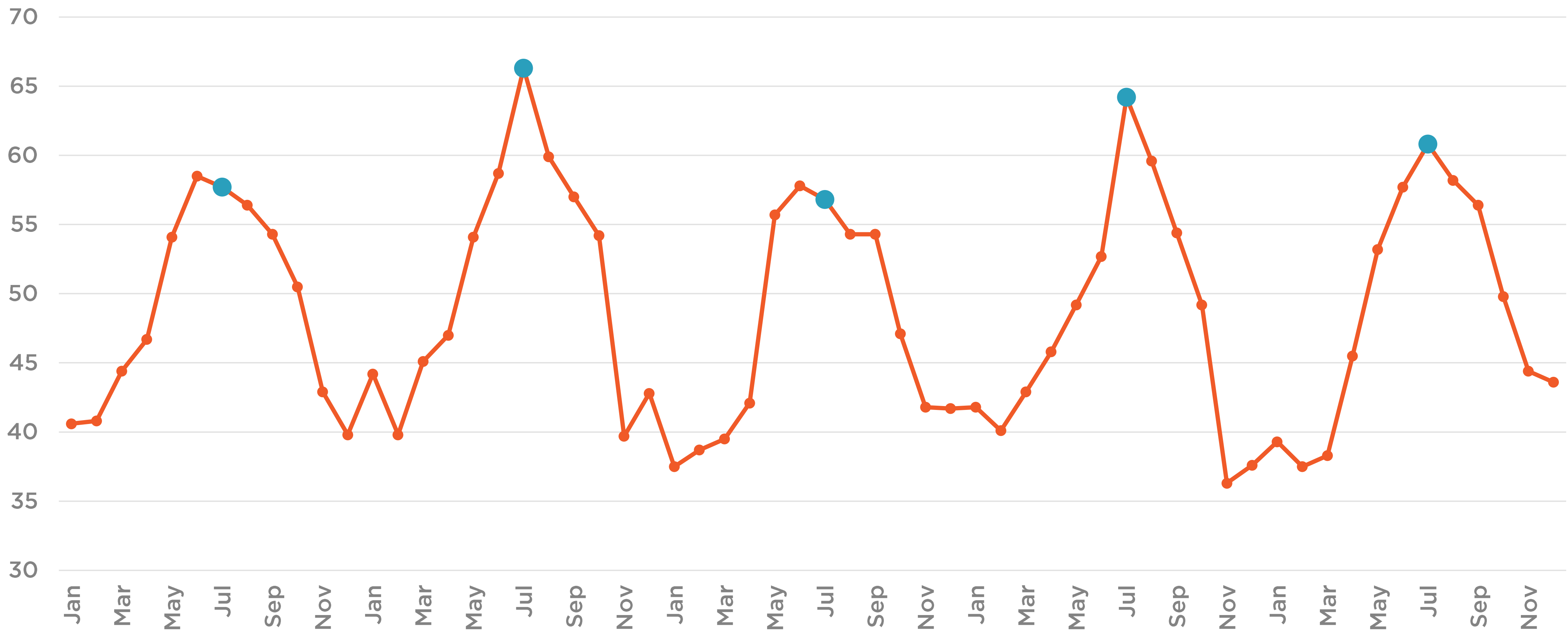
A step on the time scale is called a lag

Trend and seasonality are visual indicators of autocorrelation

Trend as an Indicator of Autocorrelation



Seasonality as an Indicator of Autocorrelation



What Does the Story Tell?

Autocorrelation as a consequence of the data story

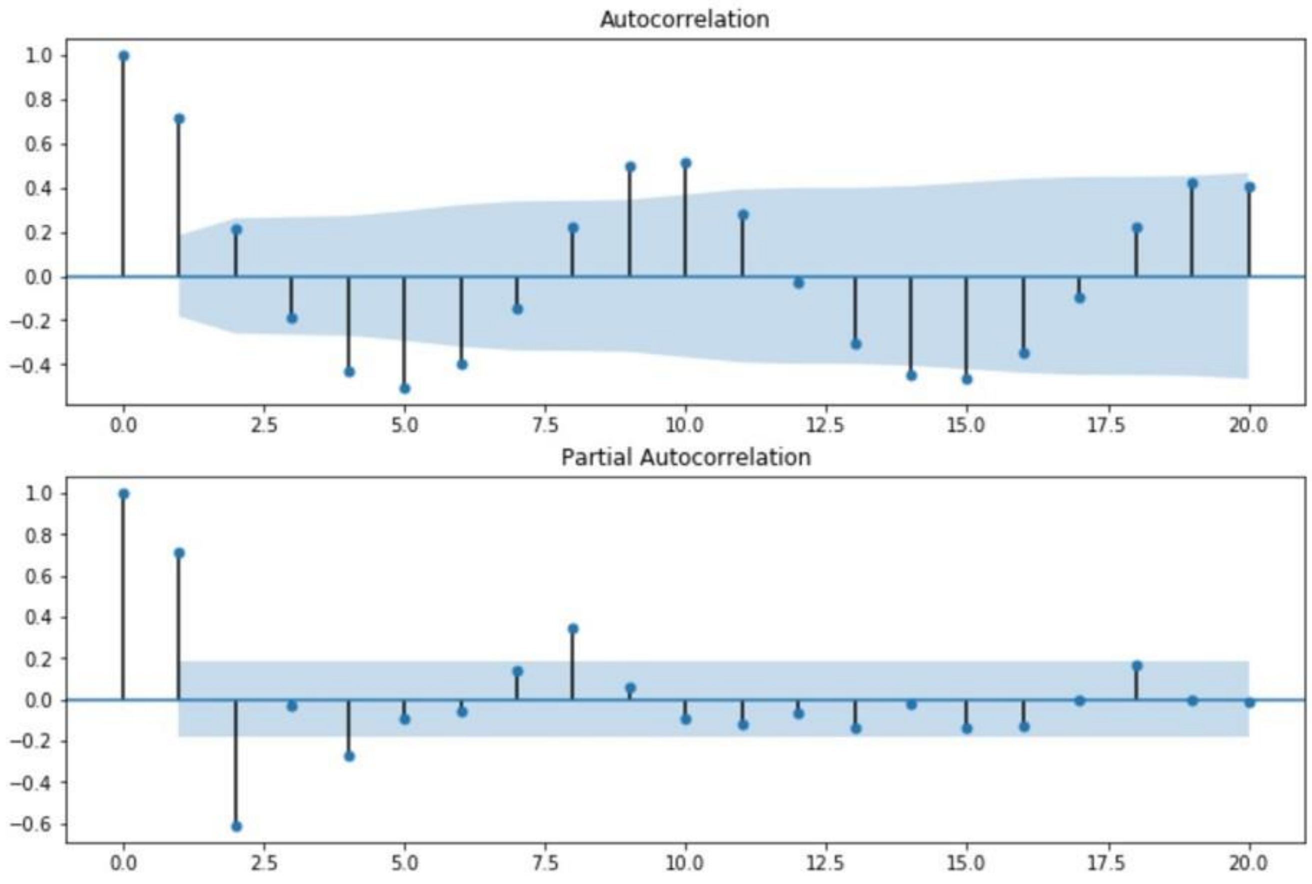
Past elimination of members influences the future state of the population

ACF and PACF plots

Visual representation of autocorrelation

The relevant lags with autocorrelation

PACF is adjusted for all earlier lags



Determining the Number of Lags

Non-seasonal

Twenty lags work fine for 95% confidence (1/20)

Seasonal

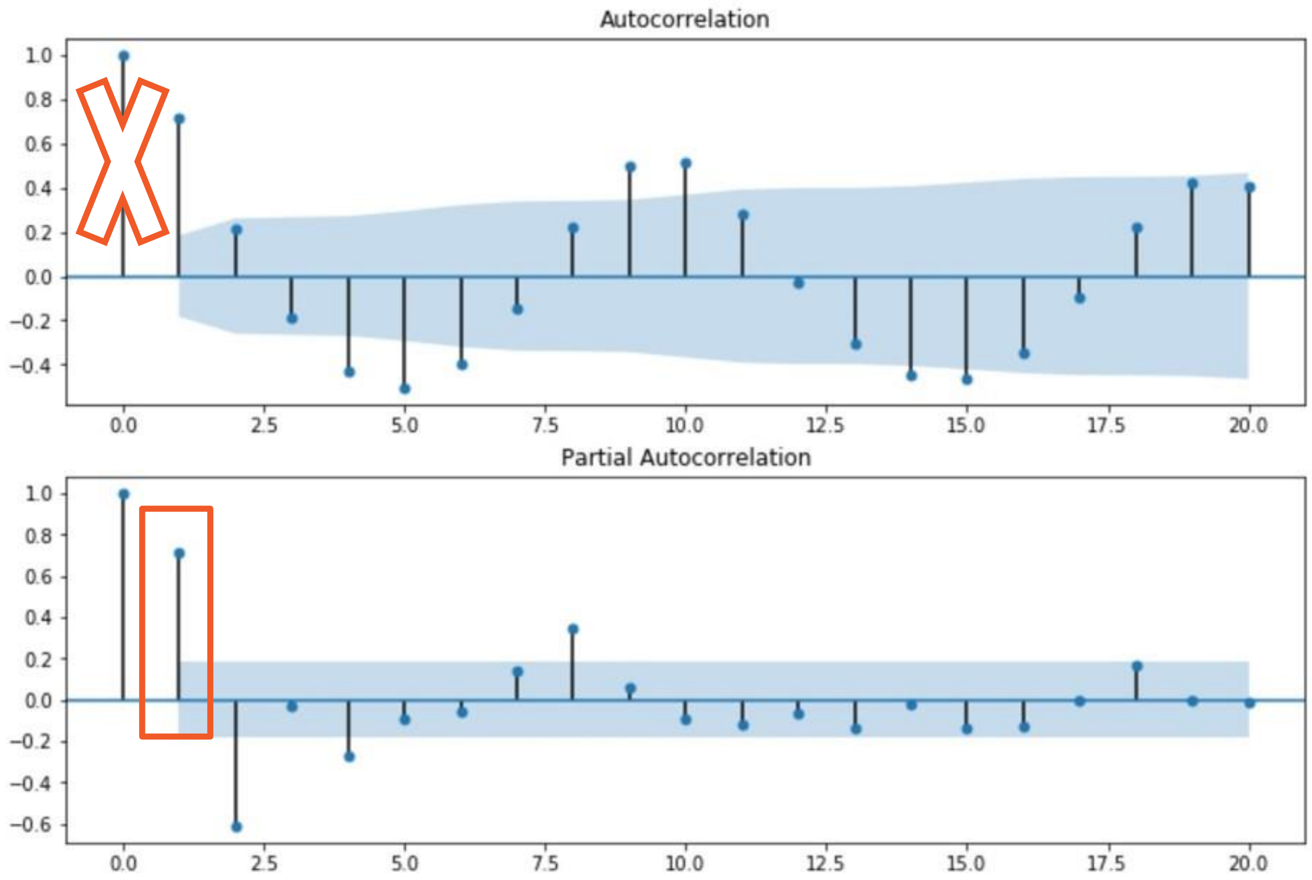
At least three seasonal cycles with extra buffer for 95% confidence (2/40)

95% confidence boundaries

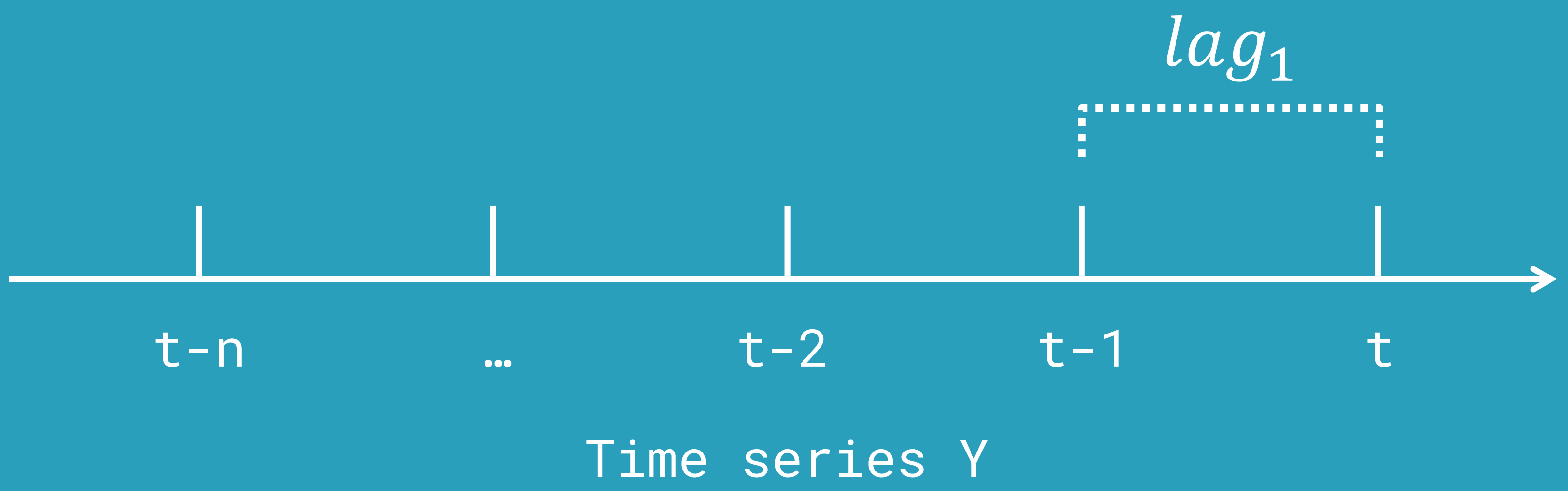
ACF plot shows autocorrelation

PACF is significant at lags 1, 2, 4 and 8

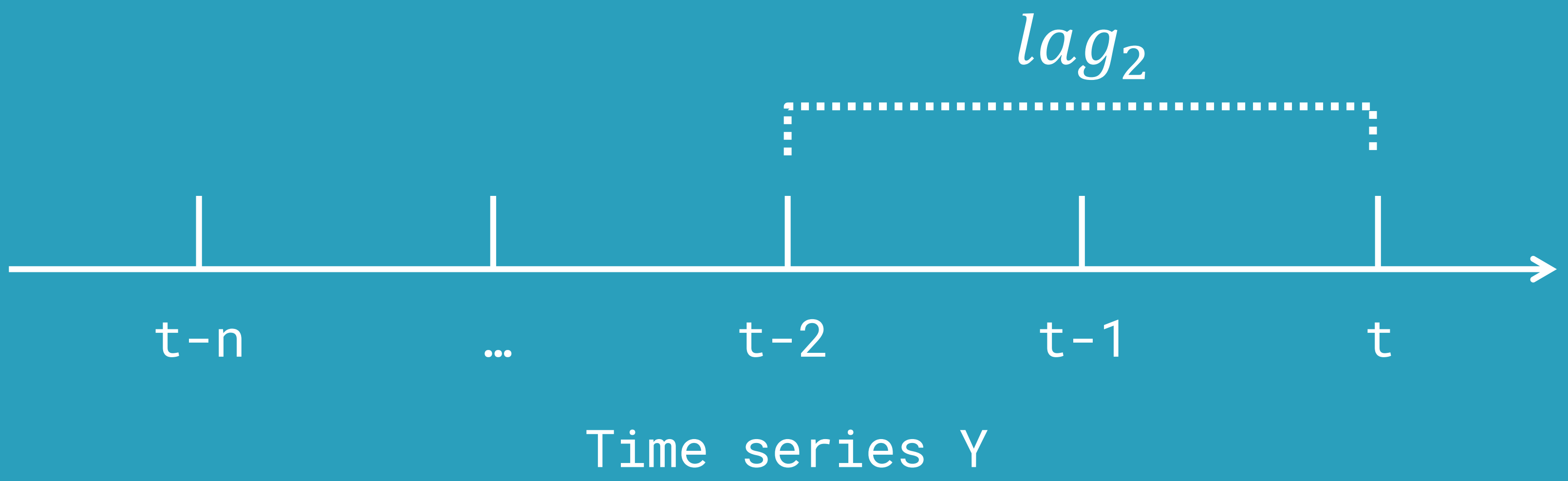
Use both ACF and PACF plots



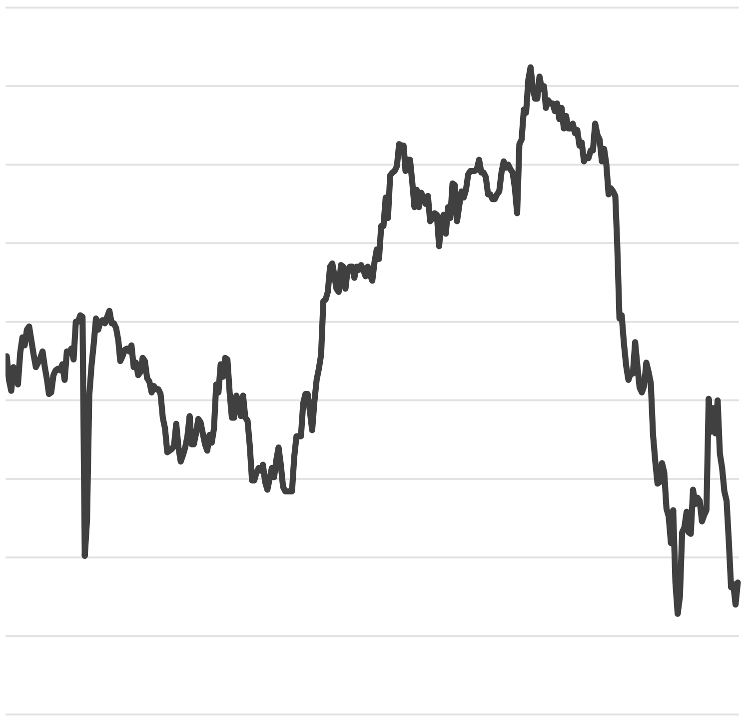
Correlation Between Observations



Correlation Between Observations



Moving Averages and Smoothers



Distractions in the pattern (outliers, extreme values) might hinder the analysis

Smoothers show the middle ground in the data via decimating the highs and lows

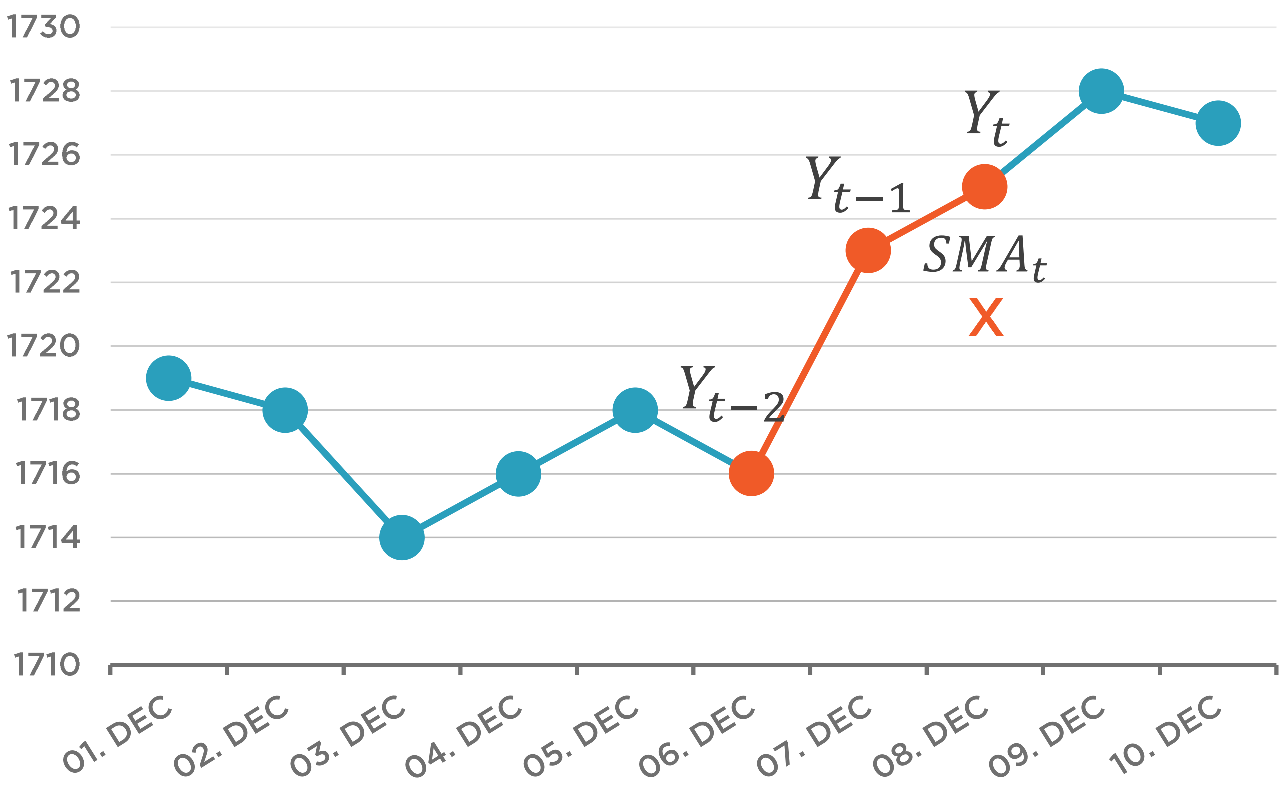
- E.g.: Finding the general trend in stock data through various time windows

Window: The number of successive observations that are combined to find the smooth value for a time point

Simple Moving Average

$$SMA_t = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-n}}{n+1}$$

Window ~ Smoothness



Main Types of Smoothers

Simple Moving Average

Calculates the rolling mean of a given time window

Window length $\sim$ Smoothness

$$SMA_t = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-n}}{n+1}$$

Exponential Moving Average

Observations can be weighted within the smoother

Smoothing factor: α

Deterioration: Recent data is more important than previous observations

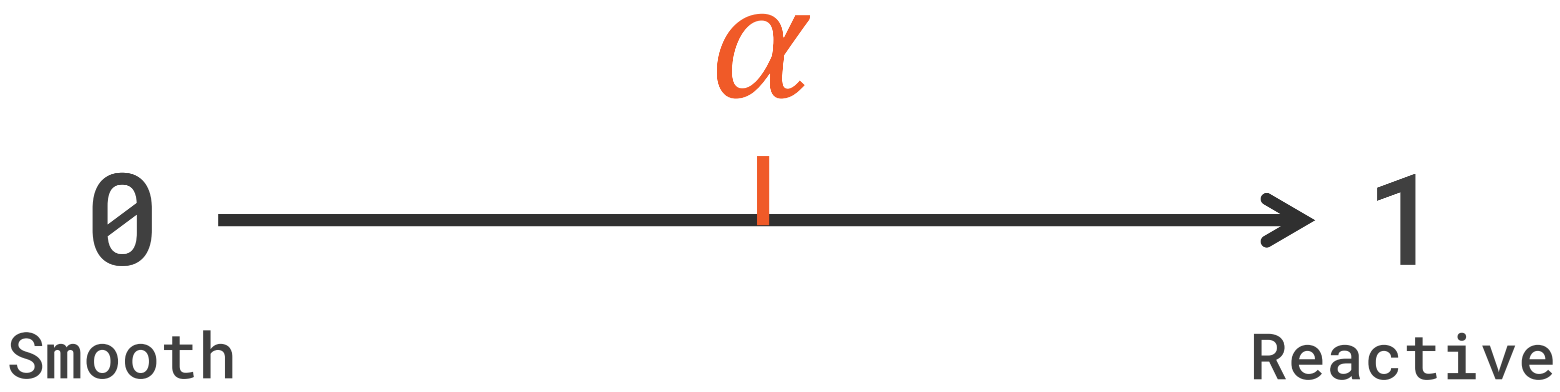
Reactiveness of EWMA: $0 < \alpha < 1$

```
mylynxts = pd.Series(mylynx_df['trappings'].values,
                     index = pd.date_range('31/12/1821' ,
                                           periods = 114,
                                           freq = 'A-DEC'))
```

Generating a Date Index

The time stamp has to be an index of dates with a frequency

The Reactiveness of EWMA



Time Series Statistics

Data visualization techniques with Matplotlib: Figures and subplots

Testing for stationarity with adfuller() from StatsModels

- Differencing non-stationary series

Testing for autocorrelation with ACF and PACF plots from StatsModels

Simple and exponential moving averages to dampen the effect of extreme values