

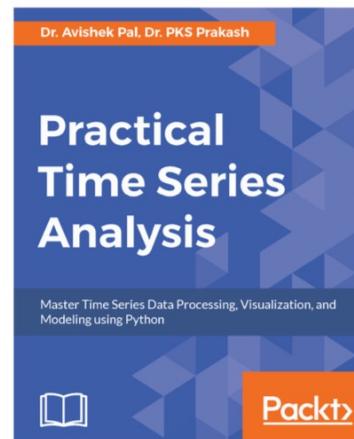
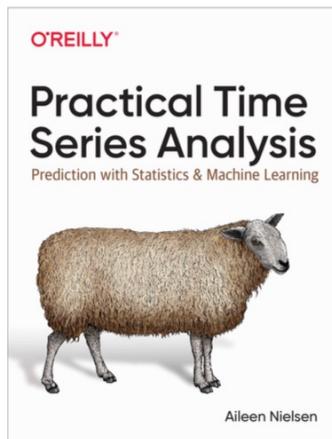
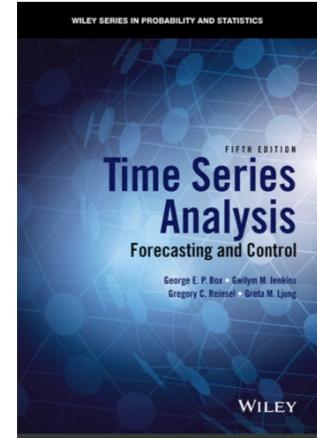
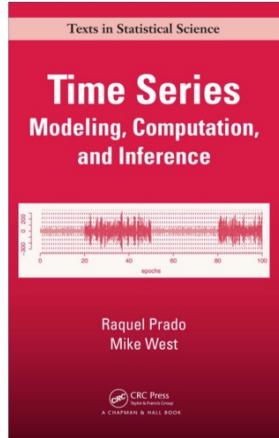
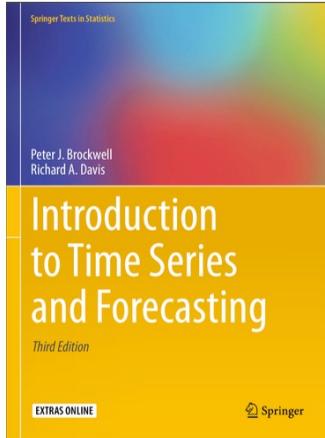
Time Series Analysis

INFO 523 - Lecture 11

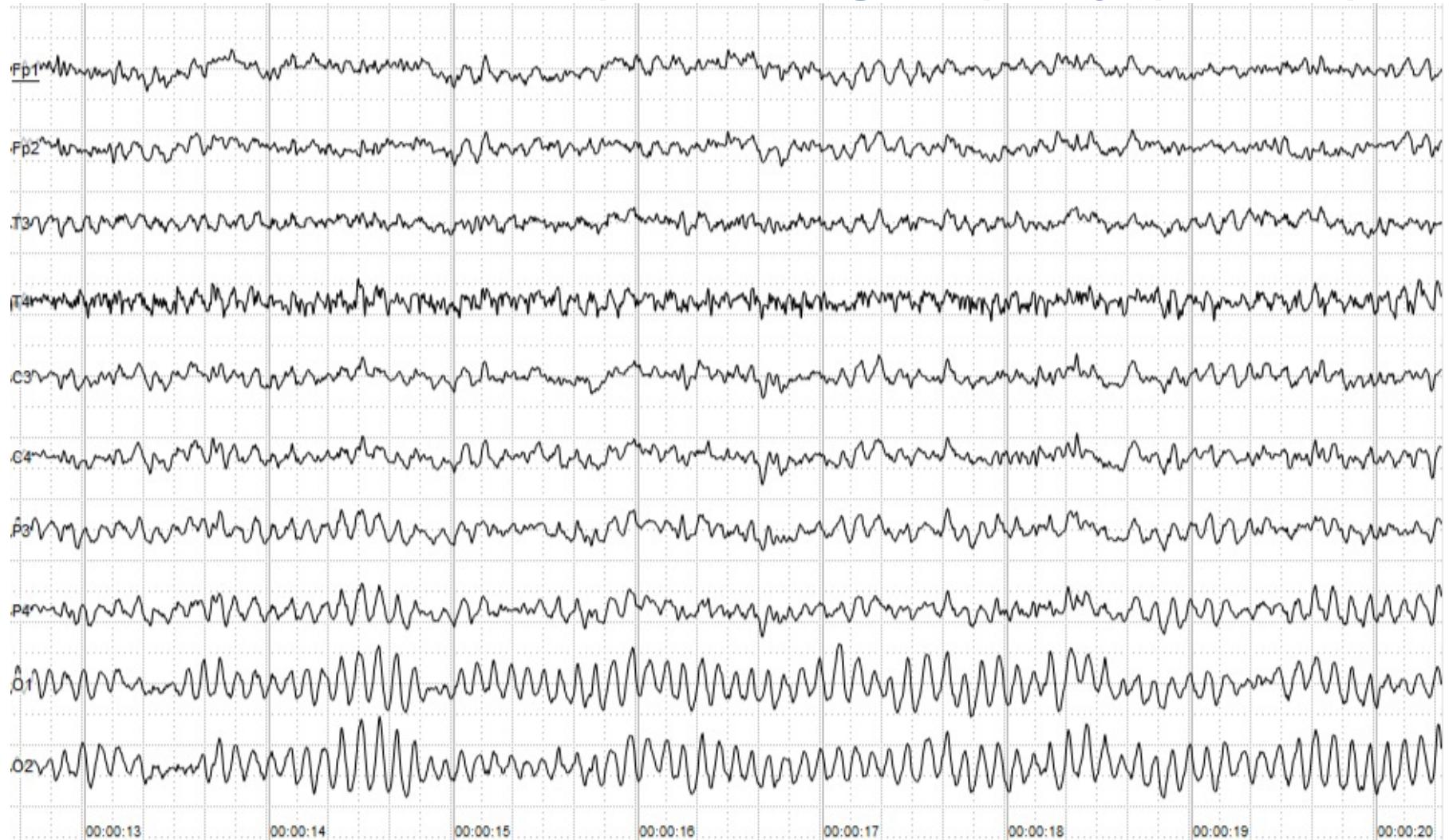
Dr. Greg Chism

Lesson 1: **Understanding Timeseries**

References

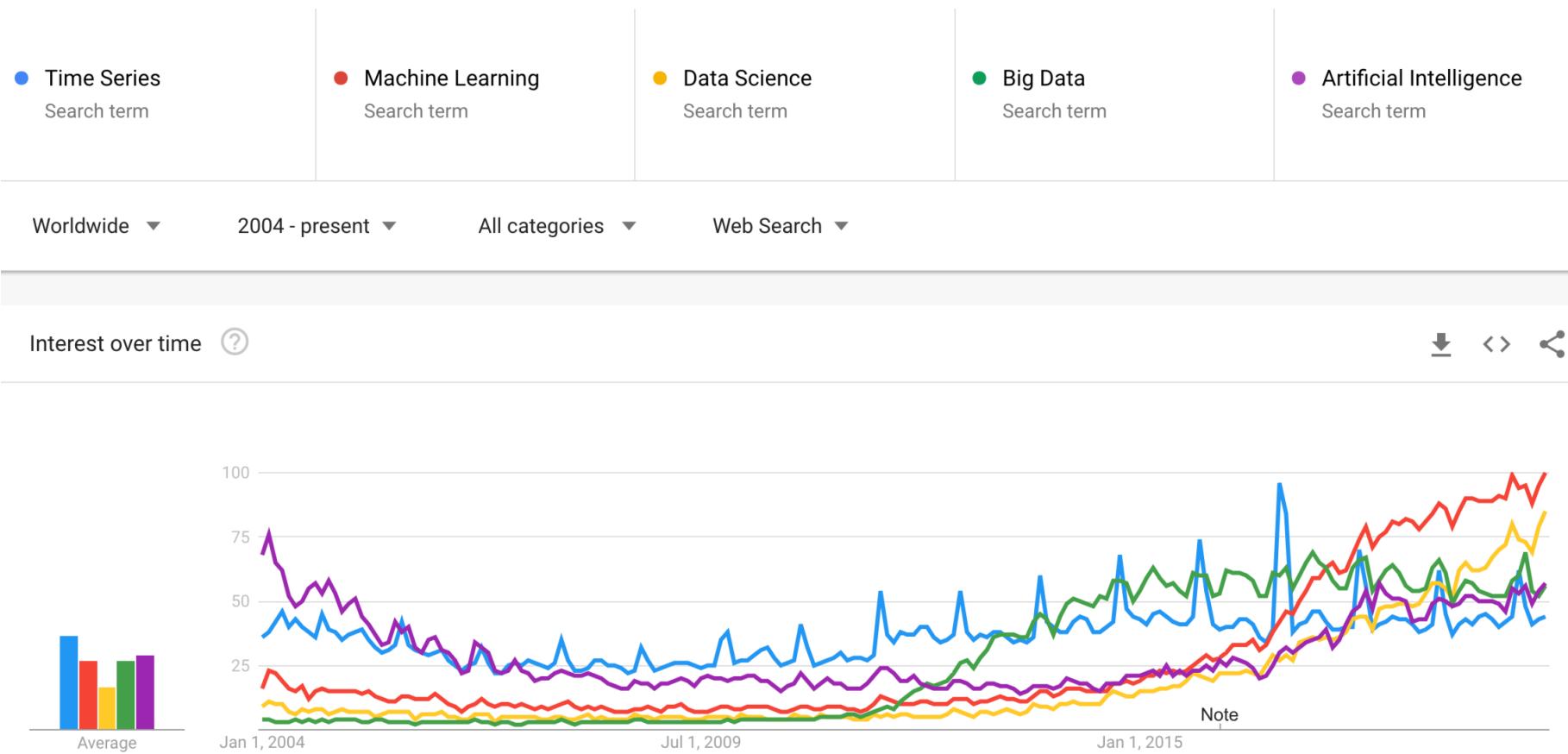


Electroencephalography(EEG)

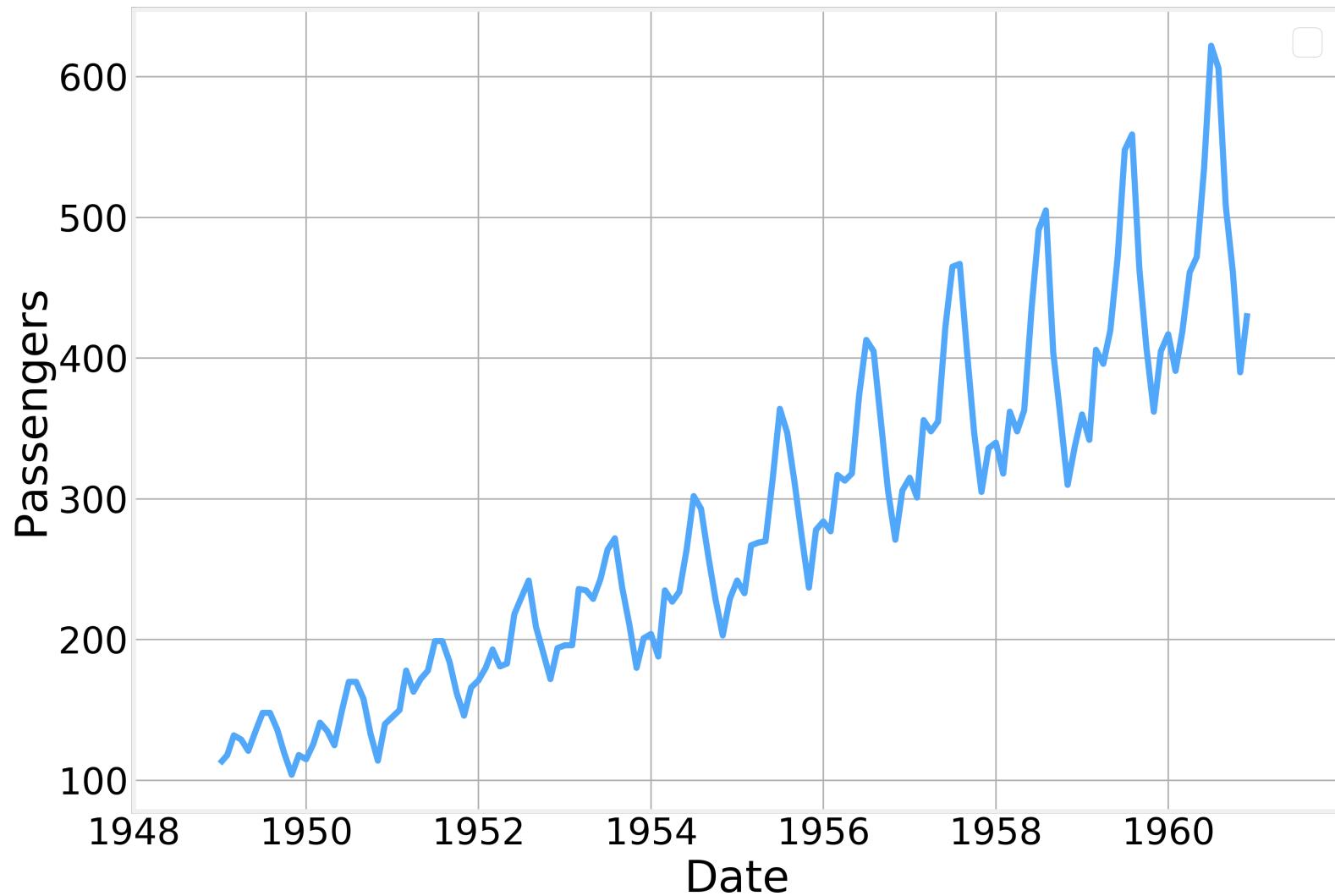


<https://en.wikipedia.org/wiki/Electroencephalography>

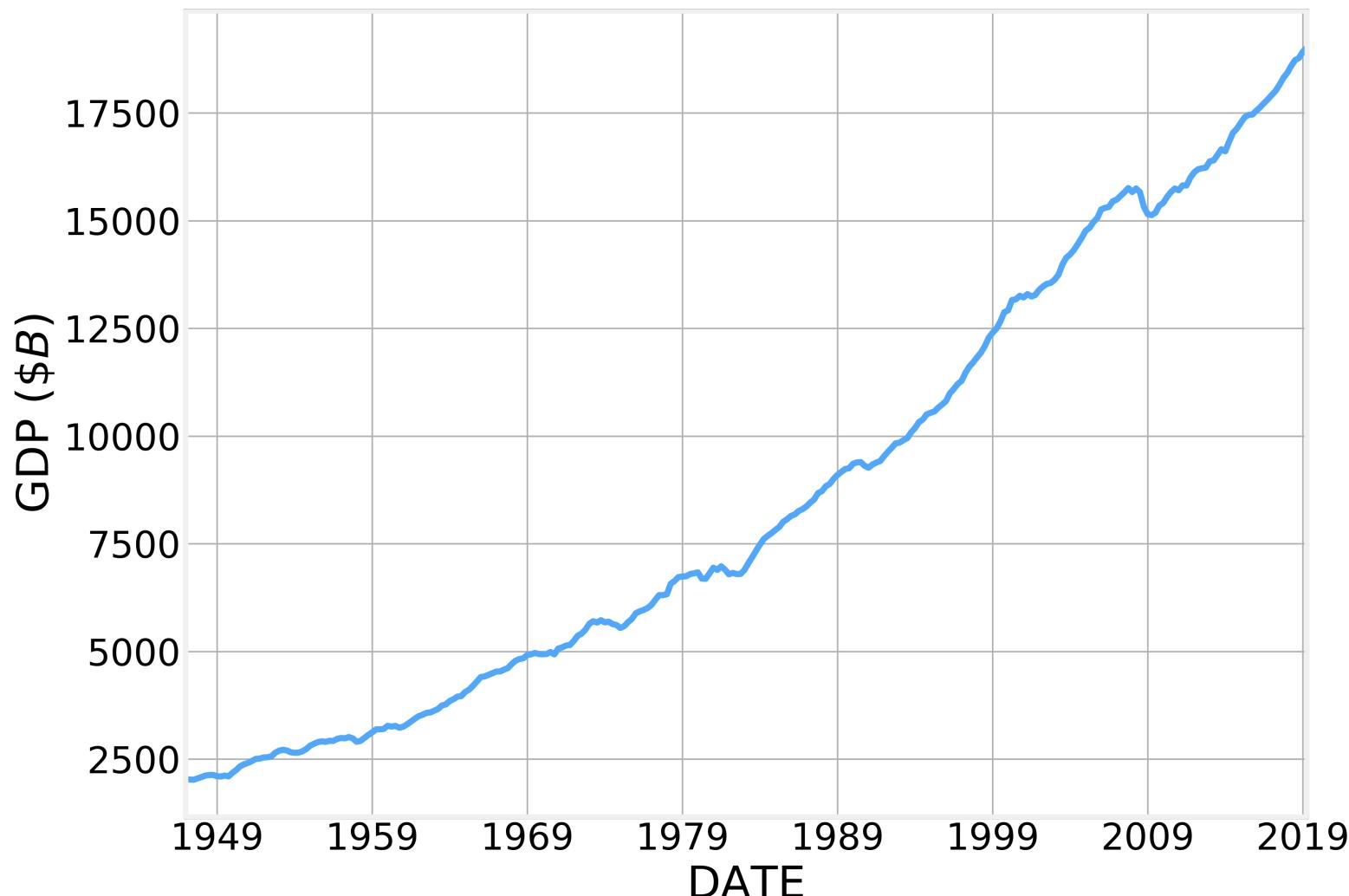
Public Interest – Search Trends



Airline Passengers

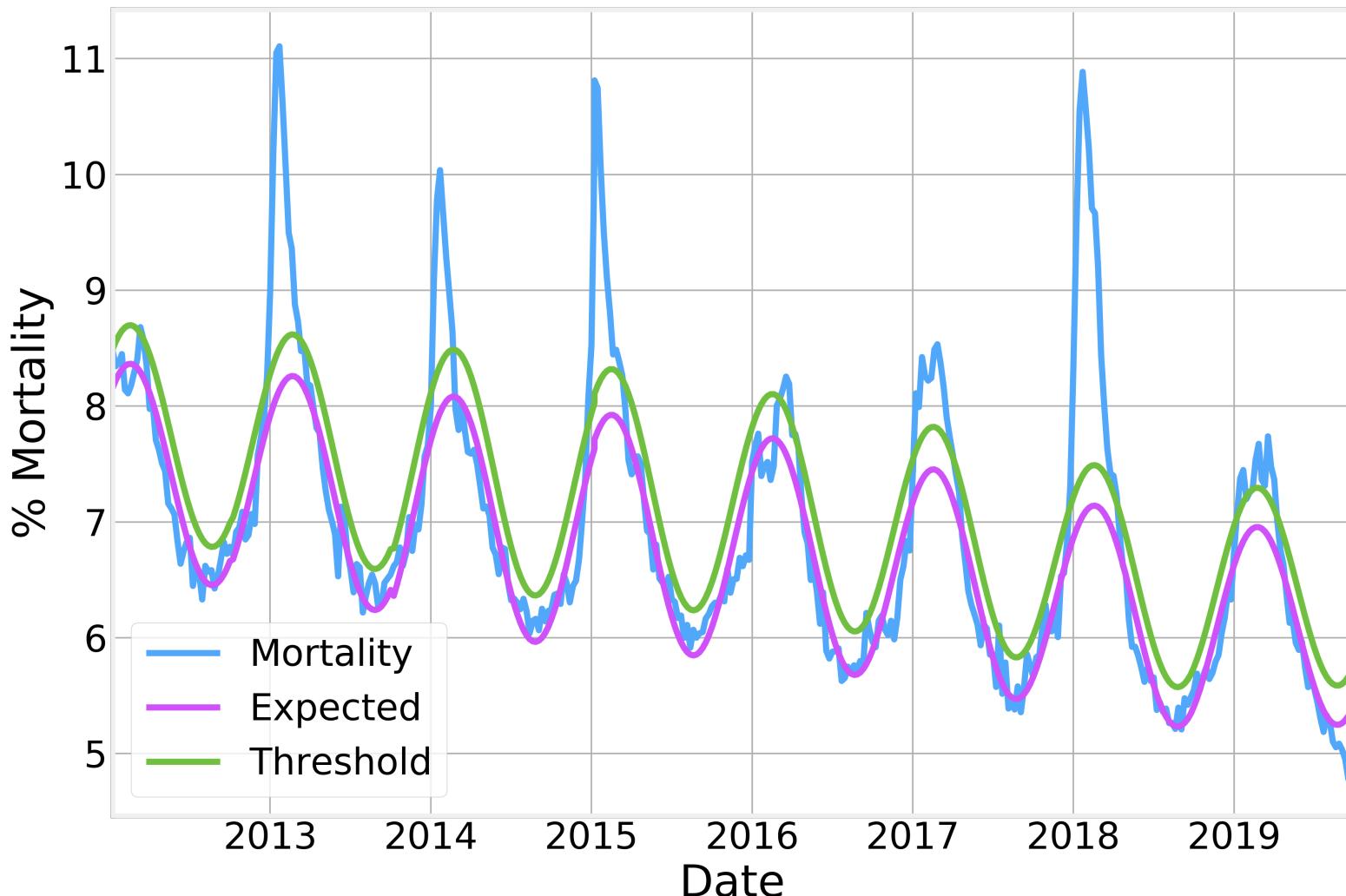


Gross Domestic Product (GDP)

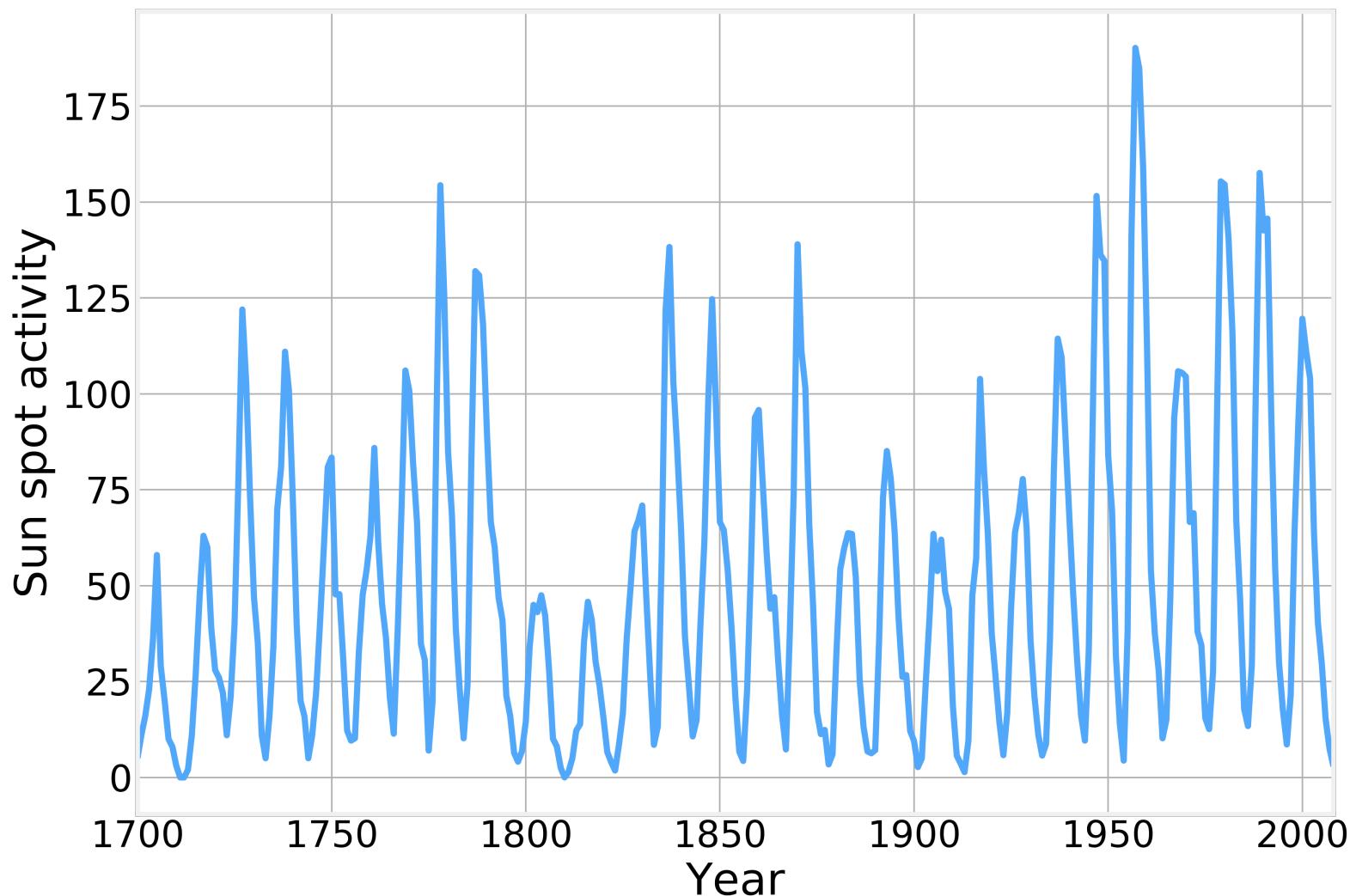


<https://fred.stlouisfed.org/series/GDPC1>

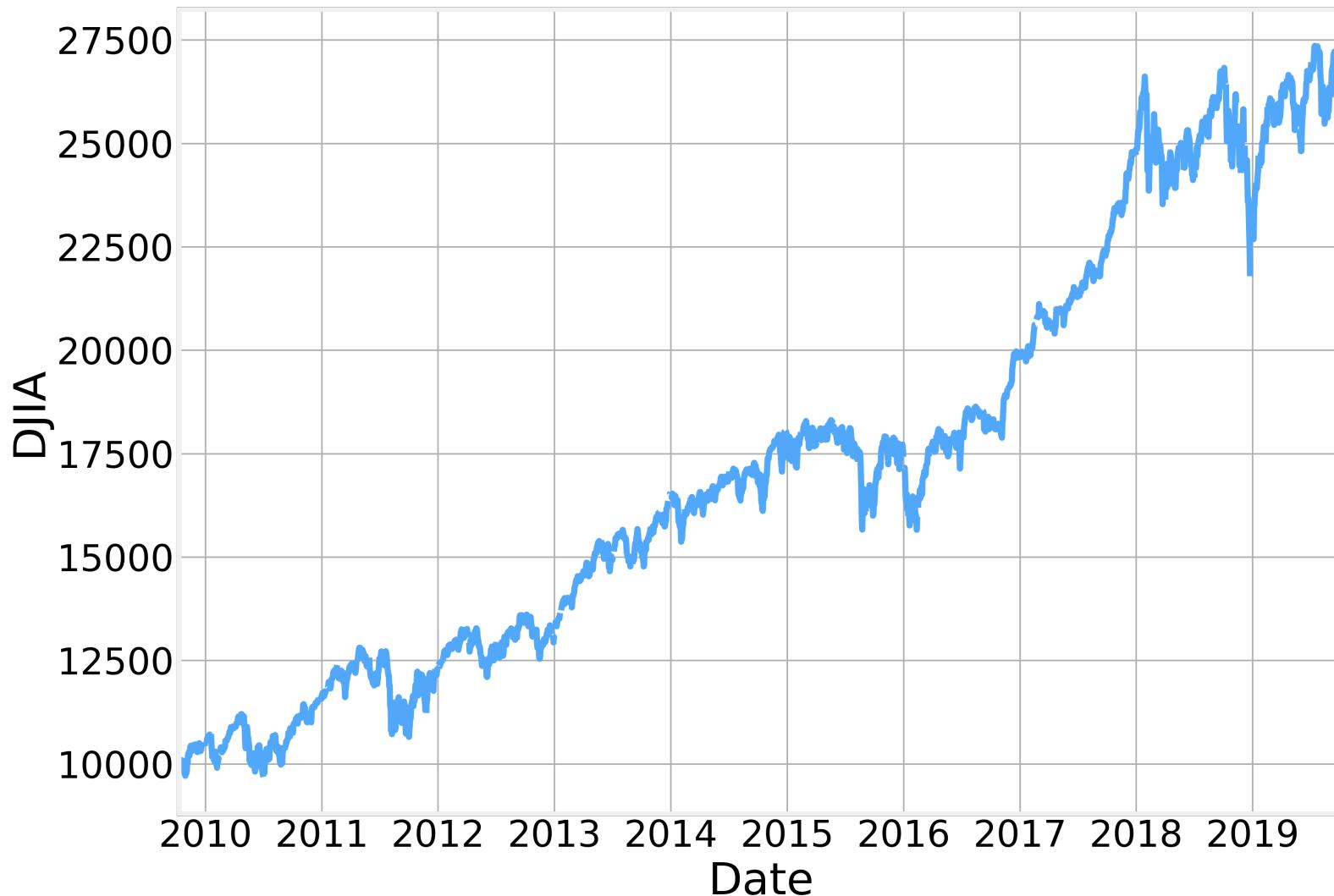
Influenza



Sunspot Activity



Stock Market - DJIA



<https://fred.stlouisfed.org/series/DJIA>

Time Series

Time Series

- A set of values measured sequentially in time

Time Series

- A set of values measured sequentially in time
- Values are typically (but not always) measured at equal intervals, x_1, x_2, x_3, x_3 , etc...

Time Series

- A set of values measured sequentially in time
- Values are typically (but not always) measured at equal intervals, x_1, x_2, x_3, x_4 , etc...
- Values can be:
 - continuous
 - discrete or symbolic (words)

Time Series

- A set of values measured sequentially in time
- Values are typically (but not always) measured at equal intervals, x_1, x_2, x_3, x_3 , etc...
- Values can be:
 - continuous
 - discrete or symbolic (words)
- Associated with empirical observation of time varying phenomena:

Time Series

- A set of values measured sequentially in time
- Values are typically (but not always) measured at equal intervals, x_1, x_2, x_3, x_3 , etc...
- Values can be:
 - continuous
 - discrete or symbolic (words)
- Associated with empirical observation of time varying phenomena:
 - Stock market prices (day, hour, minute, tick, etc...)
 - Temperatures (day, minute, second, etc...)
 - Number of patients (week, month, etc...)
 - GDP (quarter, year, etc...)

Time Series

- A set of values measured sequentially in time
- Values are typically (but not always) measured at equal intervals, x_1, x_2, x_3, x_3 , etc...
- Values can be:
 - continuous
 - discrete or symbolic (words)
- Associated with empirical observation of time varying phenomena:
 - Stock market prices (day, hour, minute, tick, etc...)
 - Temperatures (day, minute, second, etc...)
 - Number of patients (week, month, etc...)
 - GDP (quarter, year, etc...)
- Forecasting requires predicting future values based on past behavior

Time Series

- The value of the time series at time t is given by x_t

Time Series

- The value of the time series at time t is given by x_t
- The values at a given lag l are given by x_{t-l}

Time Series

- The value of the time series at time t is given by x_t
- The values at a given lag l are given by x_{t-l}
- The mean of the overall signal is μ and the corresponding running value is μ_t^w

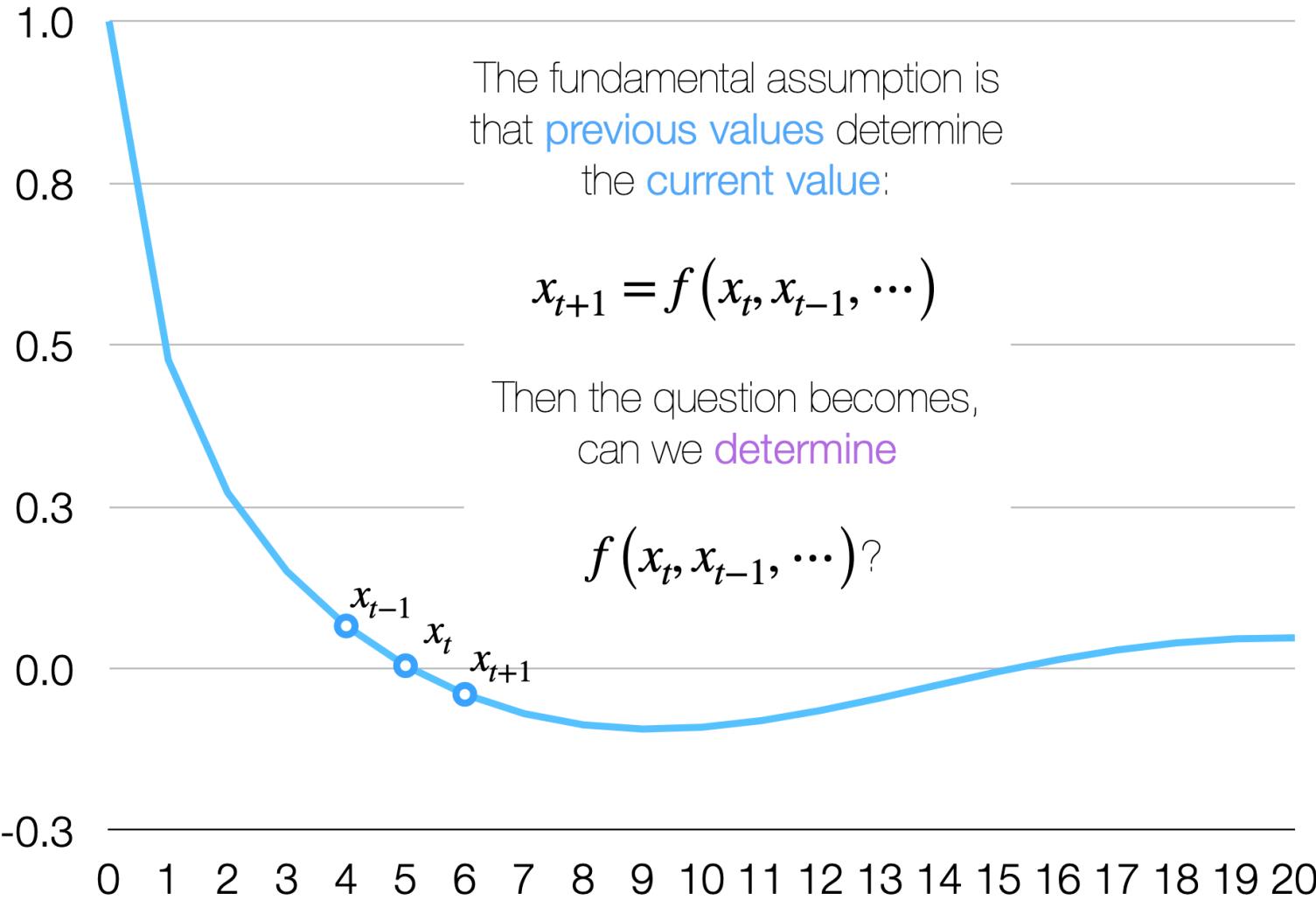
Time Series

- The value of the time series at time t is given by x_t
- The values at a given lag l are given by x_{t-l}
- The mean of the overall signal is μ and the corresponding running value is μ_t^w
- The variance of the overall signal is " and the corresponding running value is σ_t^w

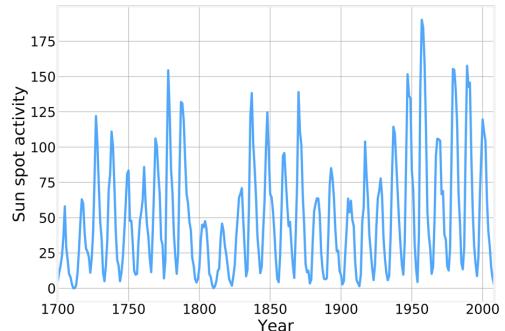
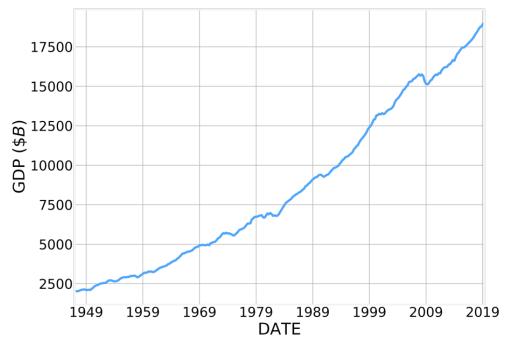
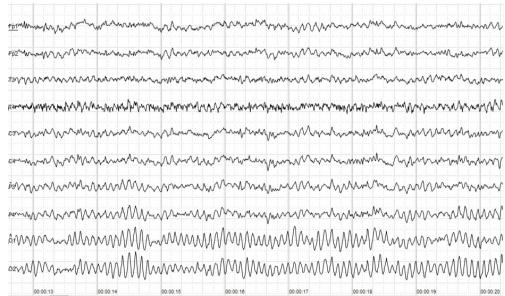
Time Series

- The value of the time series at time t is given by x_t
- The values at a given lag l are given by x_{t-l}
- The mean of the overall signal is μ and the corresponding running value is μ_t^w
- The variance of the overall signal is " and the corresponding running value is σ_t^w
- Running values are calculated over a window of width w

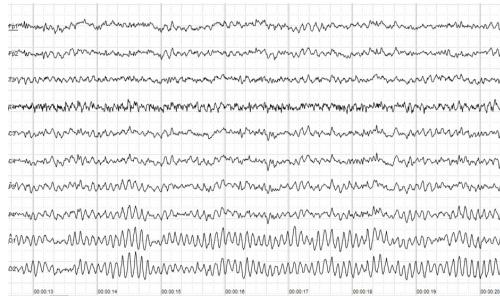
Time Series Analysis



Three fundamental behaviors

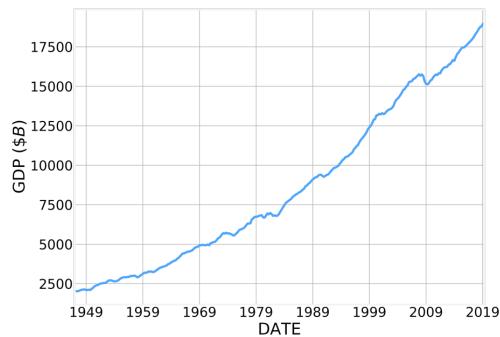


Three fundamental behaviors



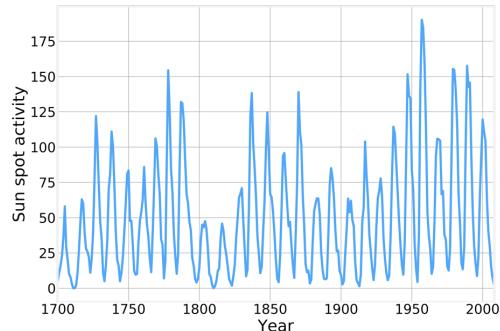
Stationarity

$$\langle x_t \rangle \approx \text{constant}$$



Trend

$$\langle x_t \rangle \approx ct$$



Seasonality

$$X_{t+T} \approx X_t$$

Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**

Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**
- In particular:
 - **Mean** - Average value stays constant

Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**
- In particular:
 - **Mean** - Average value stays constant
 - **Variance** – The width of the curve is bounded

Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**
- In particular:
 - **Mean** - Average value stays constant
 - **Variance** – The width of the curve is bounded
 - **Covariance** – Correlation between points is independent of time

Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**
- In particular:
 - **Mean** - Average value stays constant
 - **Variance** – The width of the curve is bounded
 - **Covariance** – Correlation between points is independent of time
- Stationary processes are **easier** to analyze

Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**
- In particular:
 - **Mean** - Average value stays constant
 - **Variance** – The width of the curve is bounded
 - **Covariance** – Correlation between points is independent of time
- Stationary processes are **easier** to analyze
- Many time series analysis algorithms assume the time series to be stationary

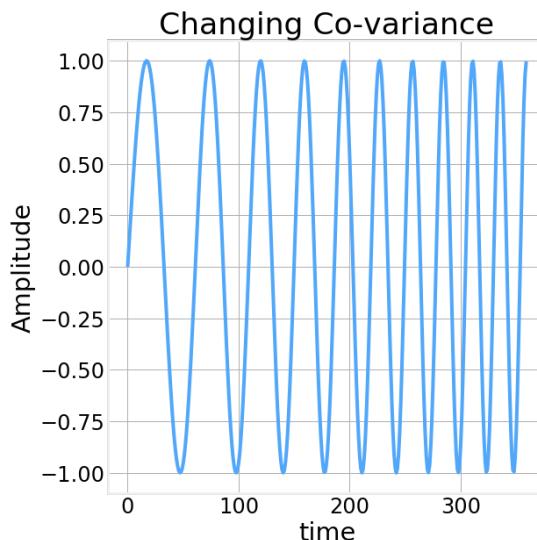
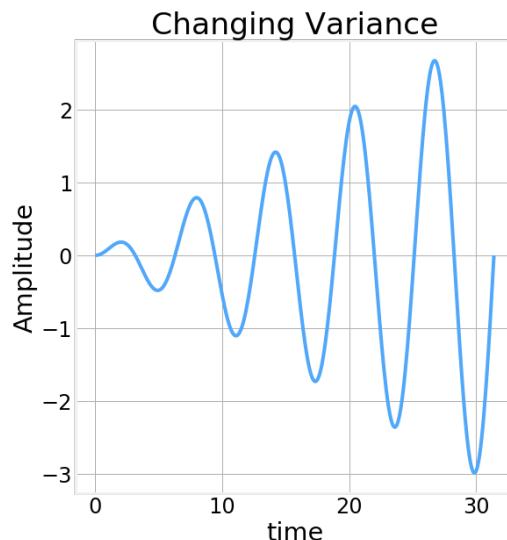
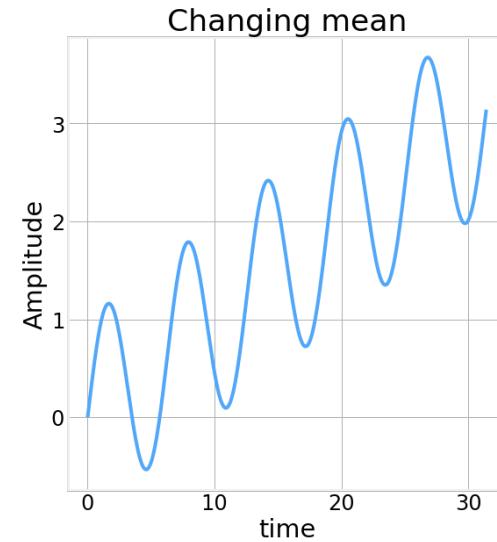
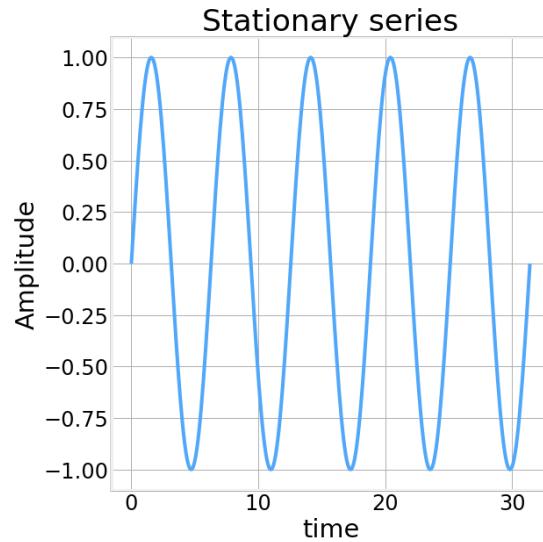
Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**
- In particular:
 - **Mean** - Average value stays constant
 - **Variance** – The width of the curve is bounded
 - **Covariance** – Correlation between points is independent of time
- Stationary processes are **easier** to analyze
- Many time series analysis algorithms assume the time series to be stationary
- Several rigorous tests for stationarity have been developed such as the **(Augmented) Dickey-Fuller** and **Hurst Exponent**

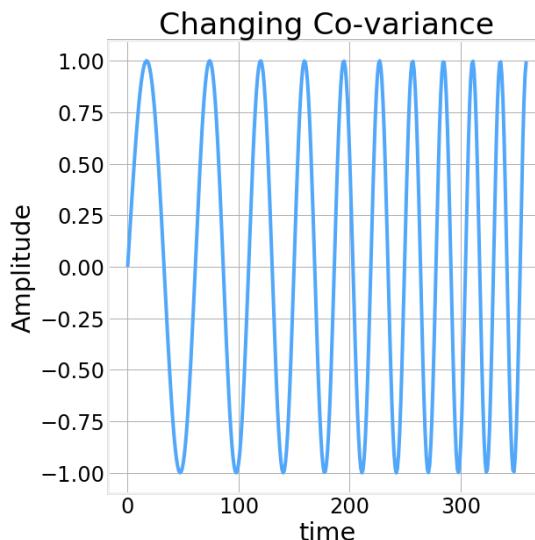
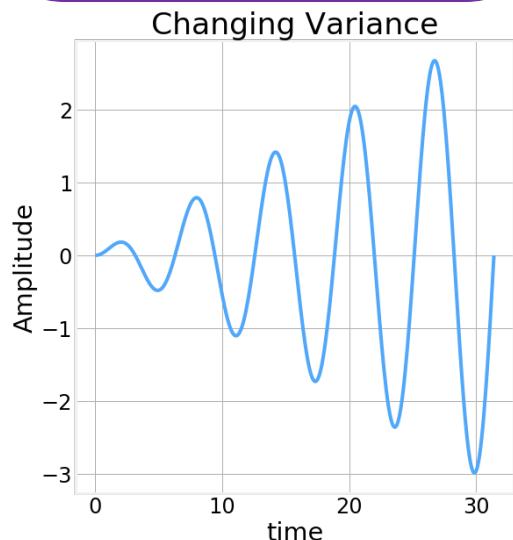
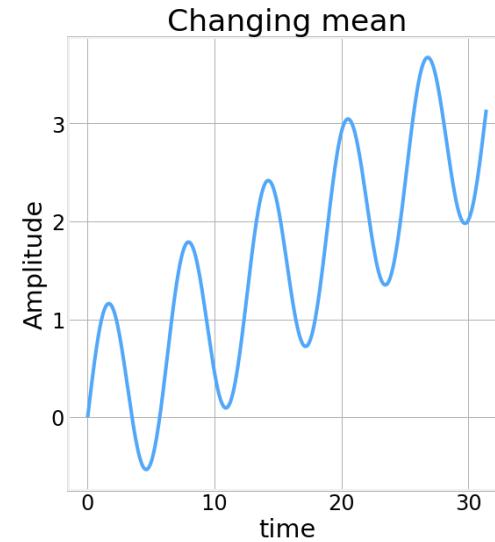
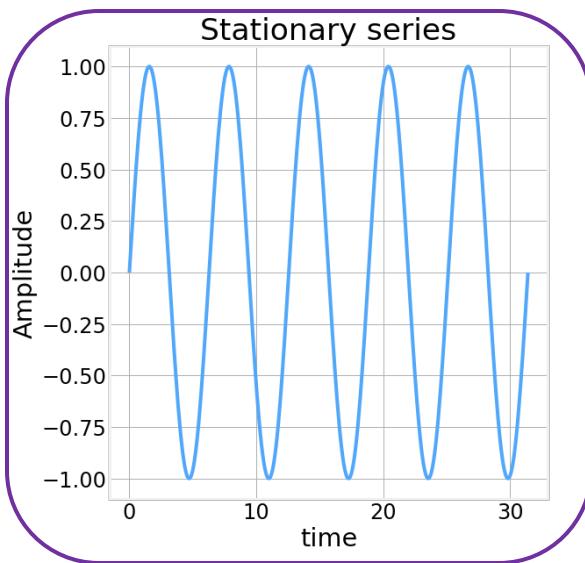
Stationarity

- A time series is said to be stationary if its basic statistical properties are **independent of time**
- In particular:
 - **Mean** - Average value stays constant
 - **Variance** – The width of the curve is bounded
 - **Covariance** – Correlation between points is independent of time
- Stationary processes are **easier** to analyze
- Many time series analysis algorithms assume the time series to be stationary
- Several rigorous tests for stationarity have been developed such as the **(Augmented) Dickey-Fuller** and **Hurst Exponent**
- Typically, the first step of any analysis is to transform the series to make it stationary

Stationarity



Stationarity



Code – Data Exploration

Trend

- Many time series have a clear trend or tendency:

Trend

- Many time series have a clear trend or tendency:
 - Stock market indices tend to go up over time

Trend

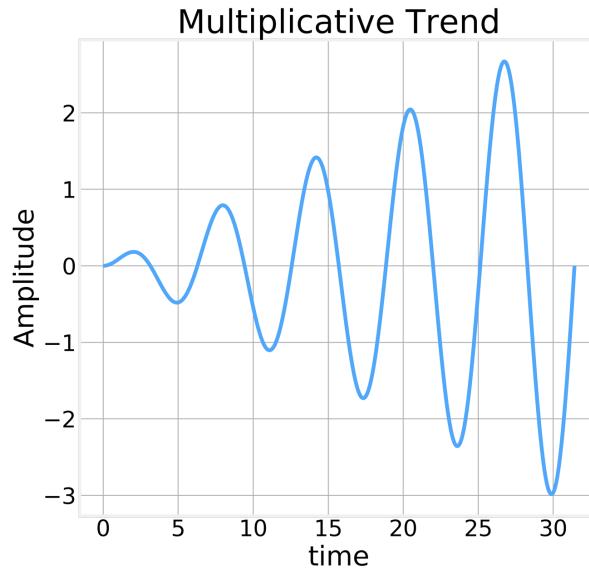
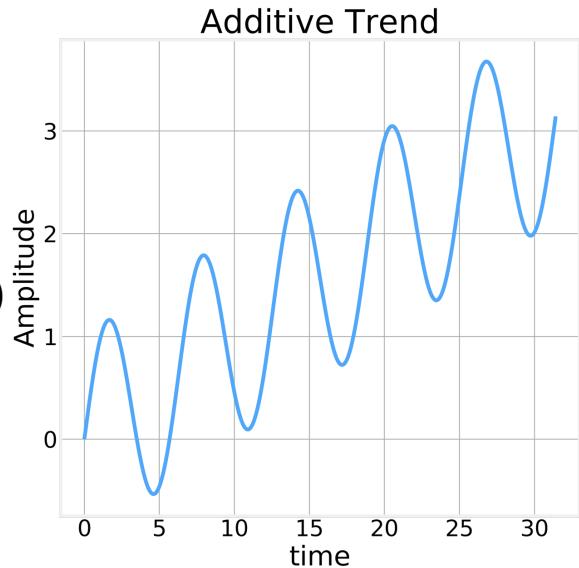
- Many time series have a clear trend or tendency:
 - Stock market indices tend to go up over time
 - Number of cases of preventable diseases tends to go down over time

Trend

- Many time series have a clear trend or tendency:
 - Stock market indices tend to go up over time
 - Number of cases of preventable diseases tends to go down over time
- Trends can be additive or multiplicative:

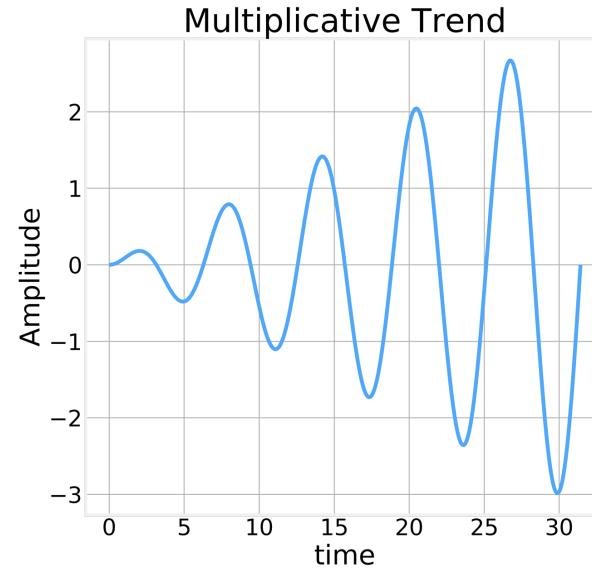
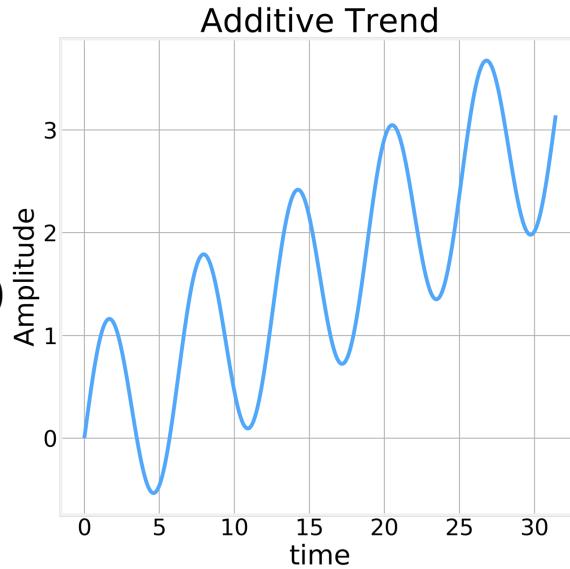
Trend

- Many time series have a clear trend or tendency:
 - Stock market indices tend to go up over time
 - Number of cases of preventable diseases tends to go down over time
- Trends can be additive or multiplicative:



Trend

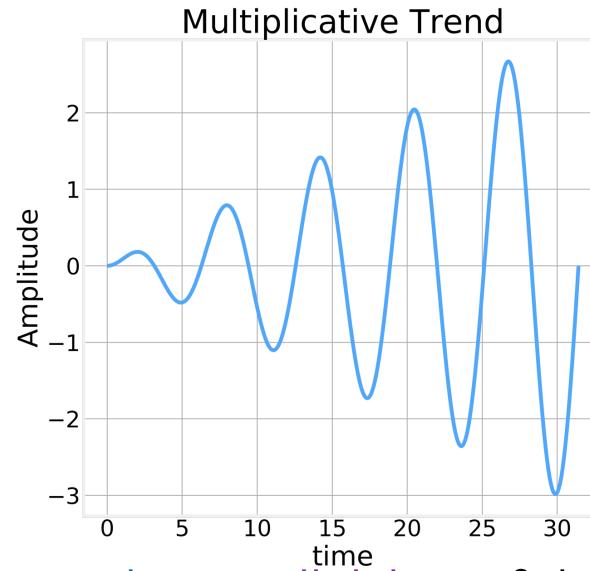
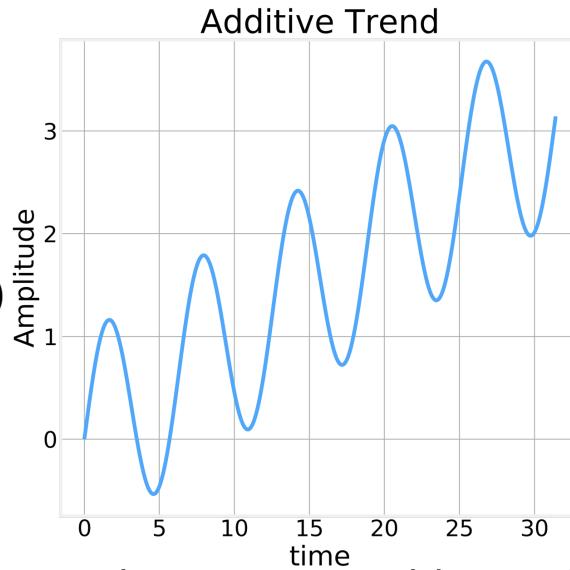
- Many time series have a clear trend or tendency:
 - Stock market indices tend to go up over time
 - Number of cases of preventable diseases tends to go down over time
- Trends can be additive or multiplicative:



- Trends can be removed by subtraction or division of the correct values

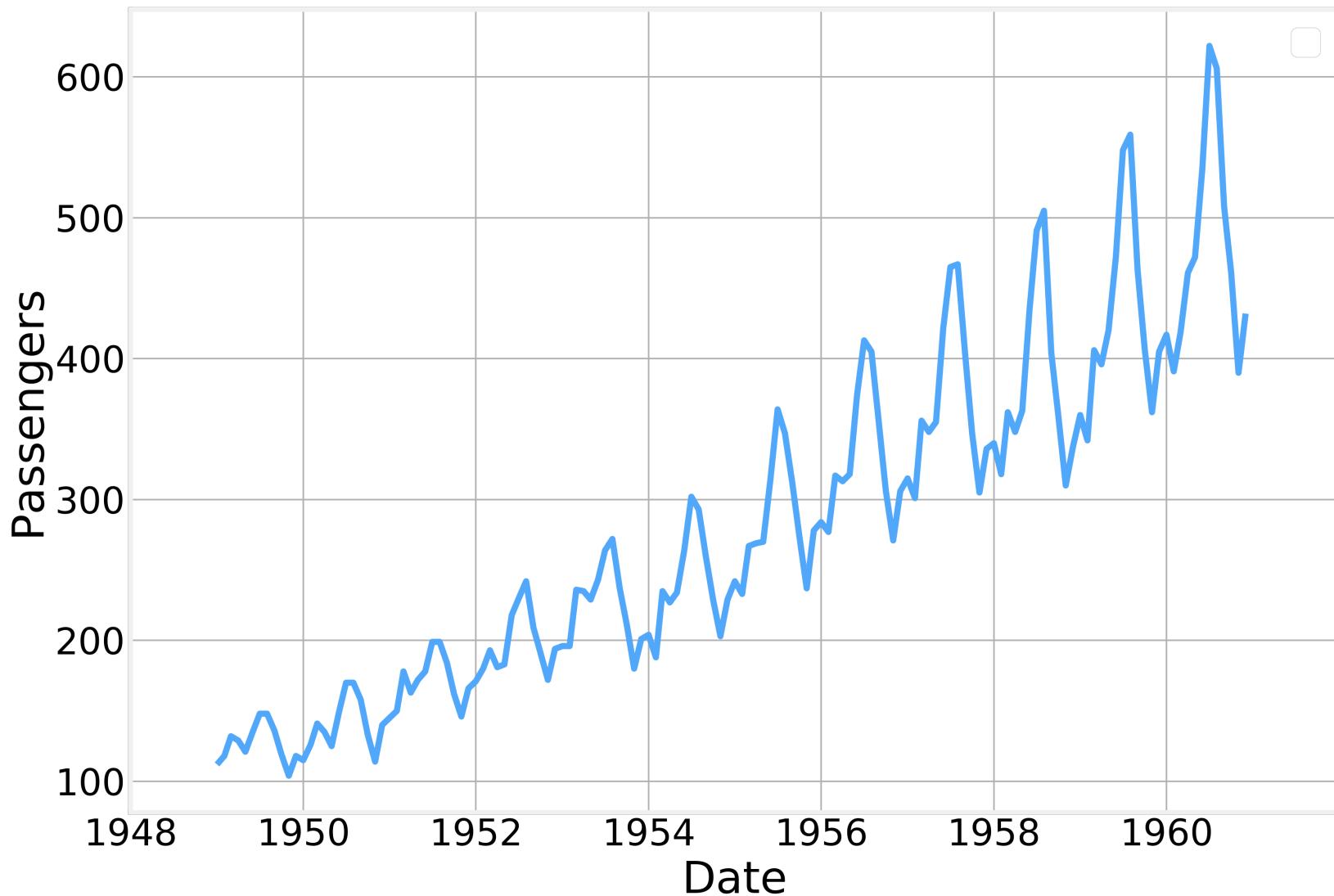
Trend

- Many time series have a clear trend or tendency:
 - Stock market indices tend to go up over time
 - Number of cases of preventable diseases tends to go down over time
- Trends can be additive or multiplicative:

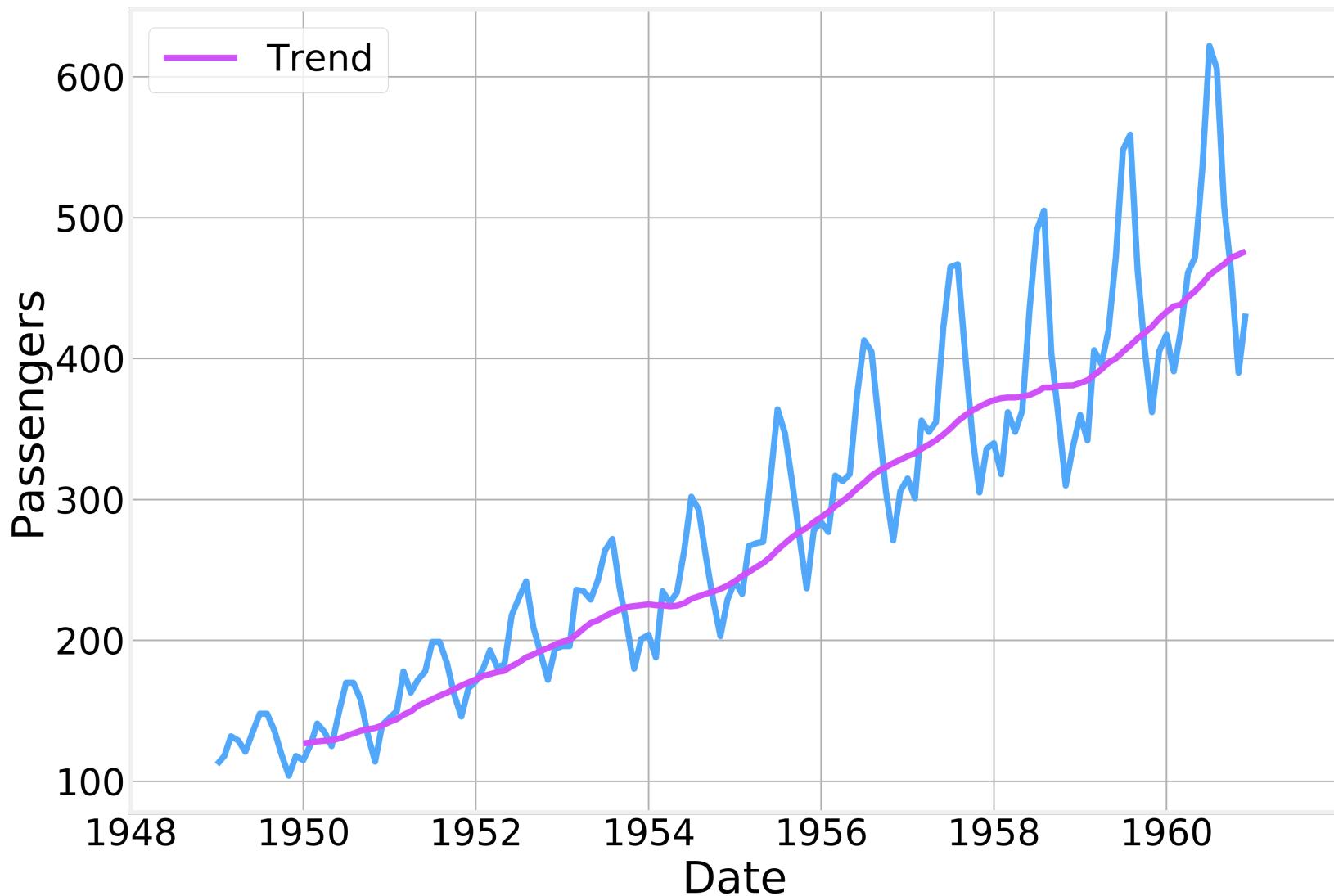


- Trends can be removed by subtraction or division of the correct values
- One way to determine the trend is to find the running average

Trend



Trend



Seasonality

Seasonality

- Many of the phenomena we might be interested in vary in time in a **cyclical** or **seasonal** fashion

Seasonality

- Many of the phenomena we might be interested in vary in time in a **cyclical** or **seasonal** fashion
 - Ice-cream sales peak in the **summer** and drop in the **winter**

Seasonality

- Many of the phenomena we might be interested in vary in time in a **cyclical** or **seasonal** fashion
 - Ice-cream sales peak in the **summer** and drop in the **winter**
 - Number of cell phone calls made is larger during the day than at night

Seasonality

- Many of the phenomena we might be interested in vary in time in a **cyclical** or **seasonal** fashion
 - Ice-cream sales peak in the **summer** and drop in the **winter**
 - Number of cell phone calls made is larger during the day than at night
 - Many types of crime are more frequent at **night** than during the **day**

Seasonality

- Many of the phenomena we might be interested in vary in time in a **cyclical** or **seasonal** fashion
 - Ice-cream sales peak in the **summer** and drop in the **winter**
 - Number of cell phone calls made is larger during the day than at night
 - Many types of crime are more frequent at **night** than during the **day**
 - Visits to museums are more frequent in the **weekend** than in **weekdays**

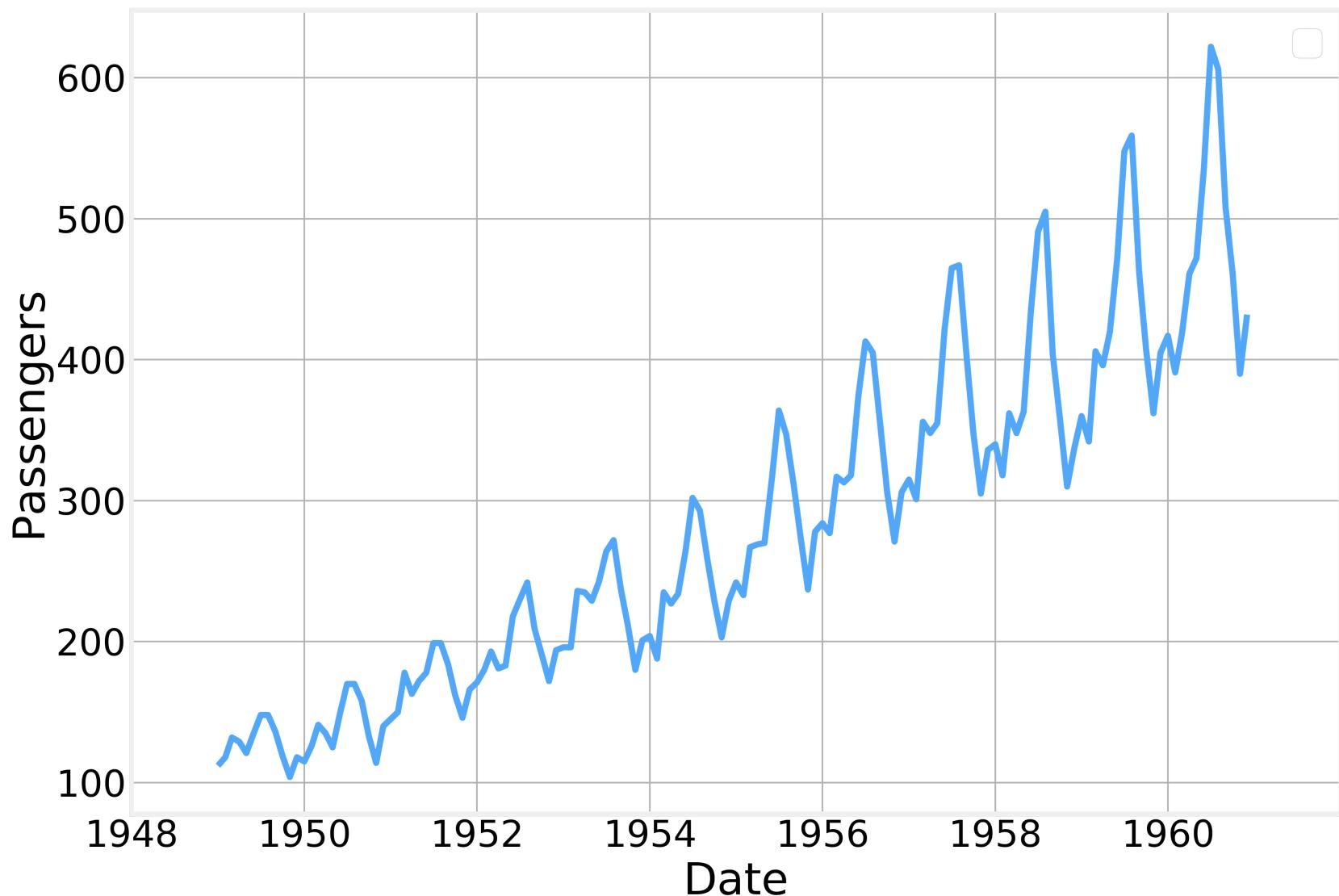
Seasonality

- Many of the phenomena we might be interested in vary in time in a **cyclical** or **seasonal** fashion
 - Ice-cream sales peak in the **summer** and drop in the **winter**
 - Number of cell phone calls made is larger during the day than at night
 - Many types of crime are more frequent at **night** than during the **day**
 - Visits to museums are more frequent in the **weekend** than in **weekdays**
 - The stock market grows during **bull** periods and shrinks during **bear** periods

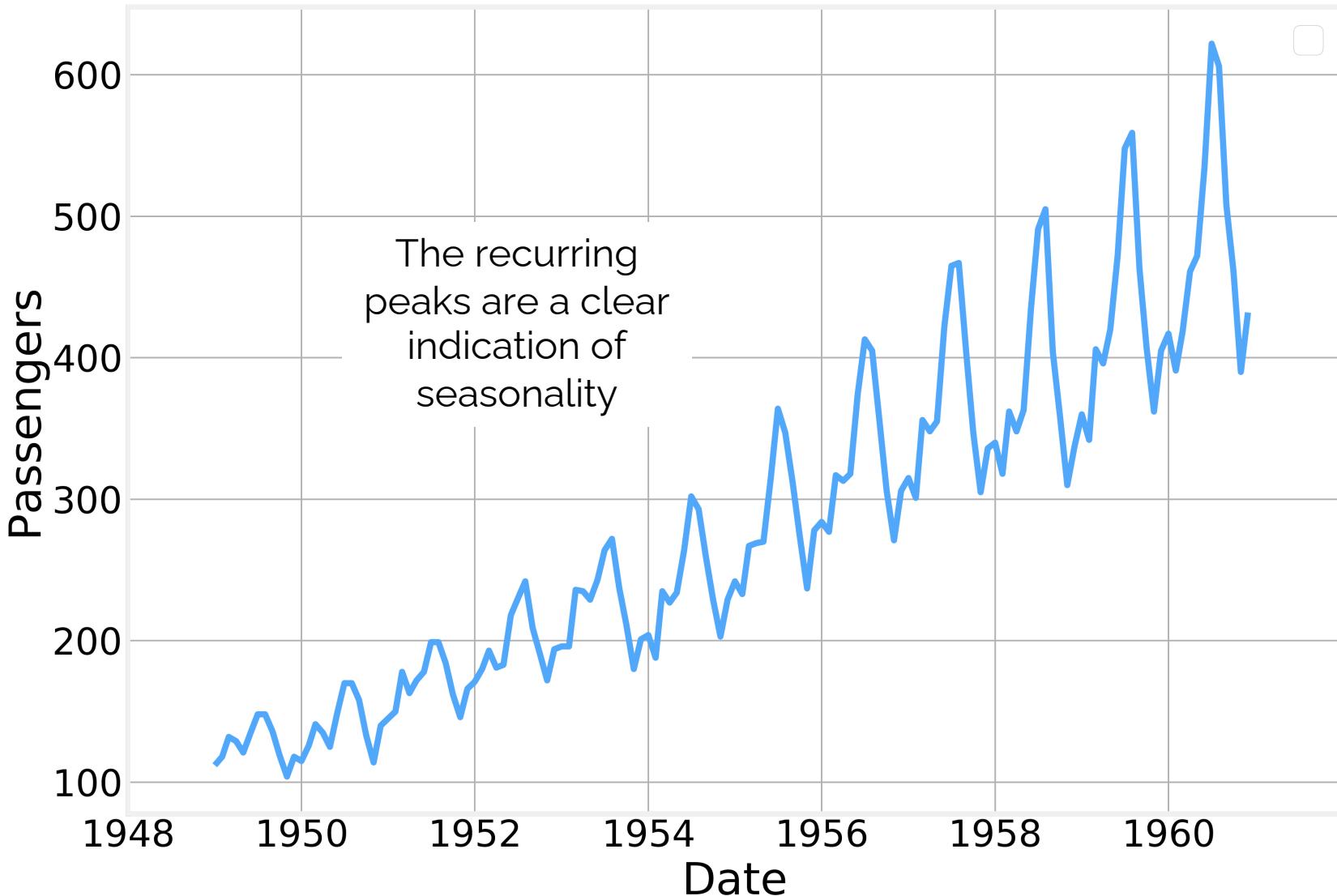
Seasonality

- Many of the phenomena we might be interested in vary in time in a **cyclical** or **seasonal** fashion
 - Ice-cream sales peak in the **summer** and drop in the **winter**
 - Number of cell phone calls made is larger during the day than at night
 - Many types of crime are more frequent at **night** than during the **day**
 - Visits to museums are more frequent in the **weekend** than in **weekdays**
 - The stock market grows during **bull** periods and shrinks during **bear** periods
- Understanding the seasonality of a time series provides important information about its long-term behavior and is extremely useful in predicting future values

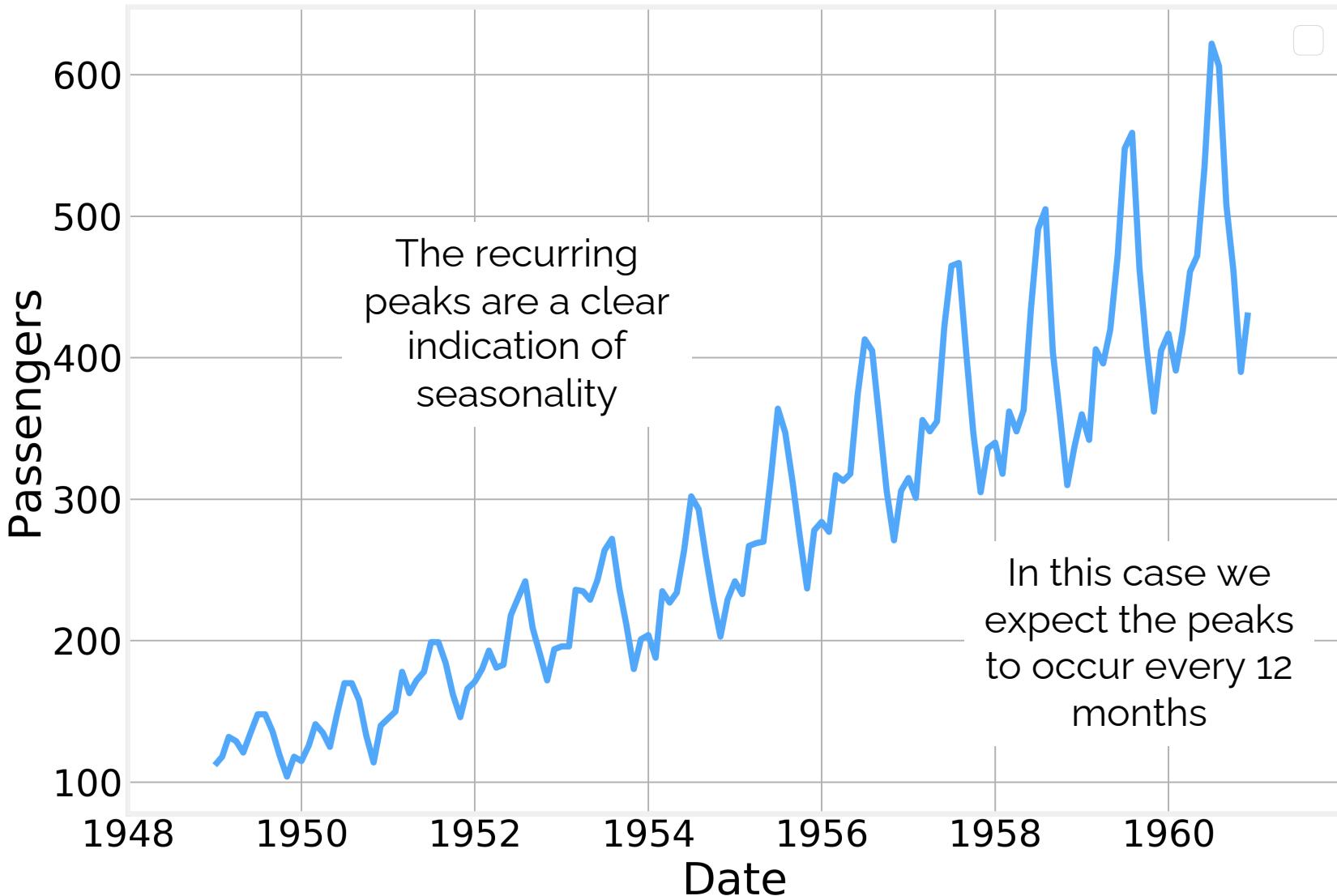
Seasonality



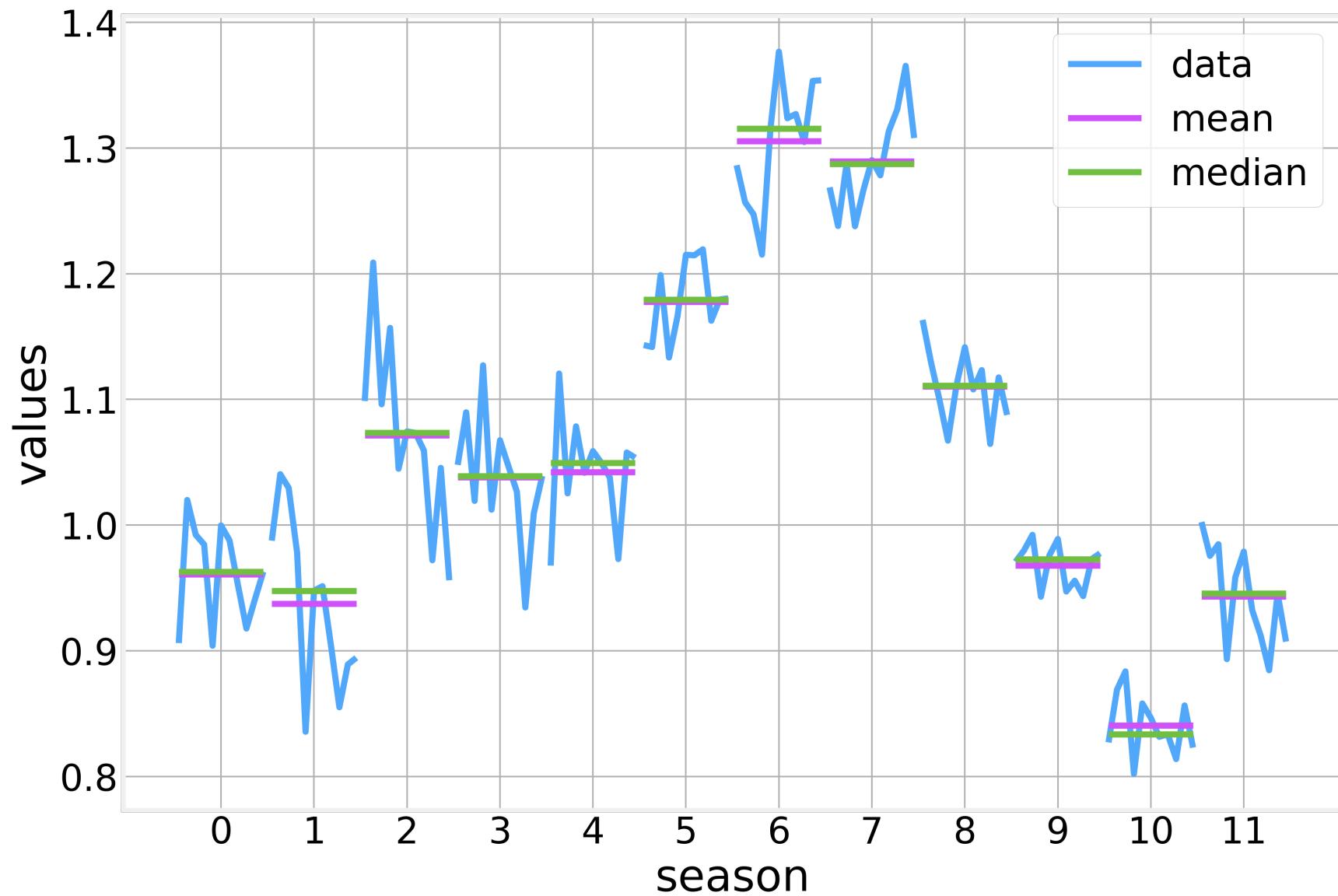
Seasonality



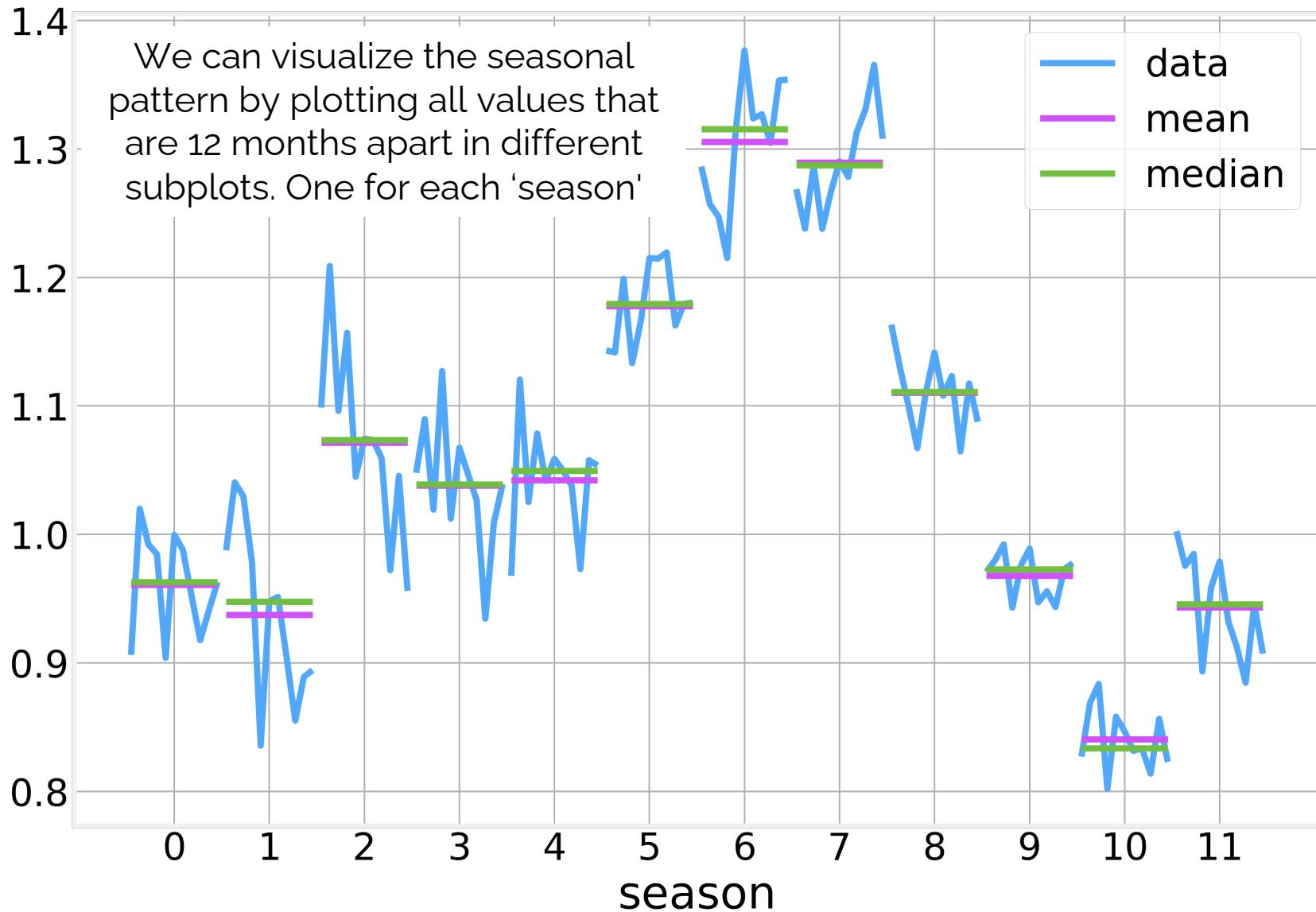
Seasonality



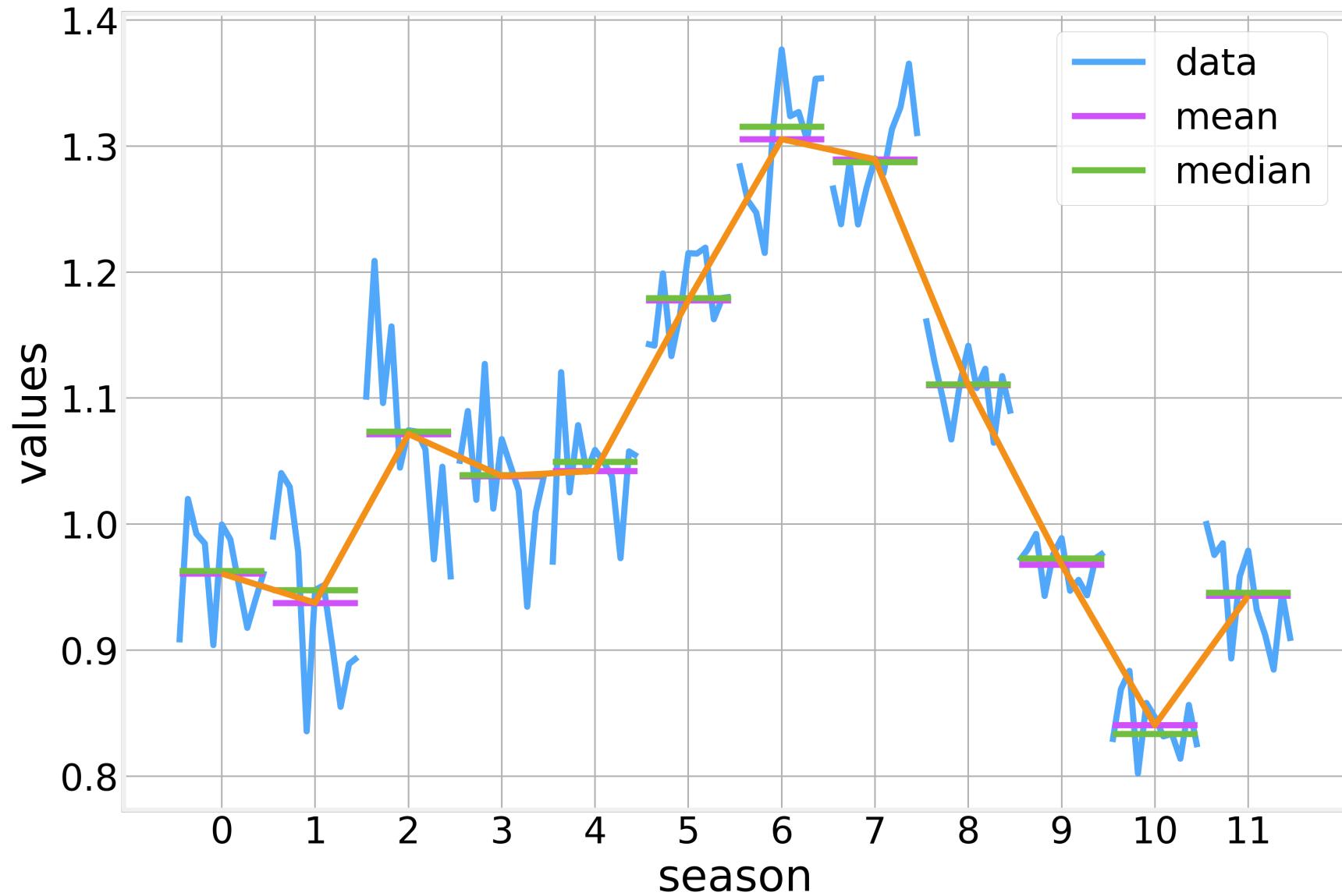
Seasonality



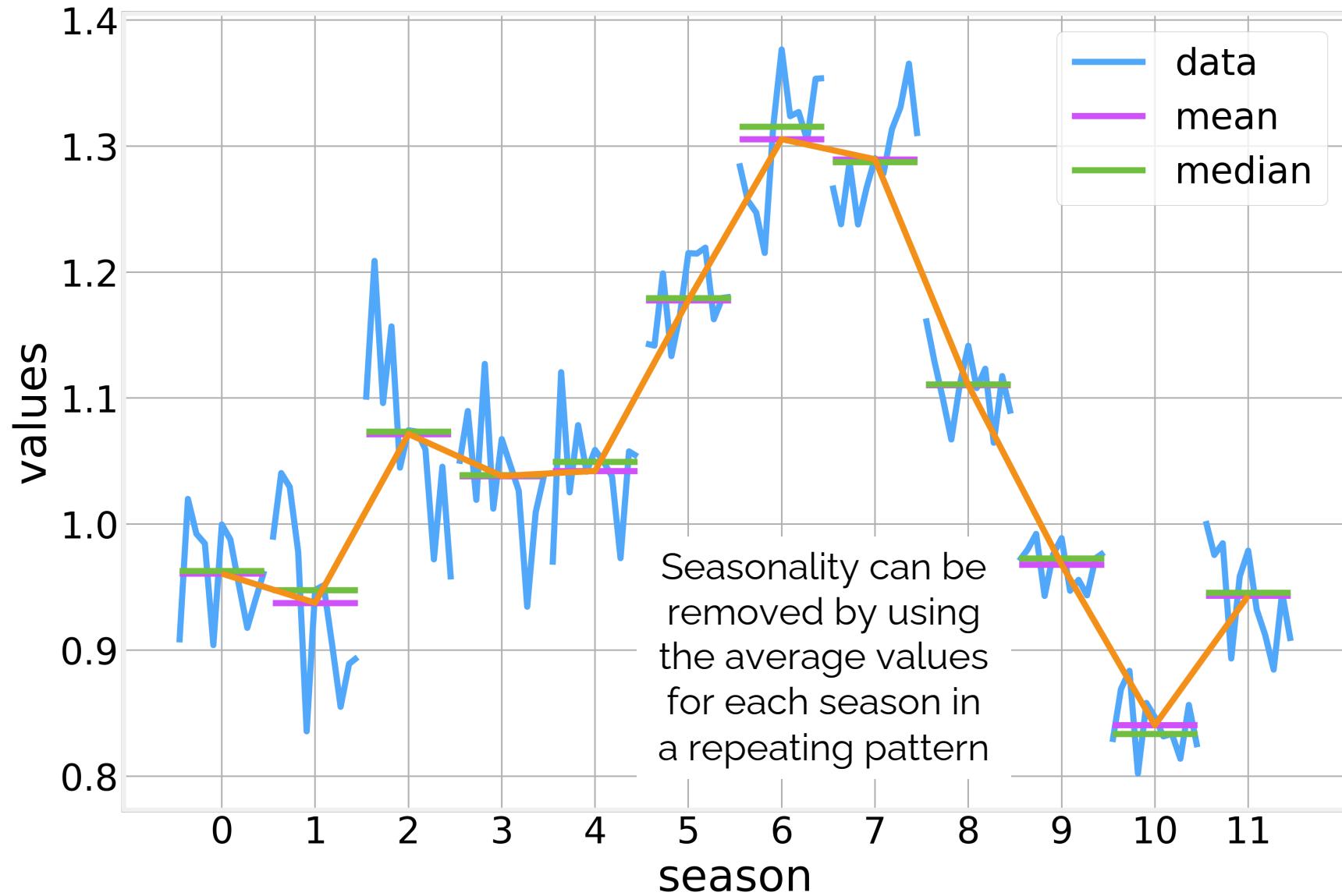
Seasonality



Seasonality



Seasonality



Time Series Decomposition

- A time series can be decomposed into **three components**:

Time Series Decomposition

- A time series can be decomposed into **three components**:
 - Trend, T_t
 - Seasonality, S_t
 - Residuals, R_t

Time Series Decomposition

- A time series can be decomposed into **three components**:
 - Trend, T_t
 - Seasonality, S_t
 - Residuals, R_t
- Decompositions can be

Time Series Decomposition

- A time series can be decomposed into **three components**:
 - Trend, T_t
 - Seasonality, S_t
 - Residuals, R_t
- Decompositions can be
 - additive - $x_t = T_t + S_t + R_t$
 - multiplicative - $x_t = Tt \cdot St \cdot Rt$

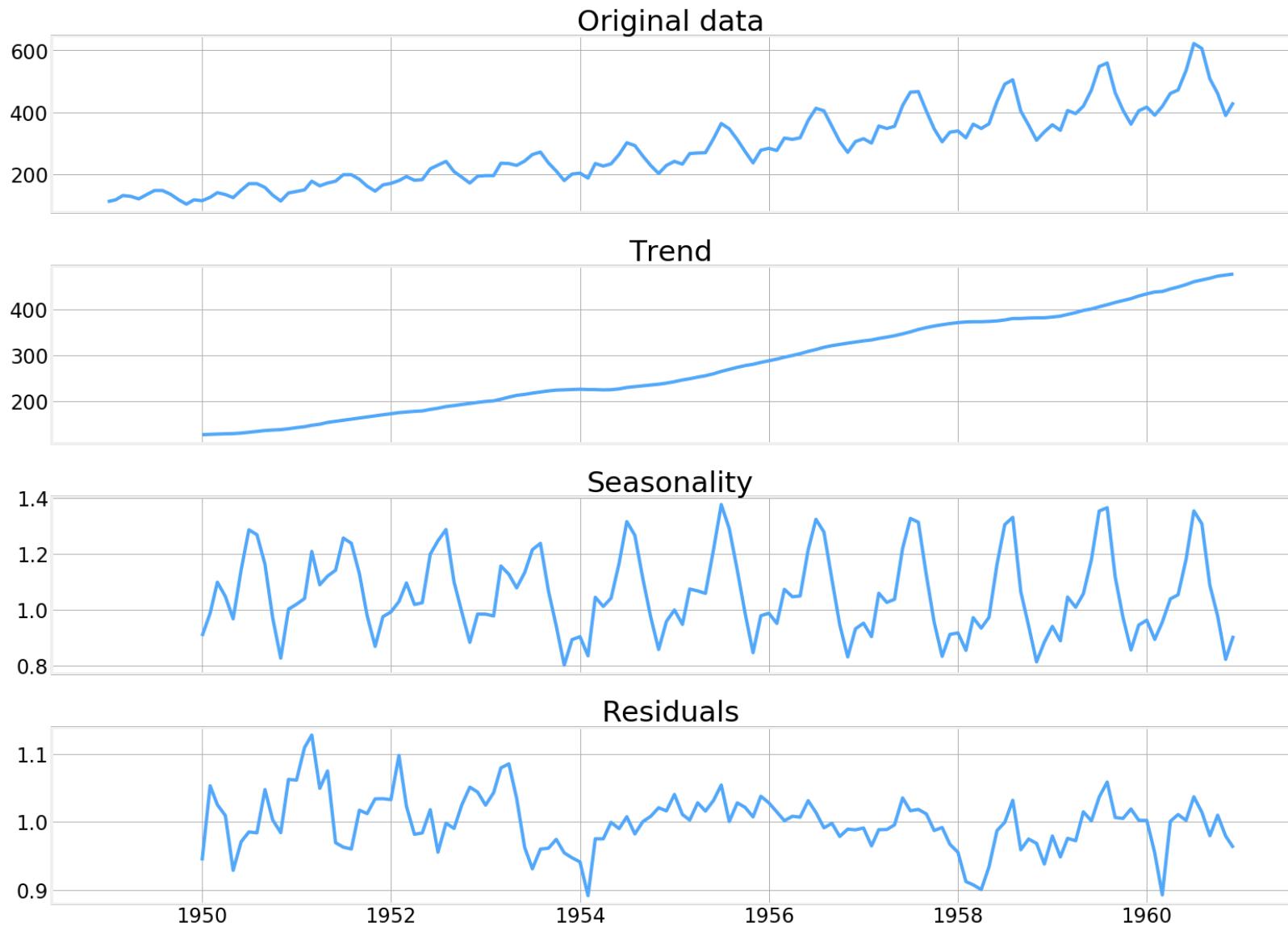
Time Series Decomposition

- A time series can be decomposed into **three components**:
 - Trend, T_t
 - Seasonality, S_t
 - Residuals, R_t
- Decompositions can be
 - additive - $x_t = T_t + S_t + R_t$
 - multiplicative - $x_t = Tt \cdot St \cdot Rt$
- The residuals are simply what is left of the original signal after we **remove the trend and the seasonality**

Time Series Decomposition

- A time series can be decomposed into **three components**:
 - Trend, T_t
 - Seasonality, S_t
 - Residuals, R_t
- Decompositions can be
 - additive - $x_t = T_t + S_t + R_t$
 - multiplicative - $x_t = Tt \cdot St \cdot Rt$
- The residuals are simply what is left of the original signal after we **remove the trend and the seasonality**
- Residuals are typically **stationary**

Time Series Decomposition



Lesson 2: Processing Timeseries Data

Lagging Values

- While analyzing time series, we often refer to values that our time series took **1, 2, 3**, etc., time steps in the past
- These are known as lagged values and denoted:

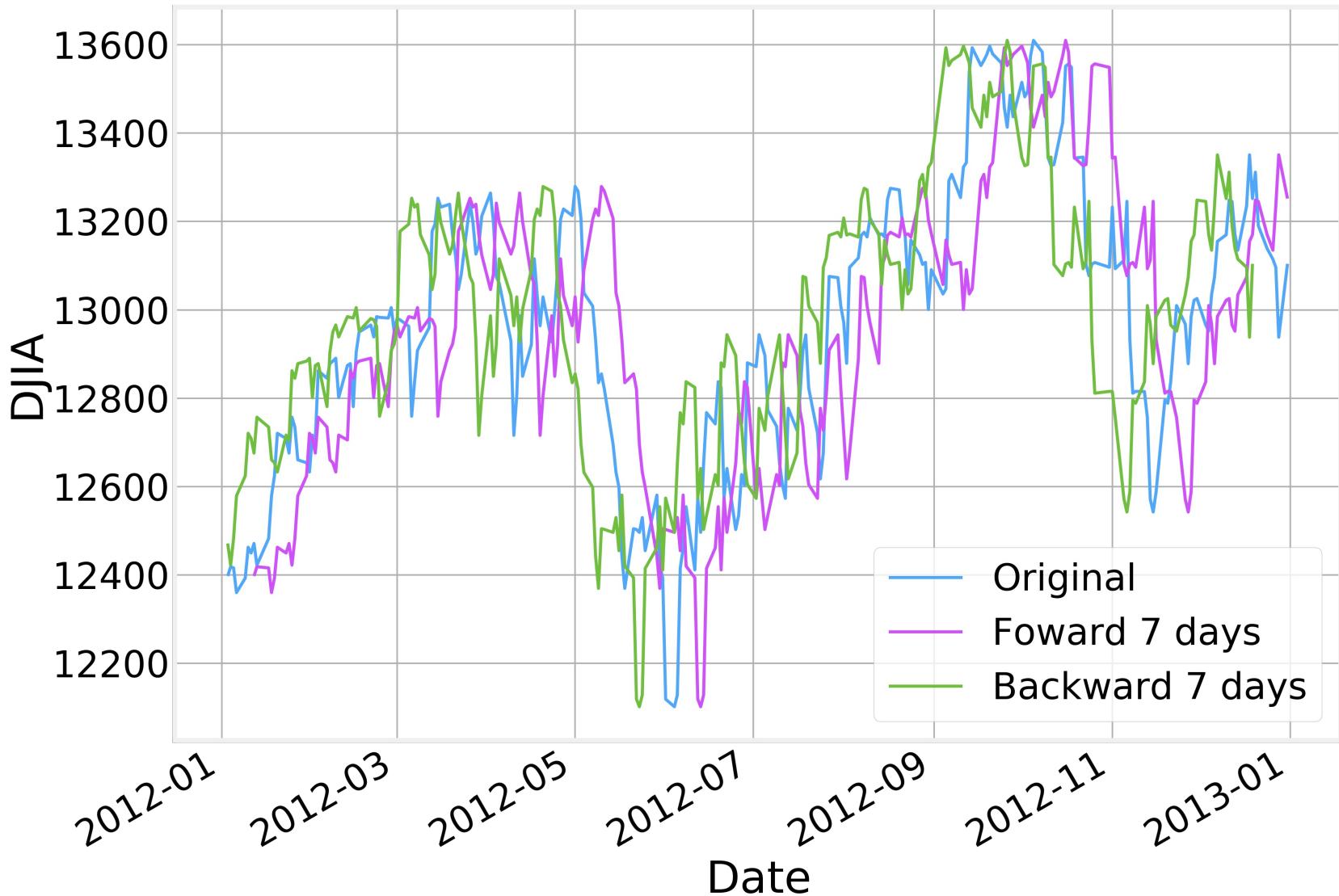
$$X_{t-l}$$

- where ***l*** is the value of the lag we are considering.

Lagging Values



Lagging Values



Differences

- Perhaps the most common use case for lagged values is for the calculation of **differences** of the form:

$$X_t - X_{t-l}$$

- Where $l \geq 1$ is the value of the lag we are interested in.
- Naturally, higher order differences can also be used, in which case, the difference of the difference is calculated:

$$y_t = X_t - X_{t-l}$$

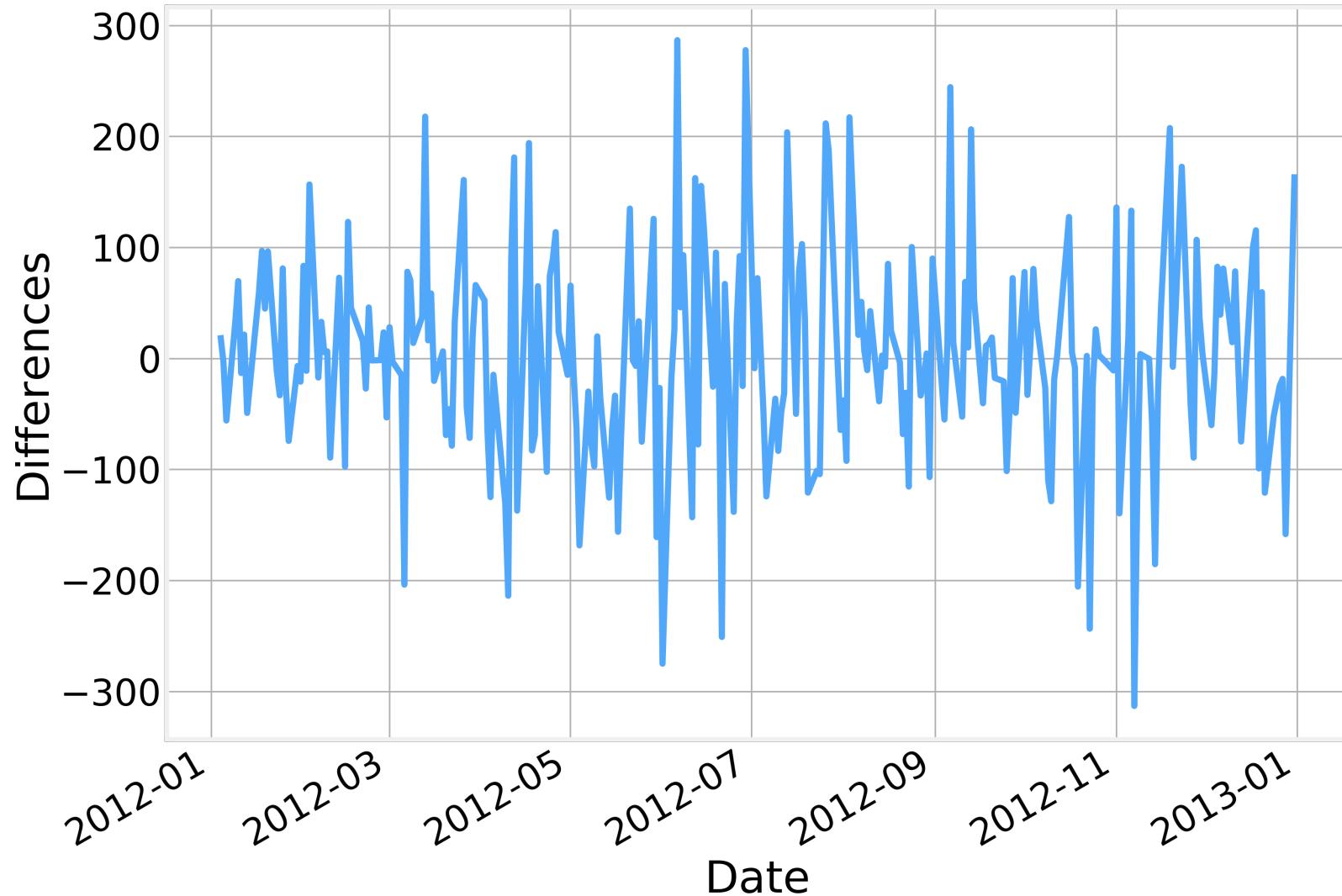
$$z_t = y_t - y_{t-l} \equiv X_t - 2X_{t-l} + X_{t-2l}$$

- This can be thought of as a discrete version of the usual derivative of a function.
- Differences are also a particularly simple way to **detrend** a time series

Differences



Differences



Windowing

- When analyzing the temporal behavior of a signal, we often need to evaluate if specific quantities are **time varying** or not

Windowing

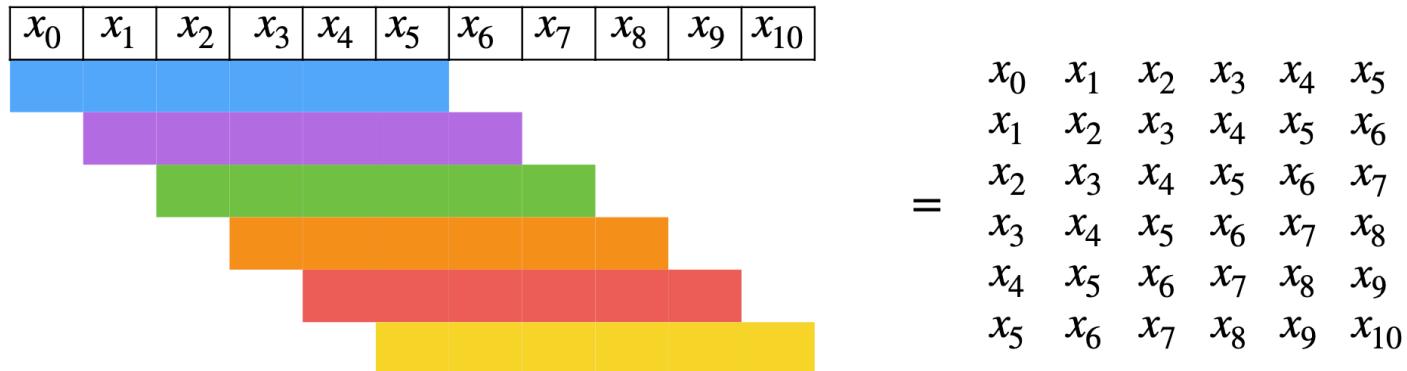
- When analyzing the temporal behavior of a signal, we often need to evaluate if specific quantities are **time varying** or not
- A common approach is to use **sliding windows** of a given length to evaluate the required values

Windowing

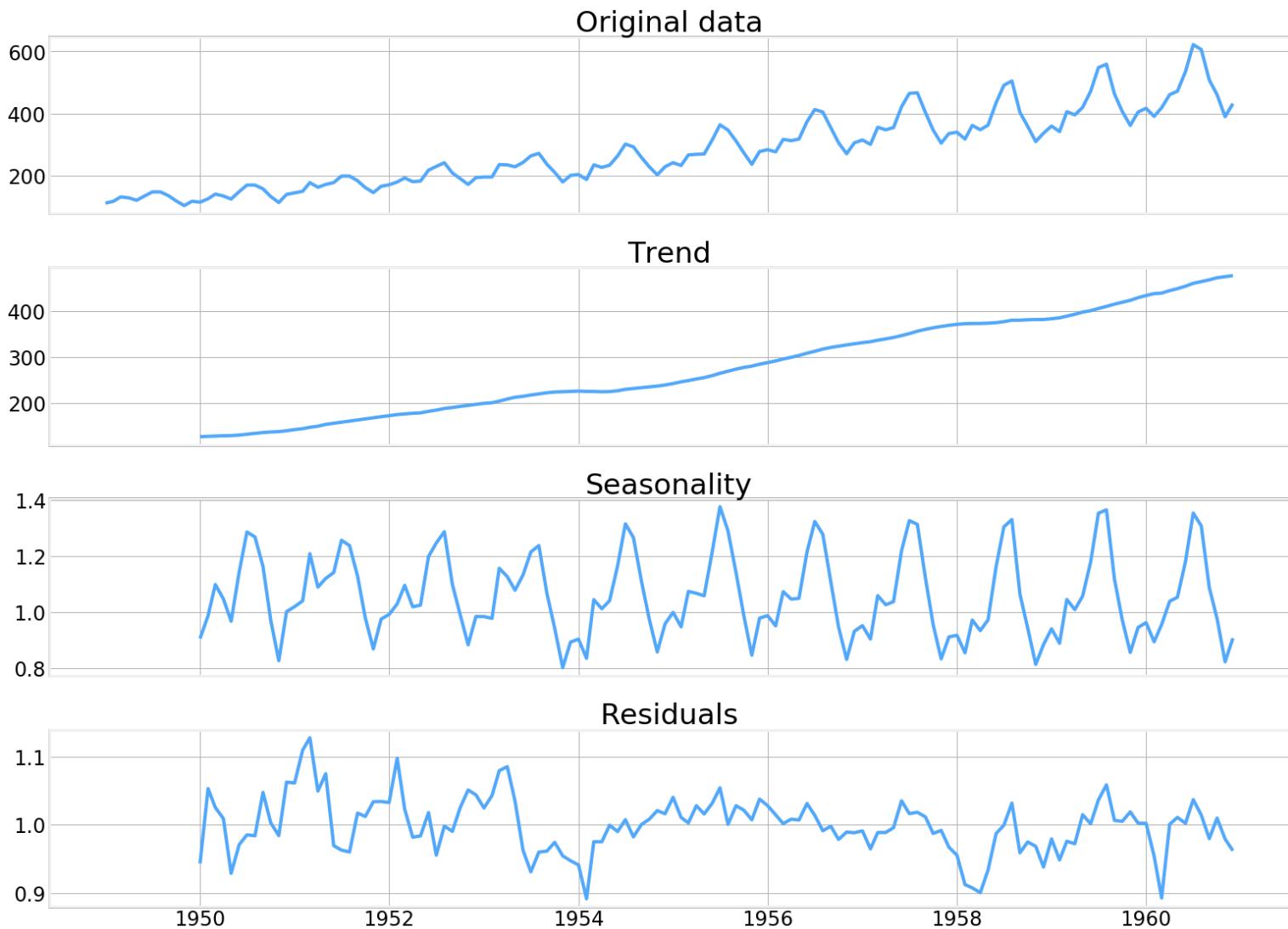
- When analyzing the temporal behavior of a signal, we often need to evaluate if specific quantities are **time varying** or not
- A common approach is to use **sliding windows** of a given length to evaluate the required values
- So, a sliding window of width **6** on a series of length **11** would look like:

Windowing

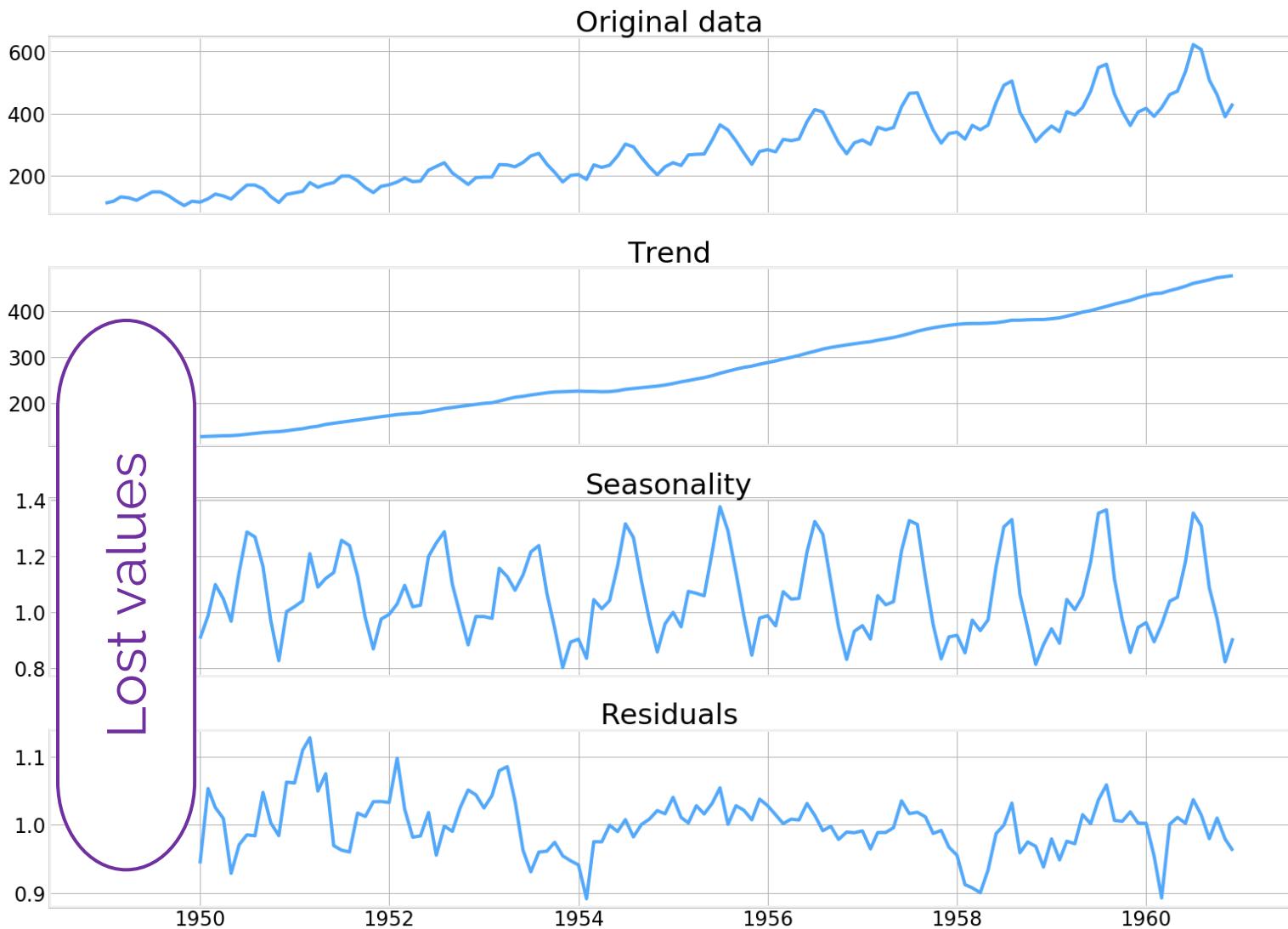
- When analyzing the temporal behavior of a signal, we often need to evaluate if specific quantities are **time varying** or not
- A common approach is to use **sliding windows** of a given length to evaluate the required values
- So, a sliding window of width **6** on a series of length **11** would look like:



Windowing

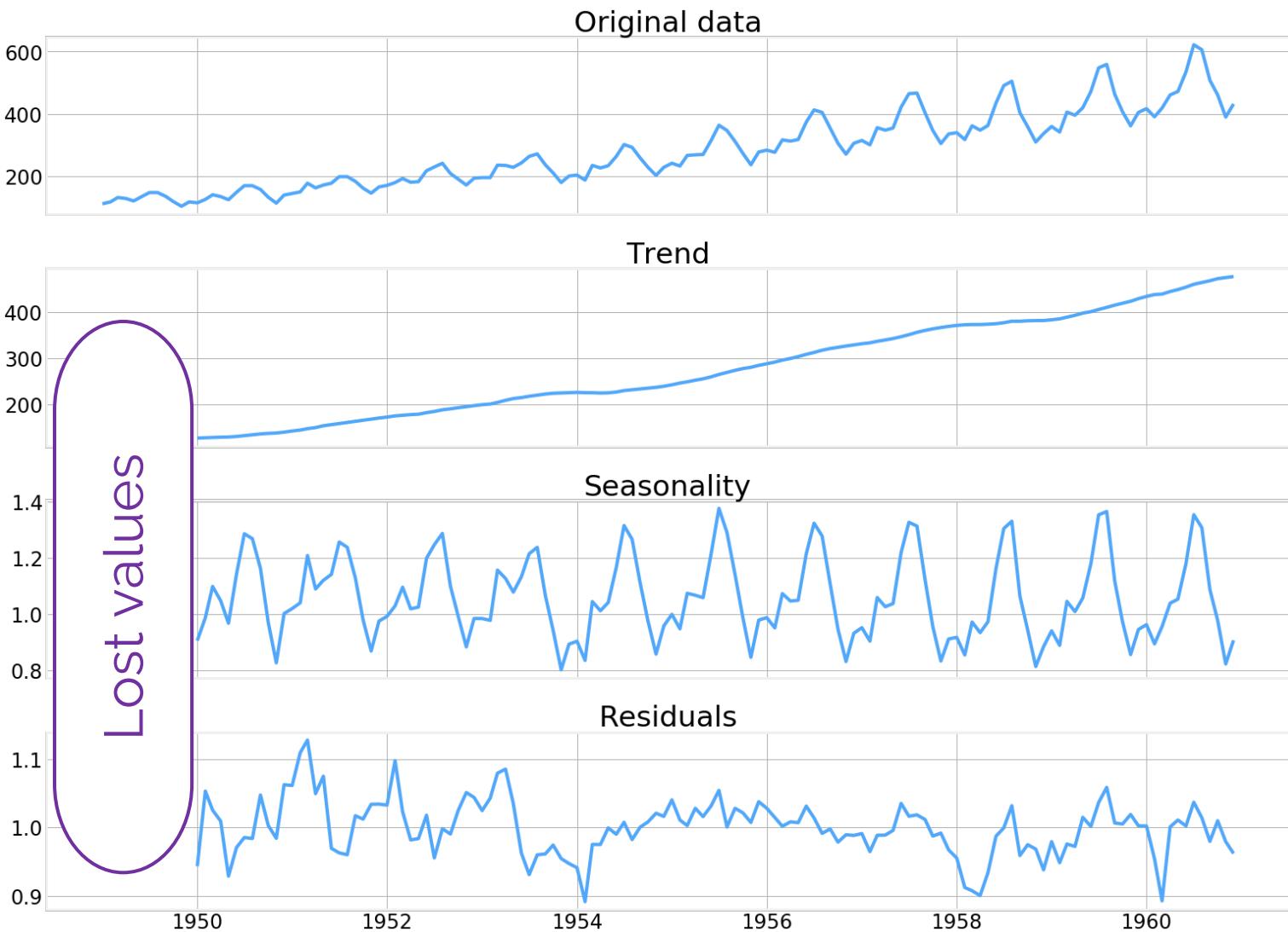


Windowing



Windowing

One common approach is to place all "lost values" at the beginning as it avoids "future leaking" when splitting the dataset



Running Values

- In the first part of the lecture, we already used running averages to detrend a time series

Running Values

- In the first part of the lecture, we already used running averages to **detrend** a time series
- Other common metrics are:
 - Variance
 - Maximum value
 - Minimum value
 - etc...

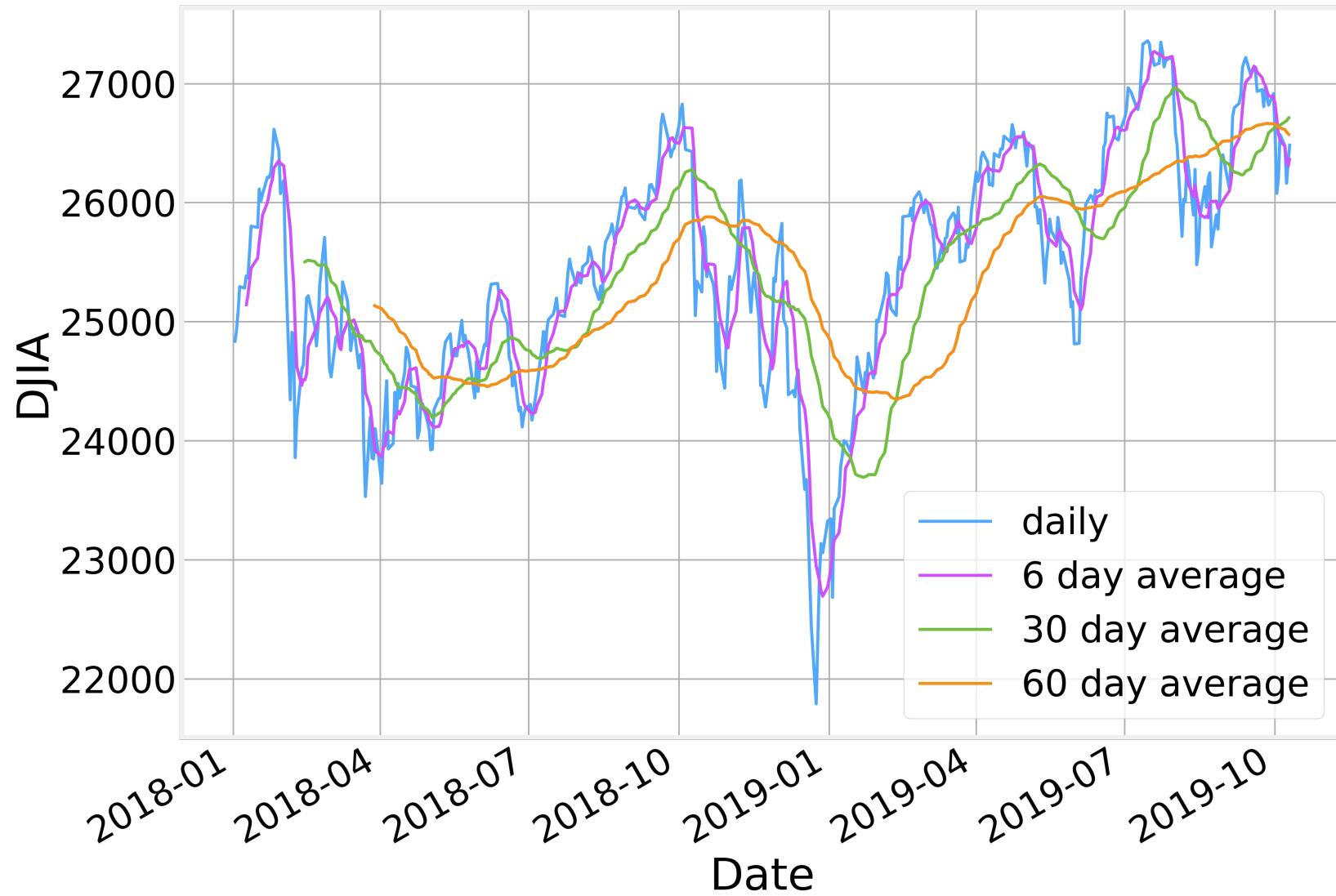
Running Values

- In the first part of the lecture, we already used running averages to **detrend** a time series
- Other common metrics are:
 - Variance
 - Maximum value
 - Minimum value
 - etc...
- One important detail to note is that while using windowing to calculate running values **we "lose"** a number of points equal to the width of the window

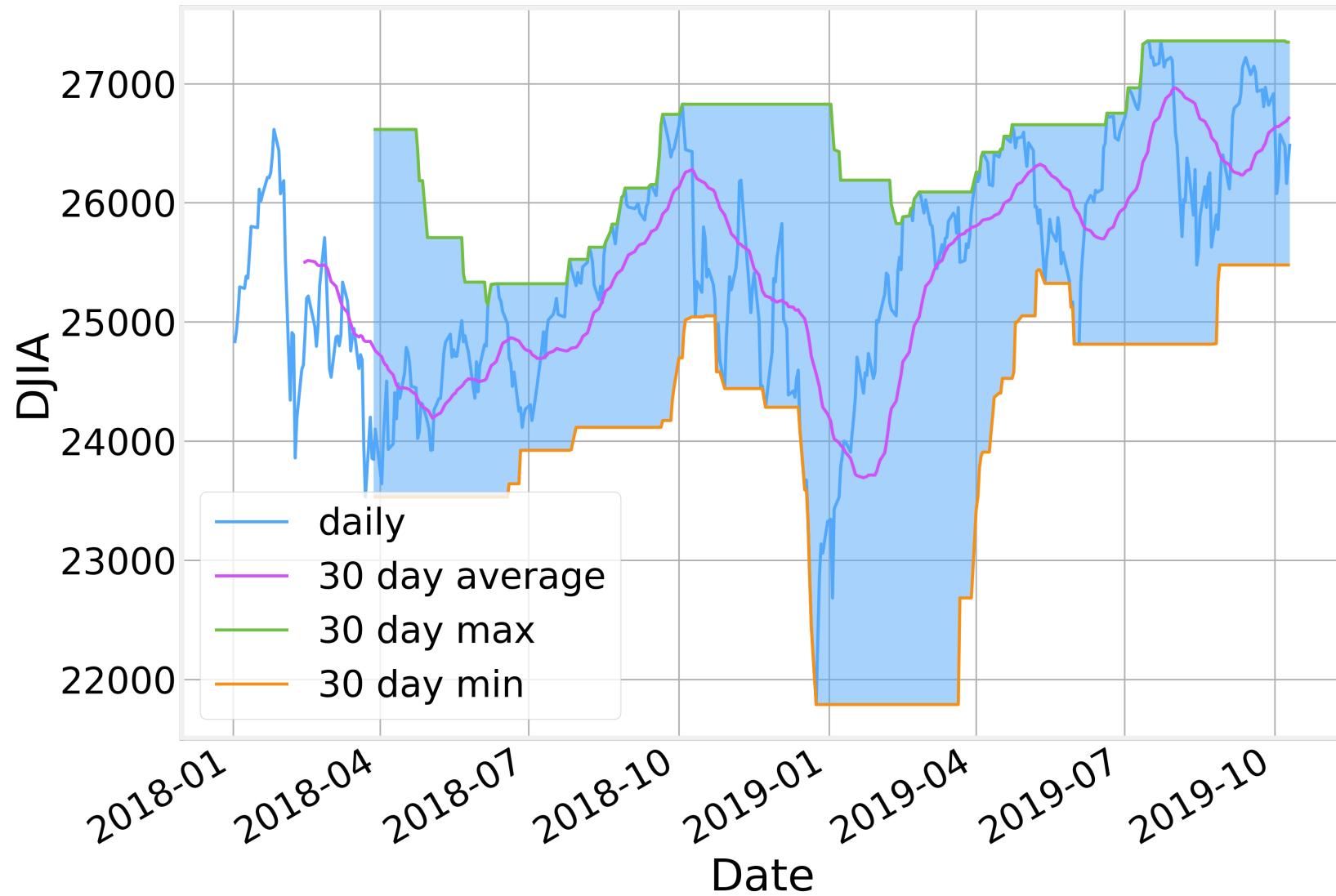
Running Values

- In the first part of the lecture, we already used running averages to **detrend** a time series
- Other common metrics are:
 - Variance
 - Maximum value
 - Minimum value
 - etc...
- One important detail to note is that while using windowing to calculate running values **we "lose"** a number of points equal to the width of the window
- Depending on the application we can choose to place the missing values in either or (or even both) extremes of the time interval

Running Values



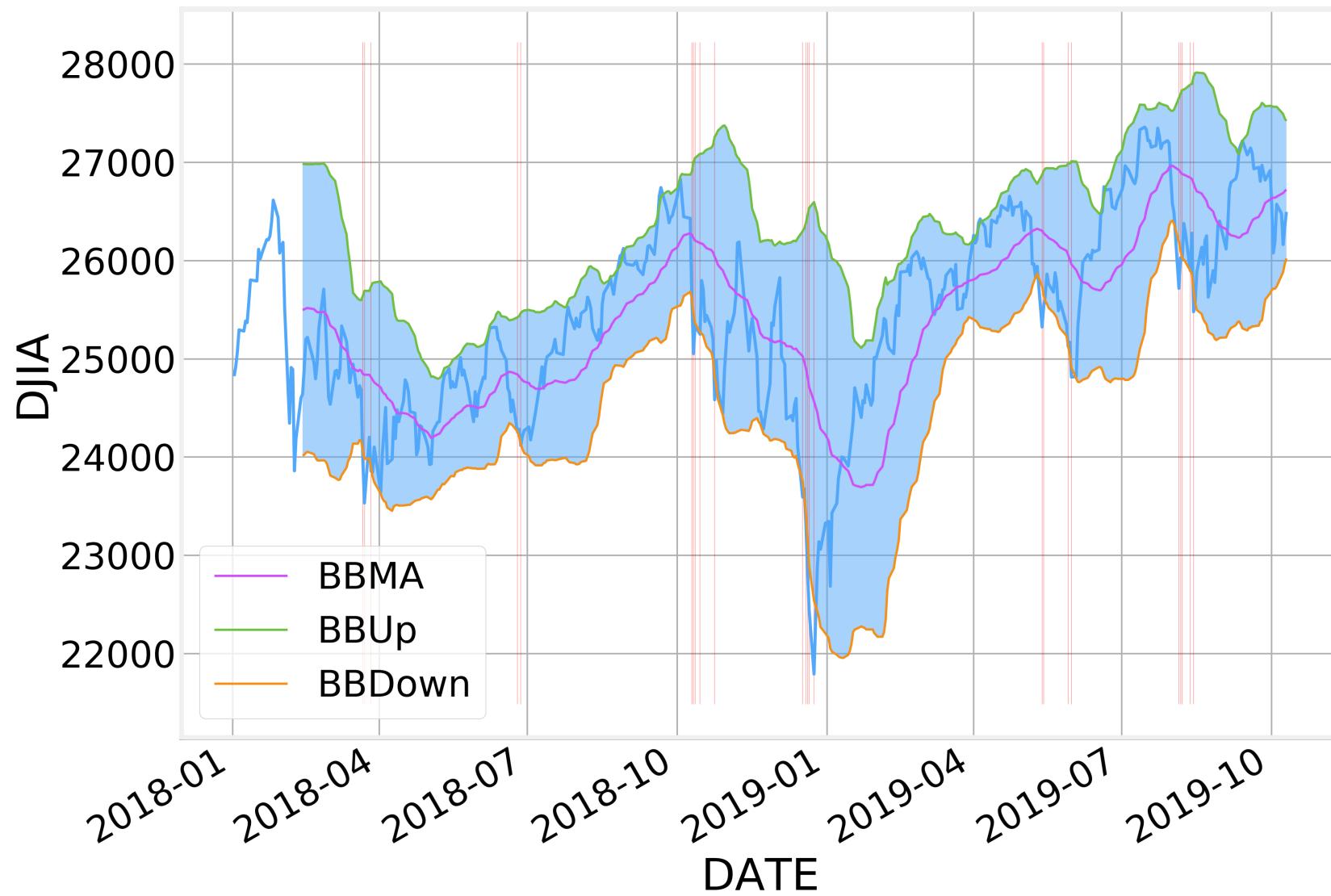
Envelopes



Bollinger Bands

- A common use for application for running values is the calculation of **Bollinger Bands**.
- Introduced by [John Bollinger](#) in the 1980s as a complement to more traditional time series technical analysis techniques.
- **Bollinger Bands** are defined by two components:
 - A N period moving average, μ_N
 - The area K standard deviations above and below the moving average $\mu_N \pm K\sigma_N$
- Both μ_N and σ_N are computed on a [running window](#) of size N
- The values N and K are application specific. For stock trading, $N=20$ and $K=2$
- Whenever the time series steps out of the Bollinger Band that's a clear indication of a **change in the temporal behavior**.

Bollinger Bands



Exponential Running Average

- One alternative to a simple running average is Exponential Smoothing

Exponential Running Average

- One alternative to a simple running average is Exponential Smoothing
- The exponentially “smooth” version of a time series is given by:

$$z_t = \alpha x_t + (1 - \alpha)z_{t-1}$$

Exponential Running Average

- One alternative to a simple running average is **Exponential Smoothing**
- The exponentially “smooth” version of a time series is given by:

$$z_t = \alpha x_t + (1 - \alpha)z_{t-1}$$

- The smaller the value of the weight α , the less influence each point has on the transformed time series.

Exponential Running Average

- One alternative to a simple running average is Exponential Smoothing
- The exponentially “smooth” version of a time series is given by:

$$z_t = \alpha x_t + (1 - \alpha)z_{t-1}$$

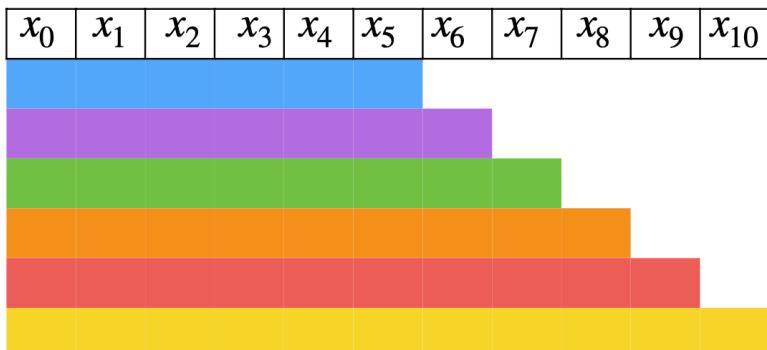
- The smaller the value of the weight α , the less influence each point has on the transformed time series.
- Each point depends implicitly on all previous points

Exponential Running Average

- One alternative to a simple running average is **Exponential Smoothing**
- The exponentially “smooth” version of a time series is given by:

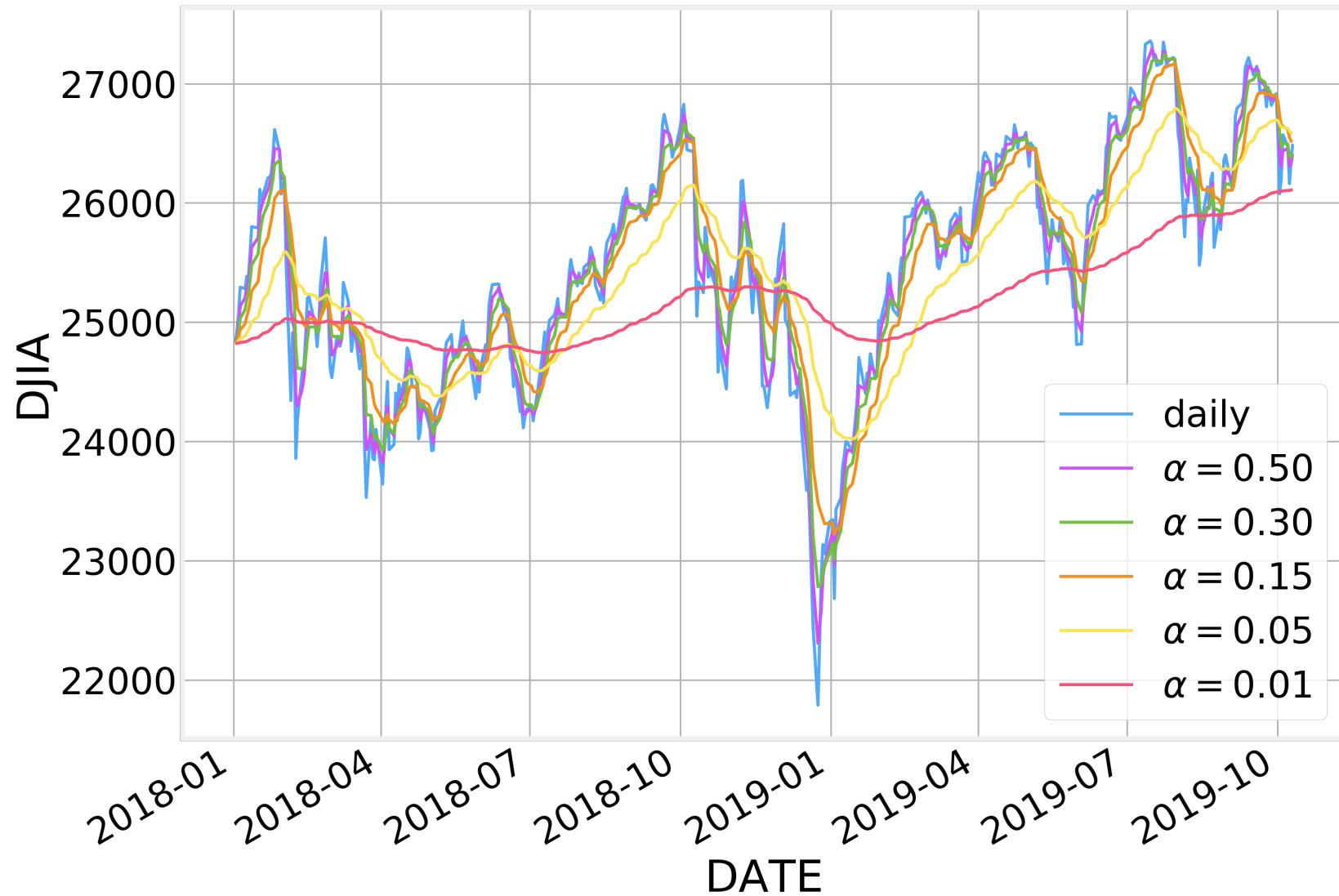
$$z_t = \alpha x_t + (1 - \alpha)z_{t-1}$$

- The smaller the value of the weight α , the less influence each point has on the transformed time series.
- Each point depends implicitly on **all previous points**

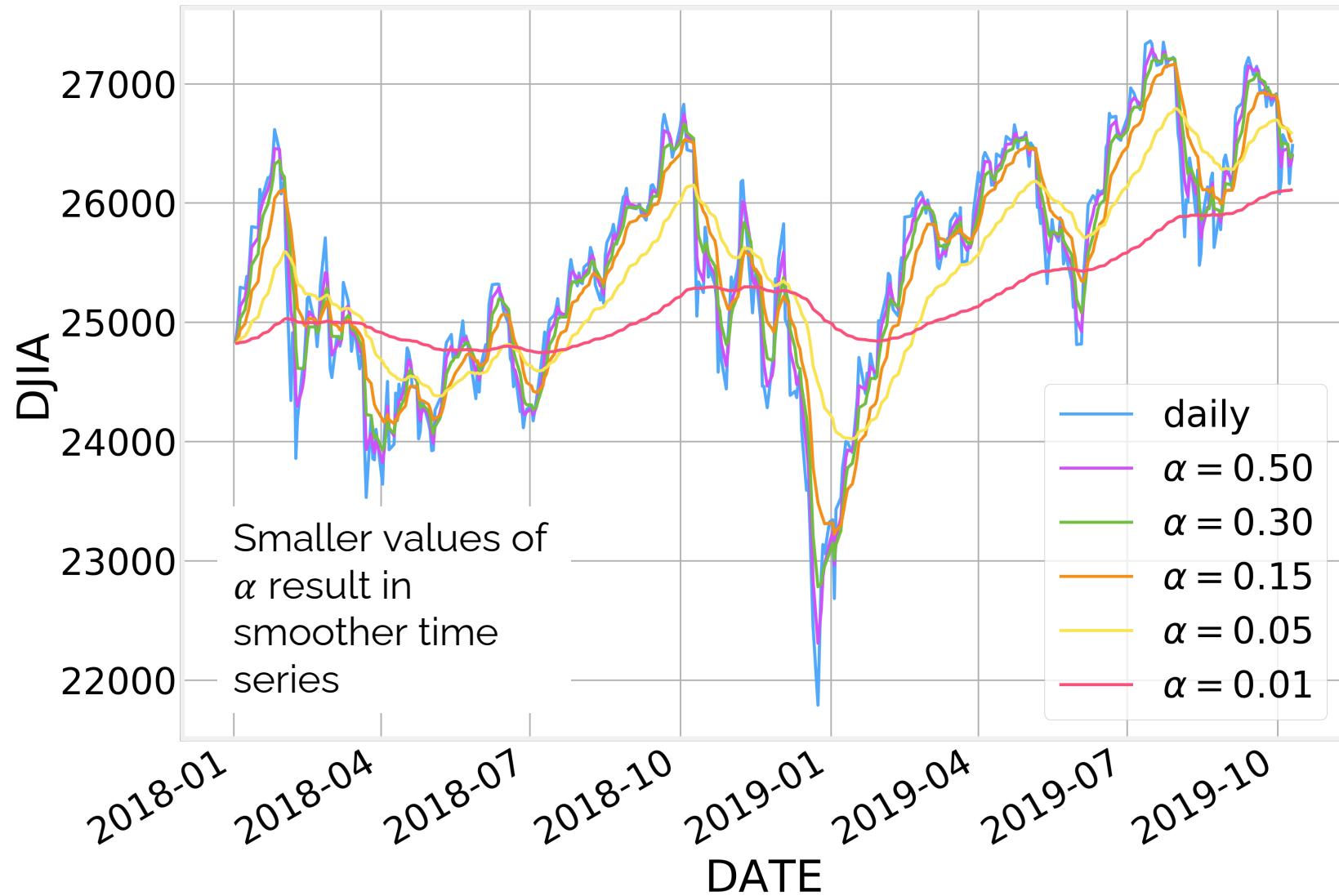


$$\begin{pmatrix} \alpha & & & & & & & & & \\ \alpha(1-\alpha)^1 & \alpha & & & & & & & & \\ \alpha(1-\alpha)^2 & & \alpha(1-\alpha)^1 & & & & & & & \\ \vdots & & \vdots & & & & & & & \ddots \\ \alpha(1-\alpha)^{n-1} & \alpha(1-\alpha)^{n-2} & \alpha(1-\alpha)^{n-3} & & & & & & & \alpha \end{pmatrix}$$

Exponential Running Average



Exponential Running Average



Forecasting

- We can also use the Exponential Moving Averages as a simple forecasting tool.

Forecasting

- We can also use the Exponential Moving Averages as a simple forecasting tool.
- The value at time $t + 1$ is given by:

$$z_{t+1} = \alpha x_t + (1 - \alpha)z_{t-1}$$

Forecasting

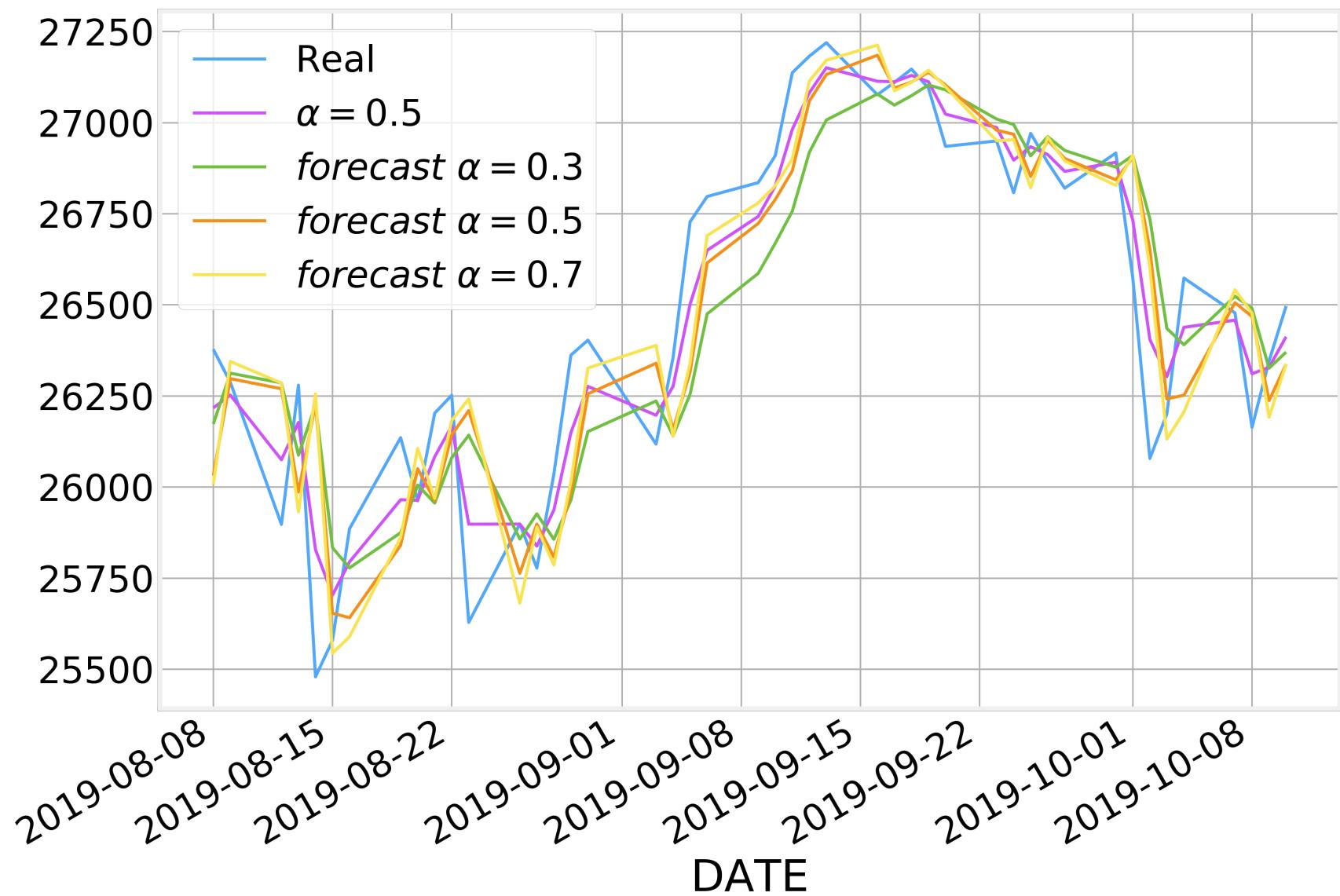
- We can also use the Exponential Moving Averages as a simple forecasting tool.
- The value at time $t + 1$ is given by:

$$z_{t+1} = \alpha x_t + (1 - \alpha)z_{t-1}$$

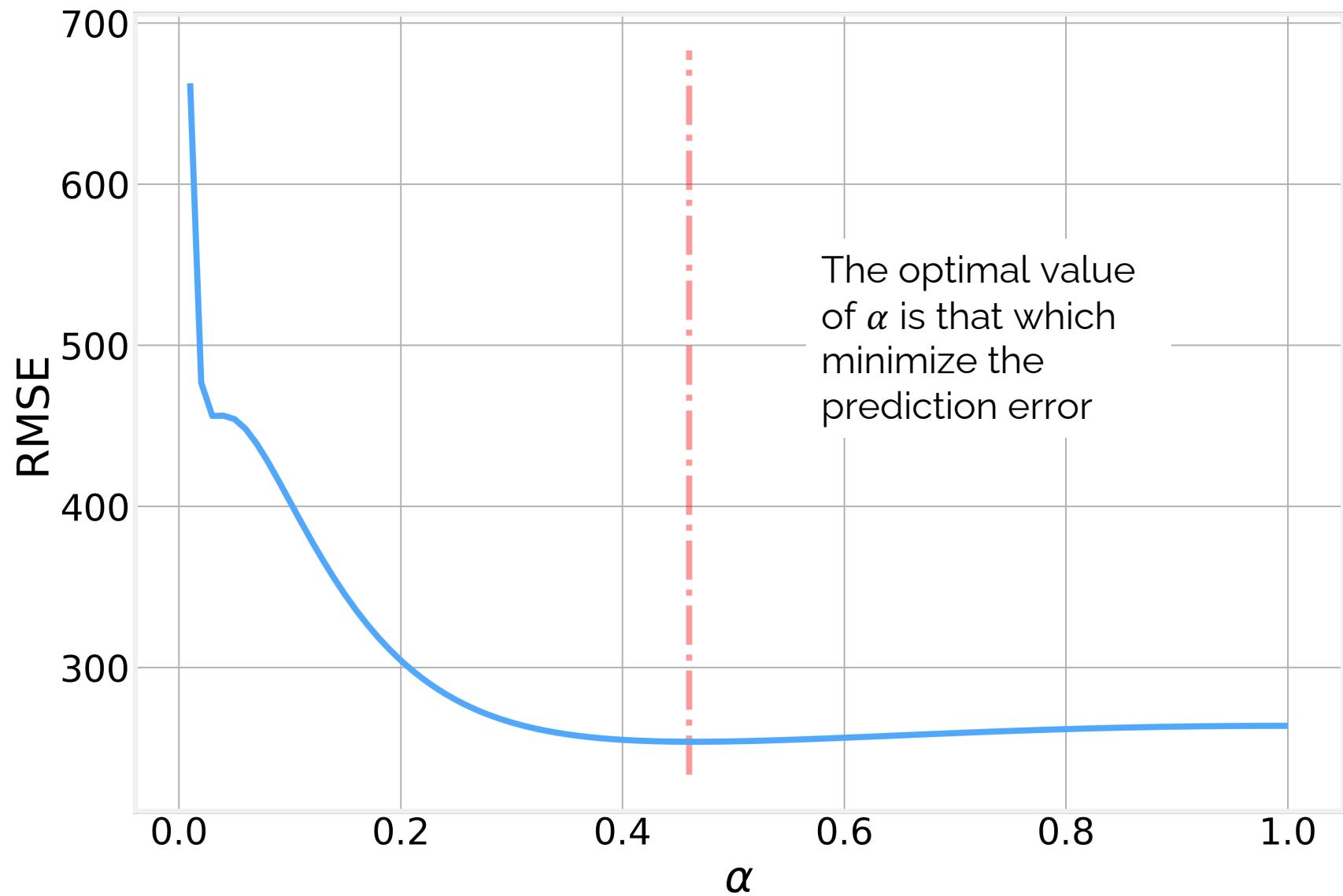
- Which we can consider to be a prediction on the value of x_{t+1} , based on the current value of z_t and some factor of our current error value $x_t - z_t$:

$$z_{t+1} = z_t + \alpha(x_t - z_t)$$

Forecasting

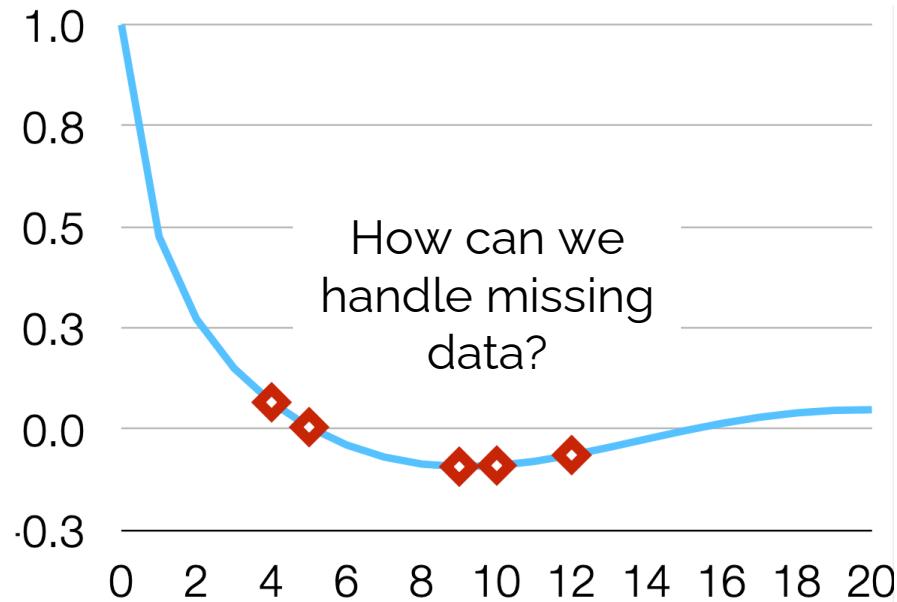


Forecasting



Fill Methods

- Sometimes the time series is incomplete



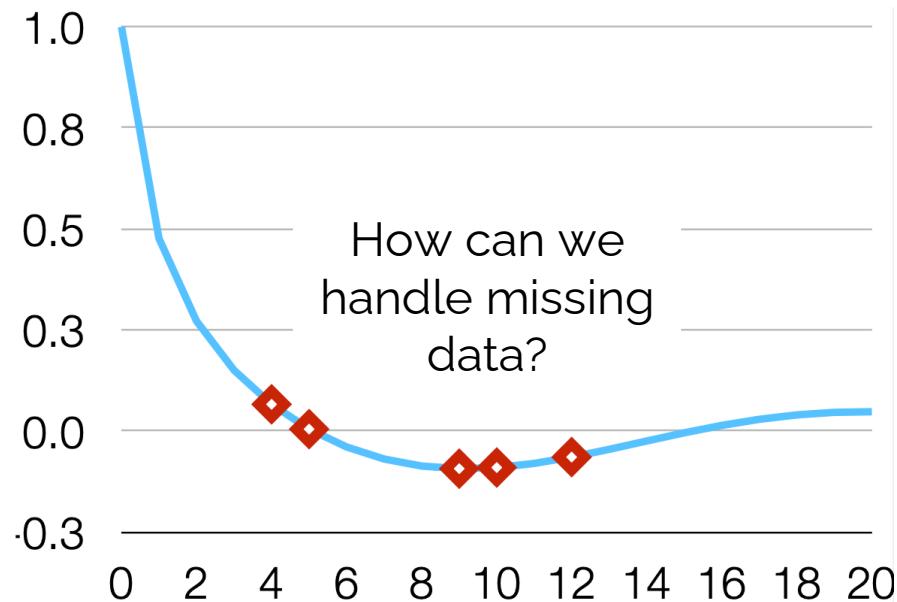
Fill Methods

- Sometimes the time series is **incomplete**
- Missing data points can be due to data corruption, data collection issues, etc.



Fill Methods

- Sometimes the time series is **incomplete**
- Missing data points can be due to data corruption, data collection issues, etc.
- Missing values are represented as **nan**



Fill Methods

- Sometimes the time series is **incomplete**
- Missing data points can be due to data corruption, data collection issues, etc.
- Missing values are represented as **nan**
- Several techniques have been developed to handle this case



Fill Methods

- Sometimes the time series is **incomplete**
- Missing data points can be due to data corruption, data collection issues, etc.
- Missing values are represented as **nan**
- Several techniques have been developed to handle this case
 - **forward fill** – keep the last valid value



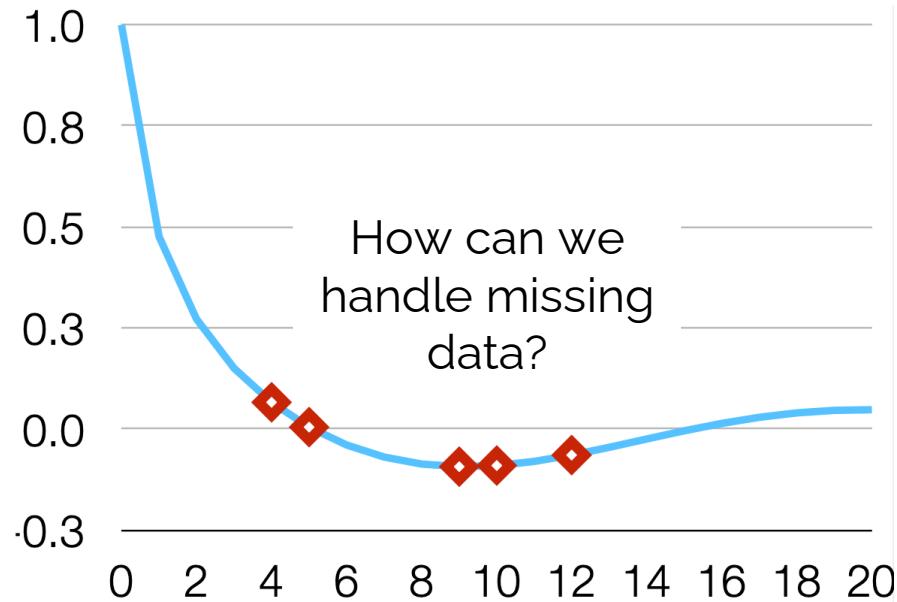
Fill Methods

- Sometimes the time series is **incomplete**
- Missing data points can be due to data corruption, data collection issues, etc.
- Missing values are represented as **nan**
- Several techniques have been developed to handle this case
 - **forward fill** – keep the last valid value
 - **back fill** – keep the next valid value



Fill Methods

- Sometimes the time series is **incomplete**
- Missing data points can be due to data corruption, data collection issues, etc.
- Missing values are represented as **nan**
- Several techniques have been developed to handle this case
 - **forward fill** – keep the last valid value
 - **back fill** – keep the next valid value
 - **interpolate** – add values by interpolating between the previous and the next value

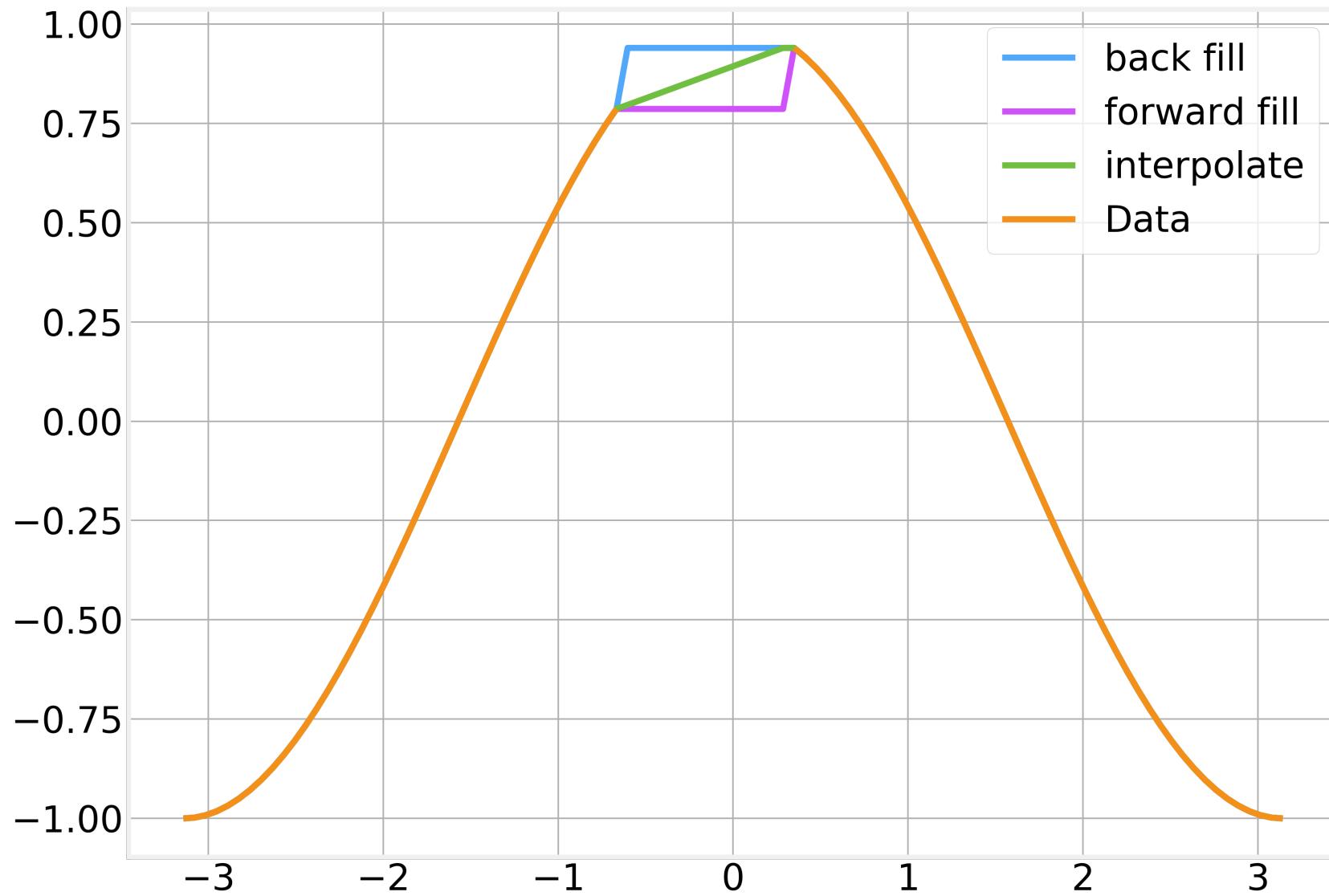


Fill Methods

- Sometimes the time series is **incomplete**
- Missing data points can be due to data corruption, data collection issues, etc.
- Missing values are represented as **nan**
- Several techniques have been developed to handle this case
 - **forward fill** – keep the last valid value
 - **back fill** – keep the next valid value
 - **interpolate** – add values by interpolating between the previous and the next value
 - **imputation** – add values based on what we expect the missing values to be



Fill Methods



Resampling

- Time series typically have an **intrinsic time scale** at which the data was collected: ticks, seconds, days, months, etc.

Resampling

- Time series typically have an **intrinsic time scale** at which the data was collected: ticks, seconds, days, months, etc.
- In many cases, our analysis requires that we **resample** the data to a different time scale

Resampling

- Time series typically have an **intrinsic time scale** at which the data was collected: ticks, seconds, days, months, etc.
- In many cases, our analysis requires that we **resample** the data to a different time scale
- Resampling to a longer timescale is relatively simple and similar to aggregation:

Resampling

- Time series typically have an **intrinsic time scale** at which the data was collected: ticks, seconds, days, months, etc.
- In many cases, our analysis requires that we **resample** the data to a different time scale
- Resampling to a longer timescale is relatively simple and similar to aggregation:
 - Transforming from daily to weekly frequency requires simply aggregating by week

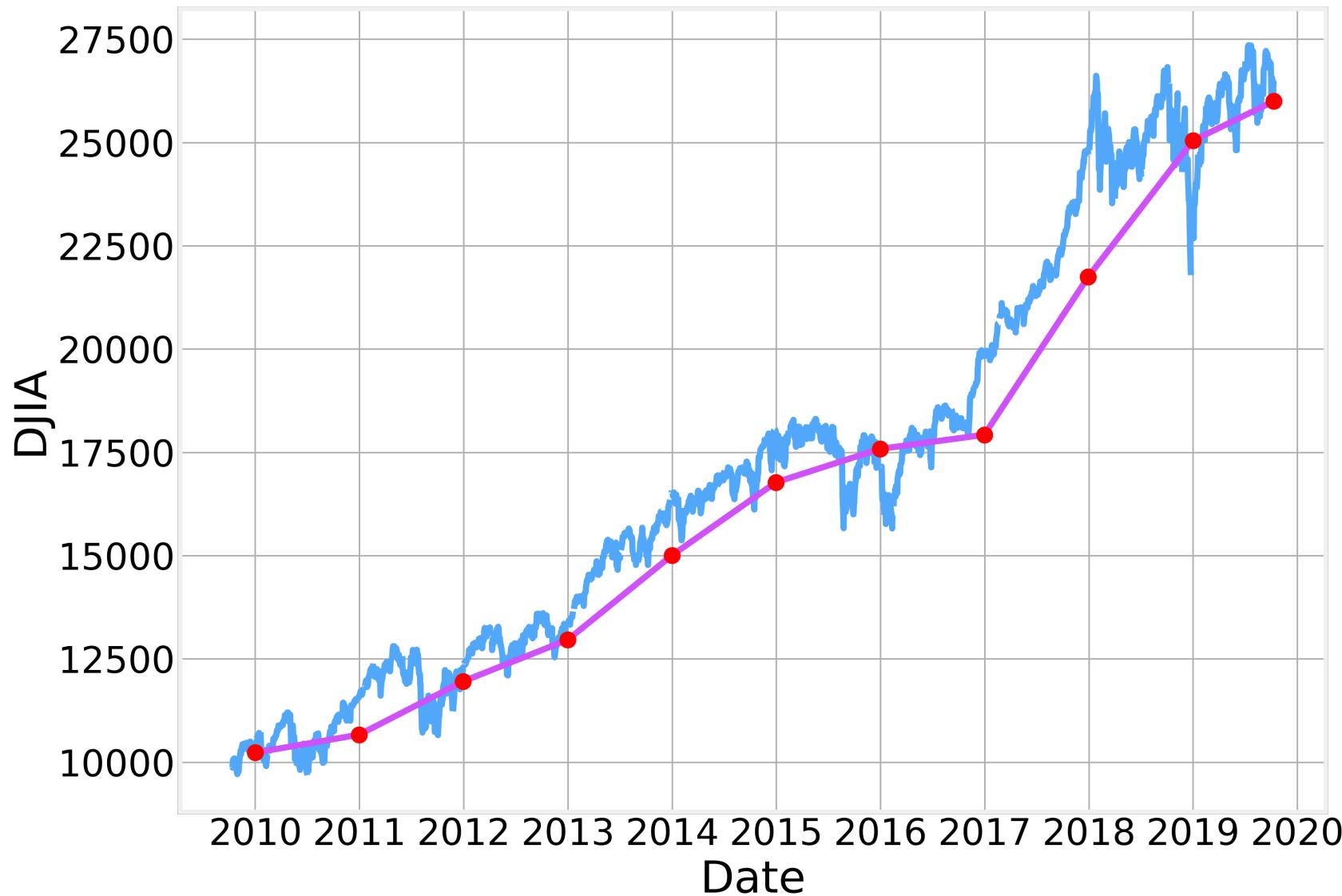
Resampling

- Time series typically have an **intrinsic time scale** at which the data was collected: ticks, seconds, days, months, etc.
- In many cases, our analysis requires that we **resample** the data to a different time scale
- Resampling to a longer timescale is relatively simple and similar to aggregation:
 - Transforming from daily to weekly frequency requires simply aggregating by week
- Resampling to shorter timescales requires **interpolation** or **imputation** to make up for the missing values

Resampling

- Time series typically have an **intrinsic time scale** at which the data was collected: ticks, seconds, days, months, etc.
- In many cases, our analysis requires that we **resample** the data to a different time scale
- Resampling to a longer timescale is relatively simple and similar to aggregation:
 - Transforming from daily to weekly frequency requires simply aggregating by week
- Resampling to shorter timescales requires **interpolation** or **imputation** to make up for the missing values
 - Going from weekly to daily frequency requires specifying how to allocate the values for each day of the week

Resampling



Bootstrapping

- Bootstrapping is another alternative to estimate statistical properties such as mean and variance

Bootstrapping

- Bootstrapping is another alternative to estimate statistical properties such as mean and variance
- Bootstrapping measures the desired property in a large number of samples (with replacement), of the observed dataset.

Bootstrapping

- Bootstrapping is another alternative to estimate statistical properties such as mean and variance
- Bootstrapping measures the desired property in a large number of samples (with replacement), of the observed dataset.
- Each sample has equal size to the observed dataset.

Bootstrapping

- Bootstrapping is another alternative to estimate statistical properties such as mean and variance
- Bootstrapping measures the desired property in a large number of samples (with replacement), of the observed dataset.
- Each sample has equal size to the observed dataset.
- From the entire population of samples, the empirical bootstrap distribution of the expected values can be obtained to provide information about the distribution in the total population

Bootstrapping

- Bootstrapping is another alternative to estimate statistical properties such as mean and variance
- Bootstrapping measures the desired property in a large number of samples (with replacement), of the observed dataset.
- Each sample has equal size to the observed dataset.
- From the entire population of samples, the empirical bootstrap distribution of the expected values can be obtained to provide information about the distribution in the total population

Also see [Jackknife Estimation](#)

Bootstrapping

- Bootstrapping is another alternative to estimate statistical properties such as mean and variance
- Bootstrapping measures the desired property in a large number of samples (with replacement), of the observed dataset.
- Each sample has equal size to the observed dataset.
- From the entire population of samples, the empirical bootstrap distribution of the expected values can be obtained to provide information about the distribution in the total population

Also see [Jackknife Estimation](#)

Bootstrapping

