Time-series Analytics

Giacomo Ziffer

giacomo.ziffer@polimi.it

POLITECNICO
MILANO 1863

A few words about me

Giacomo Ziffer, Dr

PhD Student @ Politecnico di Milano
CEO & Co-founder @ motus ml

Researcher in the **Time-Evolving Analytics** field, focusing on applying Streaming Machine Learning techniques to (un)structured data streams with **concept drifts** and **temporal dependence**

Introduction

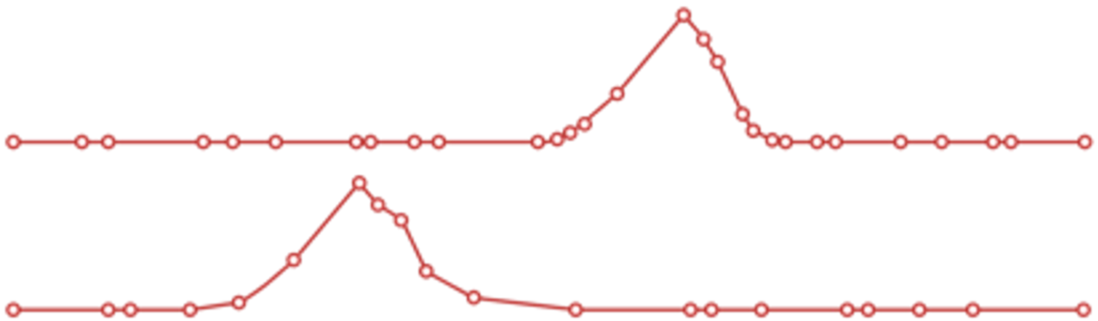
Type of data

A ***time series*** is a sequence of observations on **one** (or more) **quantitative** variable **regularly** collected **over time**.

Events vs time series

The phenomenon happens and we observe them

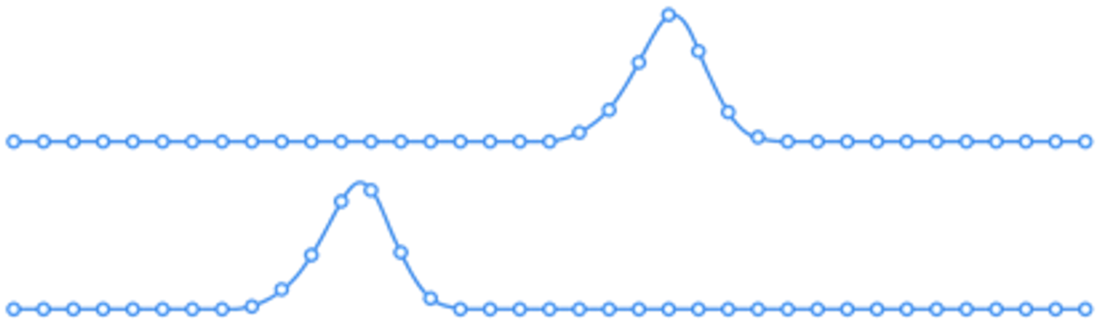
irregularly



Events

We monitor a phenomenon

regularly

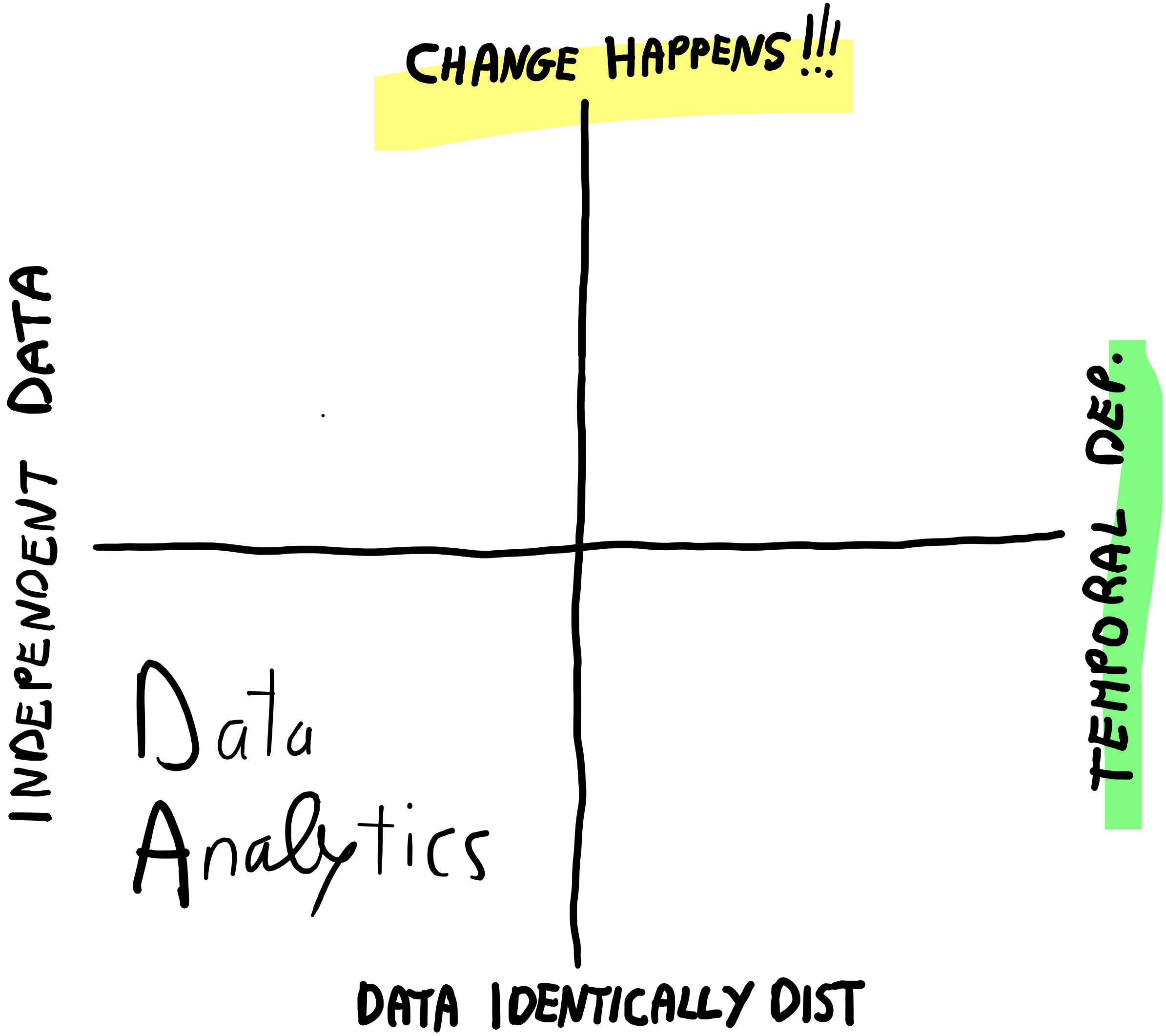


Time series

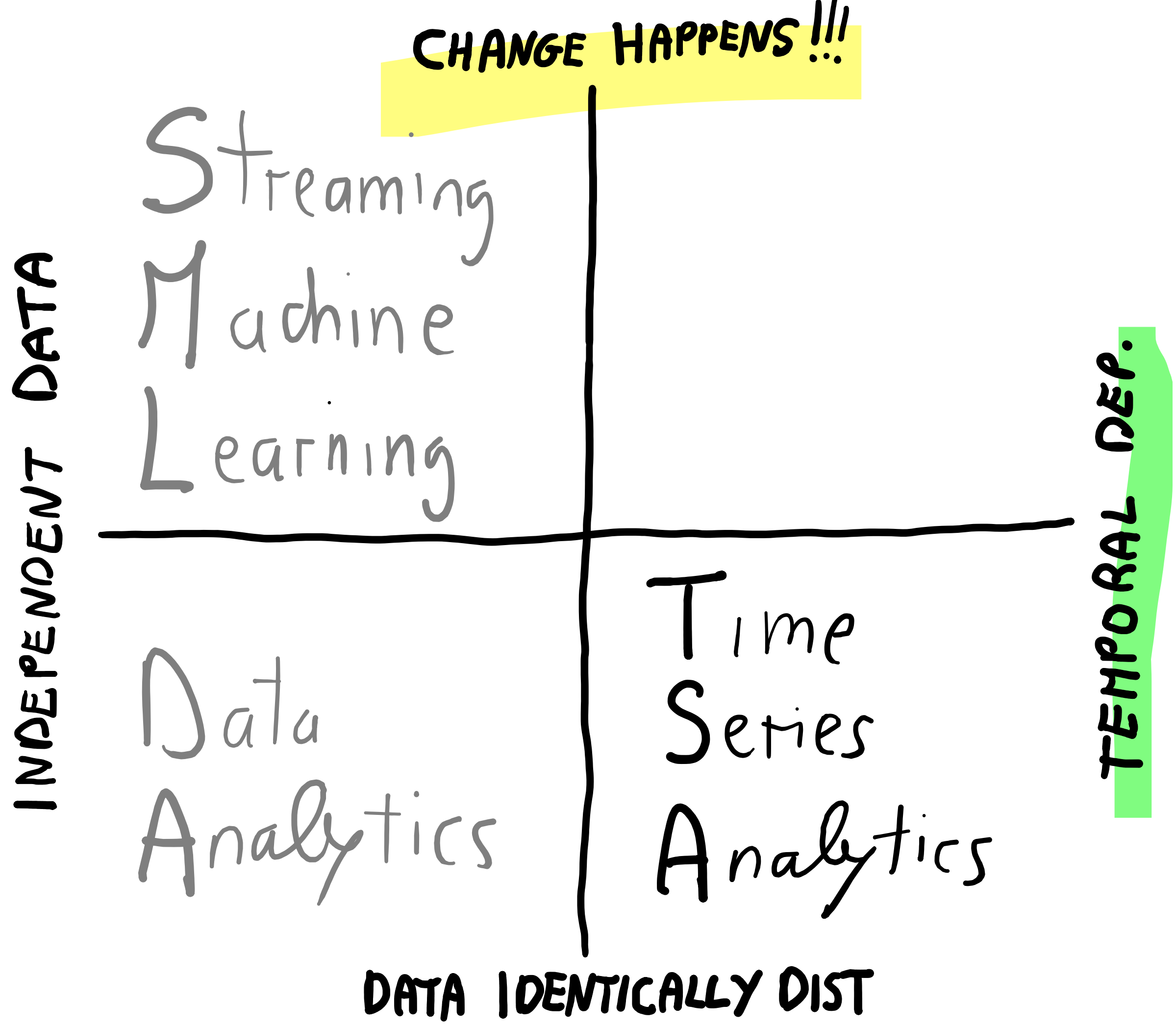
Time Series (brief recall from early lectures)

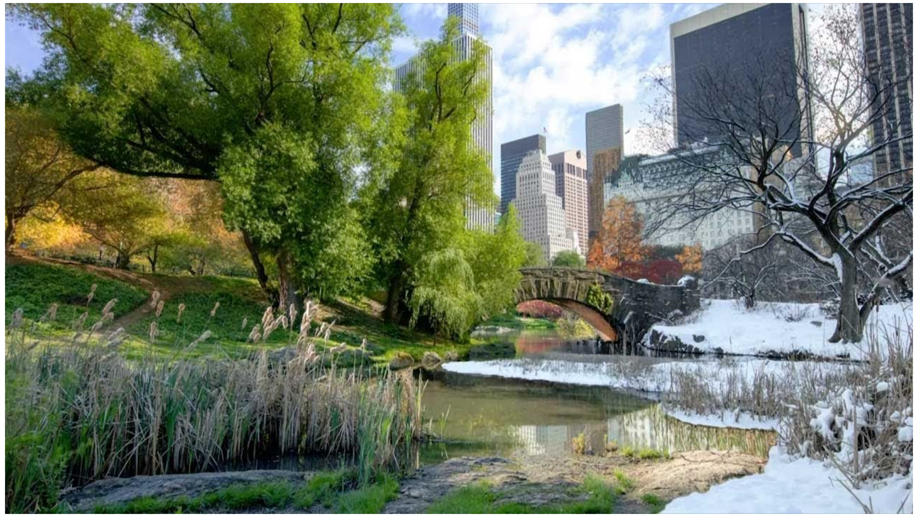
- A *time-series* is a set of observations on a quantitative variable collected over time.
- Examples
  - Dow Jones Industrial Averages
  - Historical data on sales, inventory, customer counts, interest rates, costs, etc.
  - Signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, communications engineering, …
- Businesses are often very interested in analyzing and forecasting time series variables

State-of-the-art



State-of-the-art





Explaining the past and forecast the future of a continuous flow of data without assuming data independence with

Time Series Analytics

1st foundational concept

Stationarity

Fact

If a time series is stationary,… *… it is predictable*

Photo by **cottonbro** from **Pexels**

Photo by **Bogdan Dirică** from **Pexels**

Definition

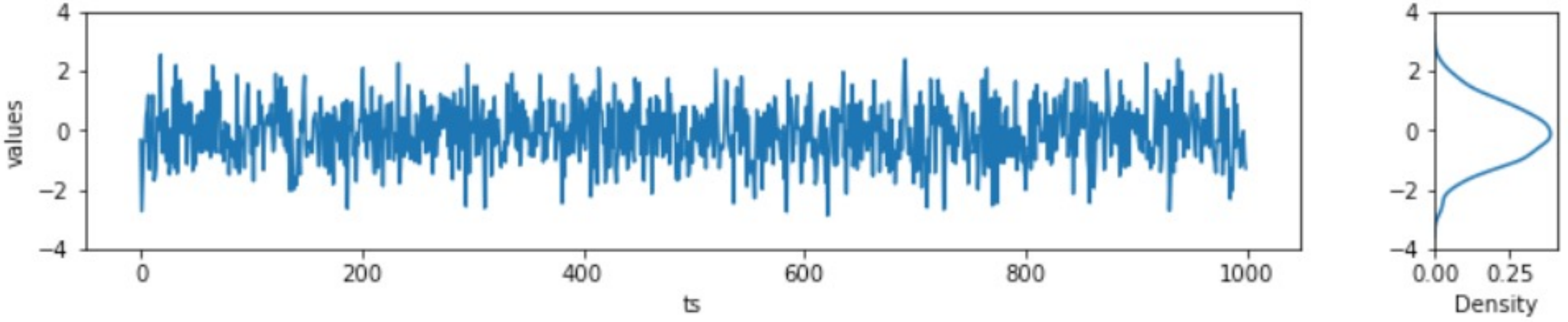
A **stationary** time series is one whose **properties do not depend on the time** at which the series is observed.

Let’s build the intuition of **stationarity**

Stationarity

White noise: the perfect time series

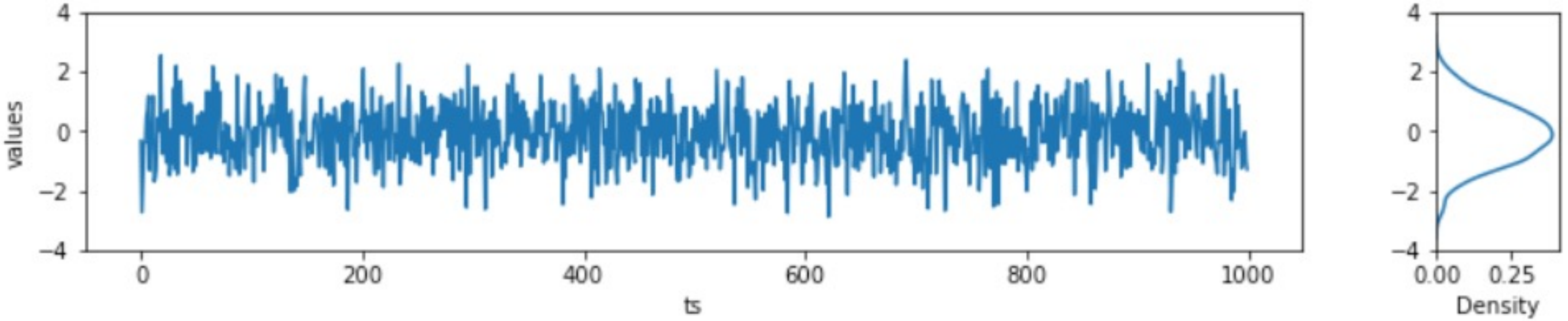
- A sequence of random numbers with zero mean and finite variance

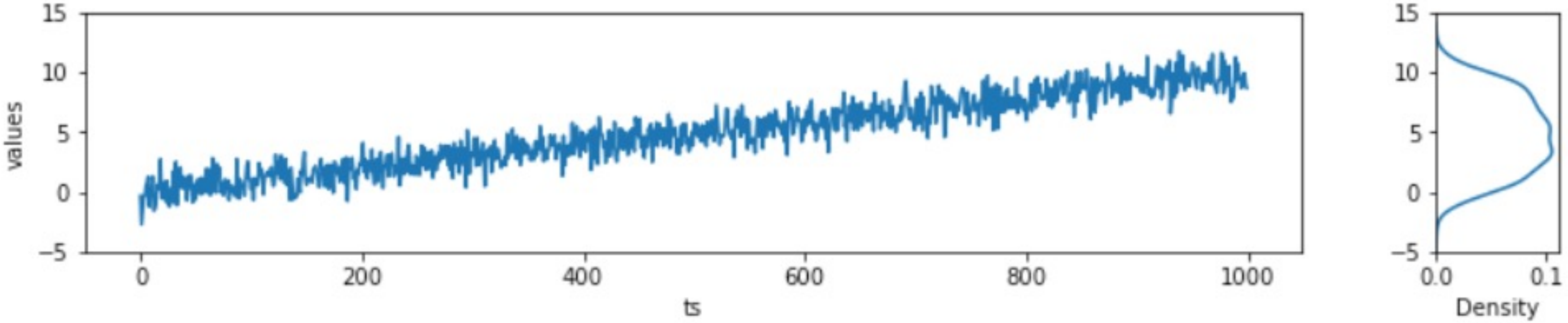


- NOTE: it is a perfect example of a stationary time series
  - If you predict 0 (the mean), you minimize the error (which is proportional to the variance)

Stationarity

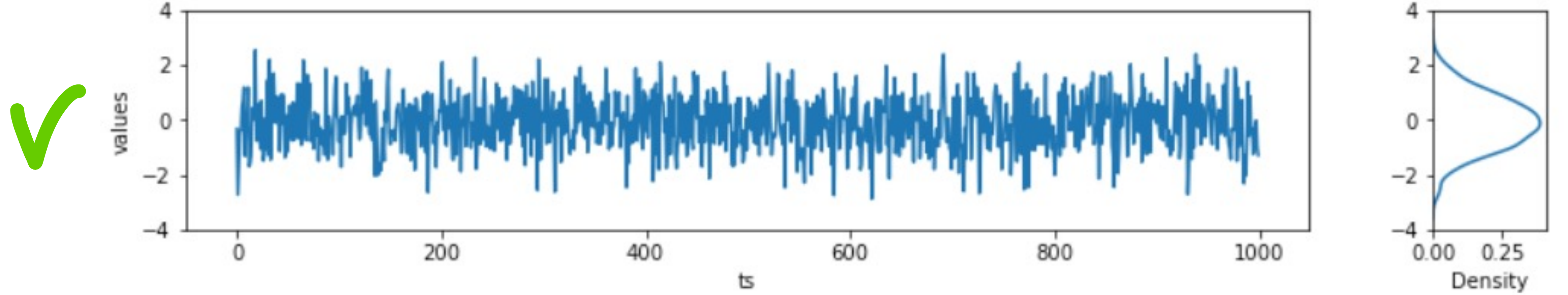
Mean is **constant** over time (a.k.a., with trend)

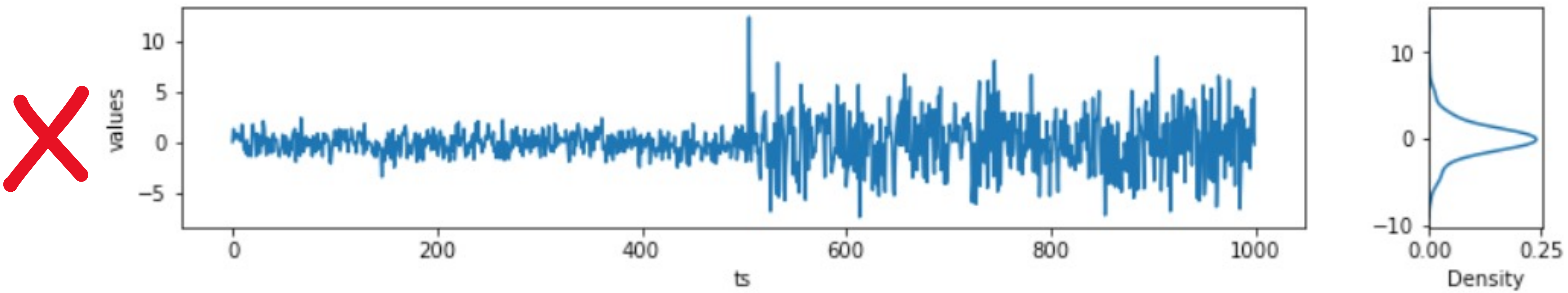




Stationarity

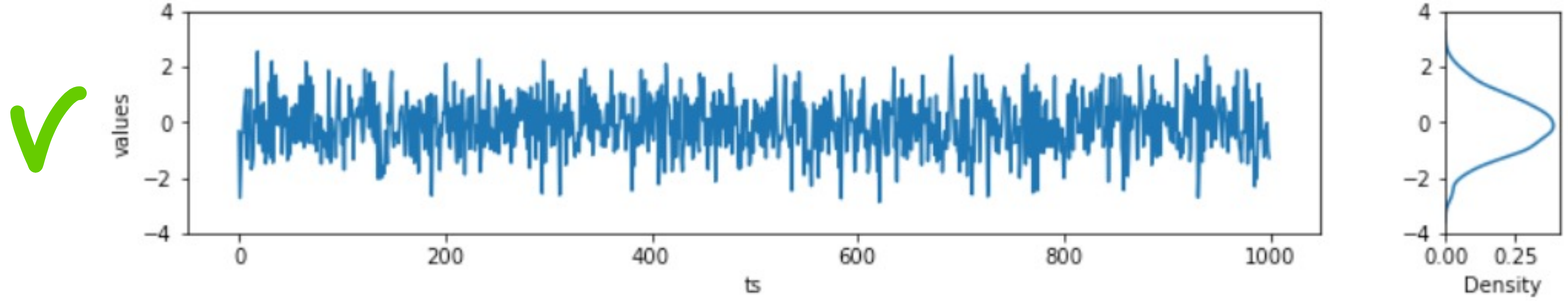
Variance is **constant** over time

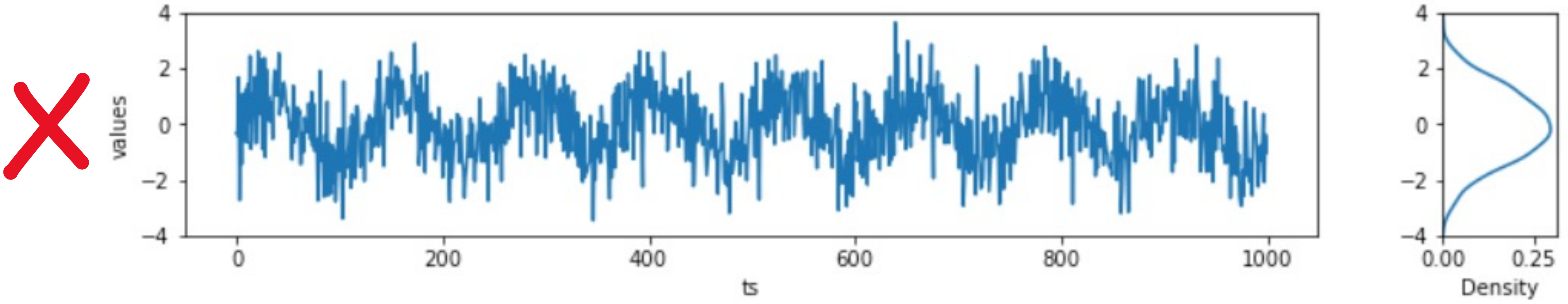




Stationarity

No repetitive pattern (a.k.a. **seasonality**)



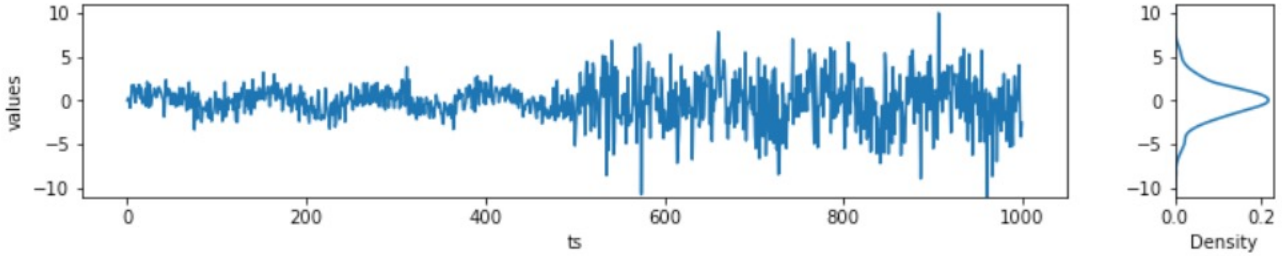


Stationarity

No combinations of the previous ones :-P

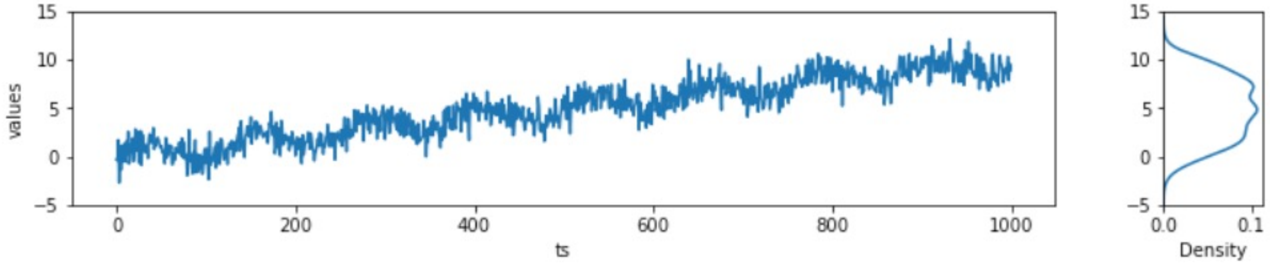
non-constant variance

+ seasonality



non-constant mean

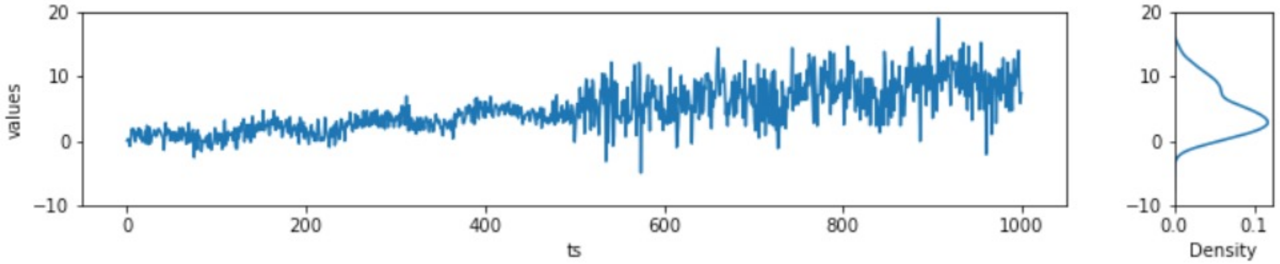
+ seasonality



Non-constant mean

+ non-constant variance

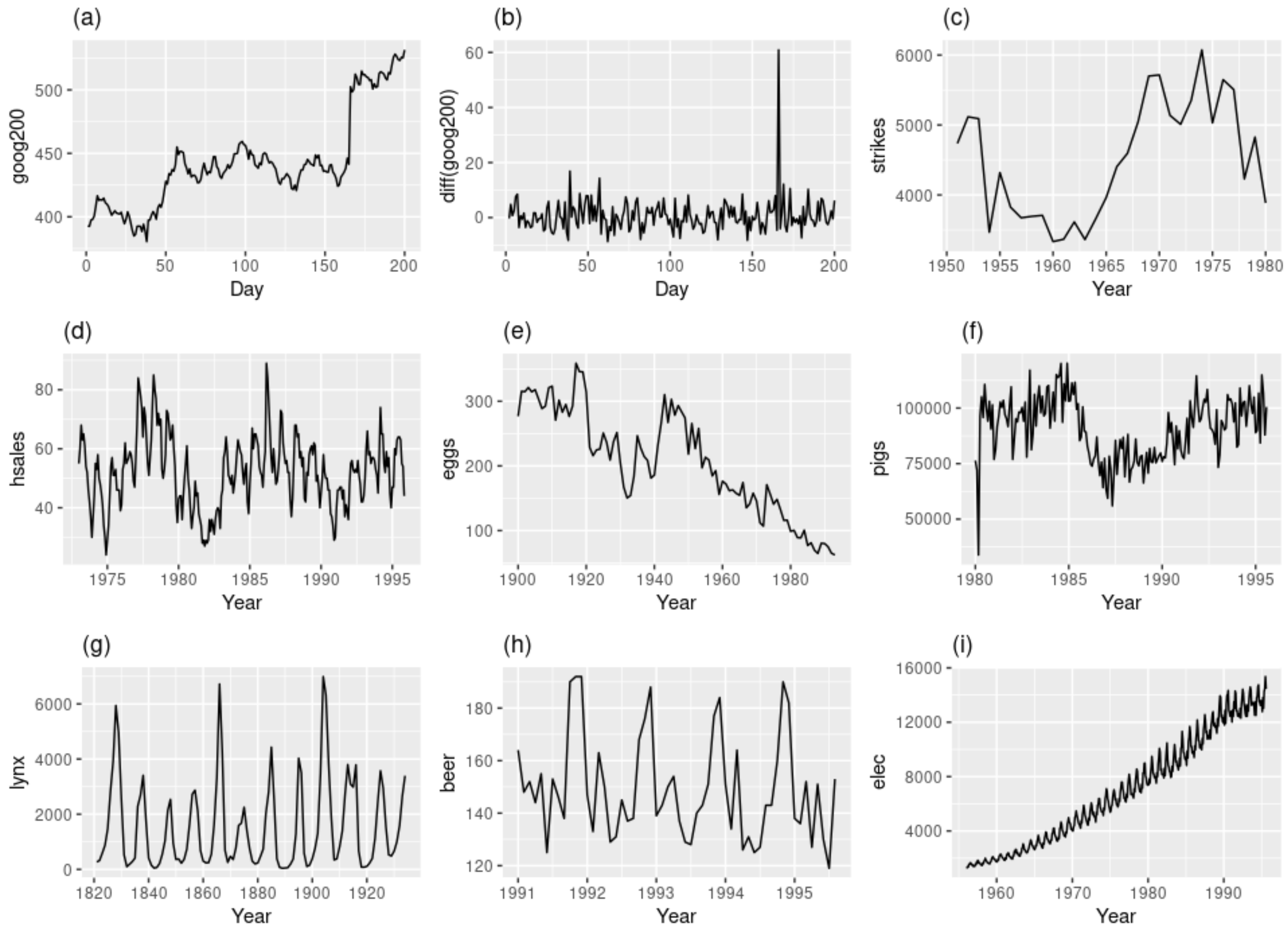
+ seasonality



Stationarity
Let's see if you got the point

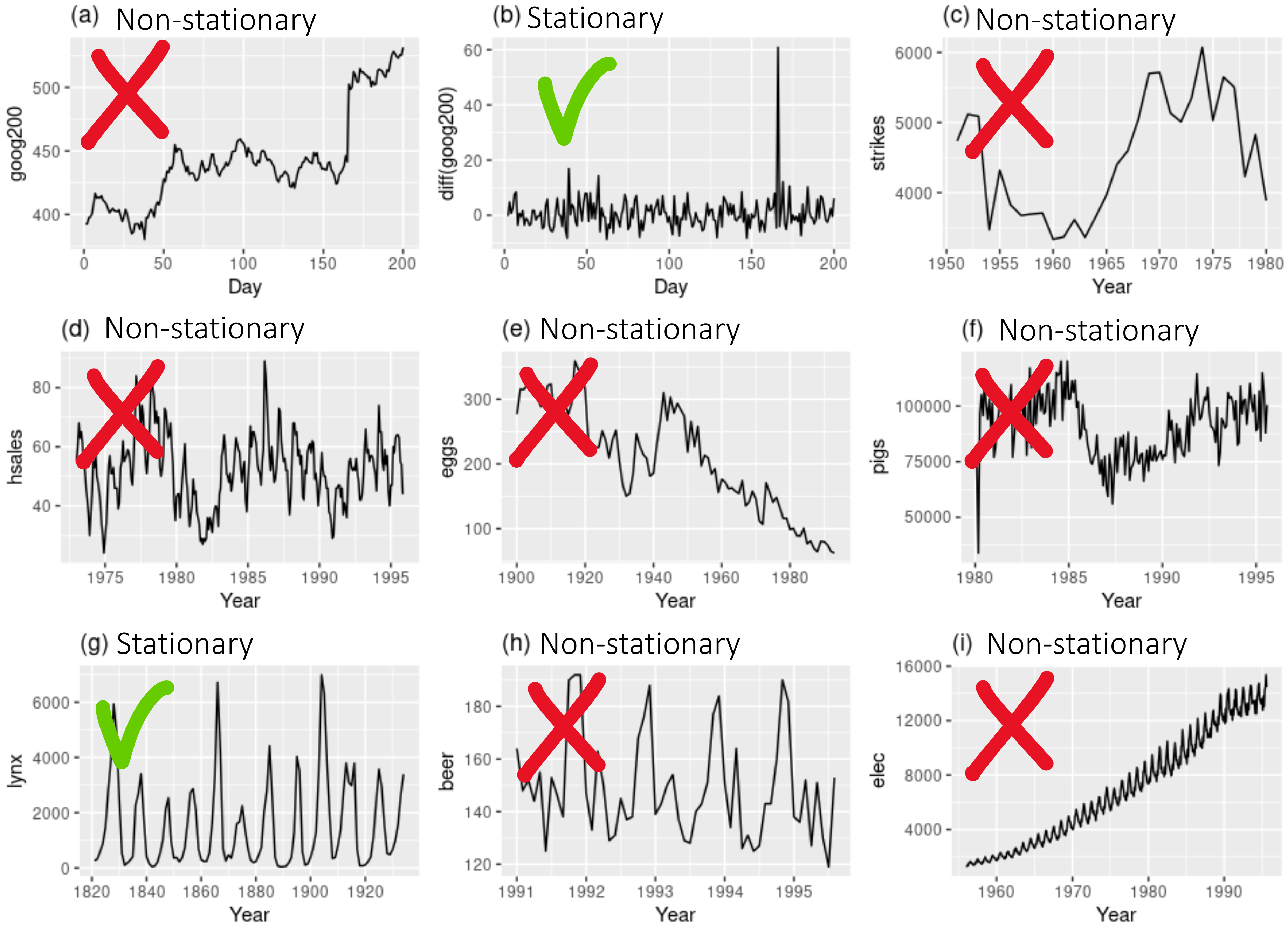
- Which of these time series is stationary?
- Why?

Quiz



stationary

non-stationary

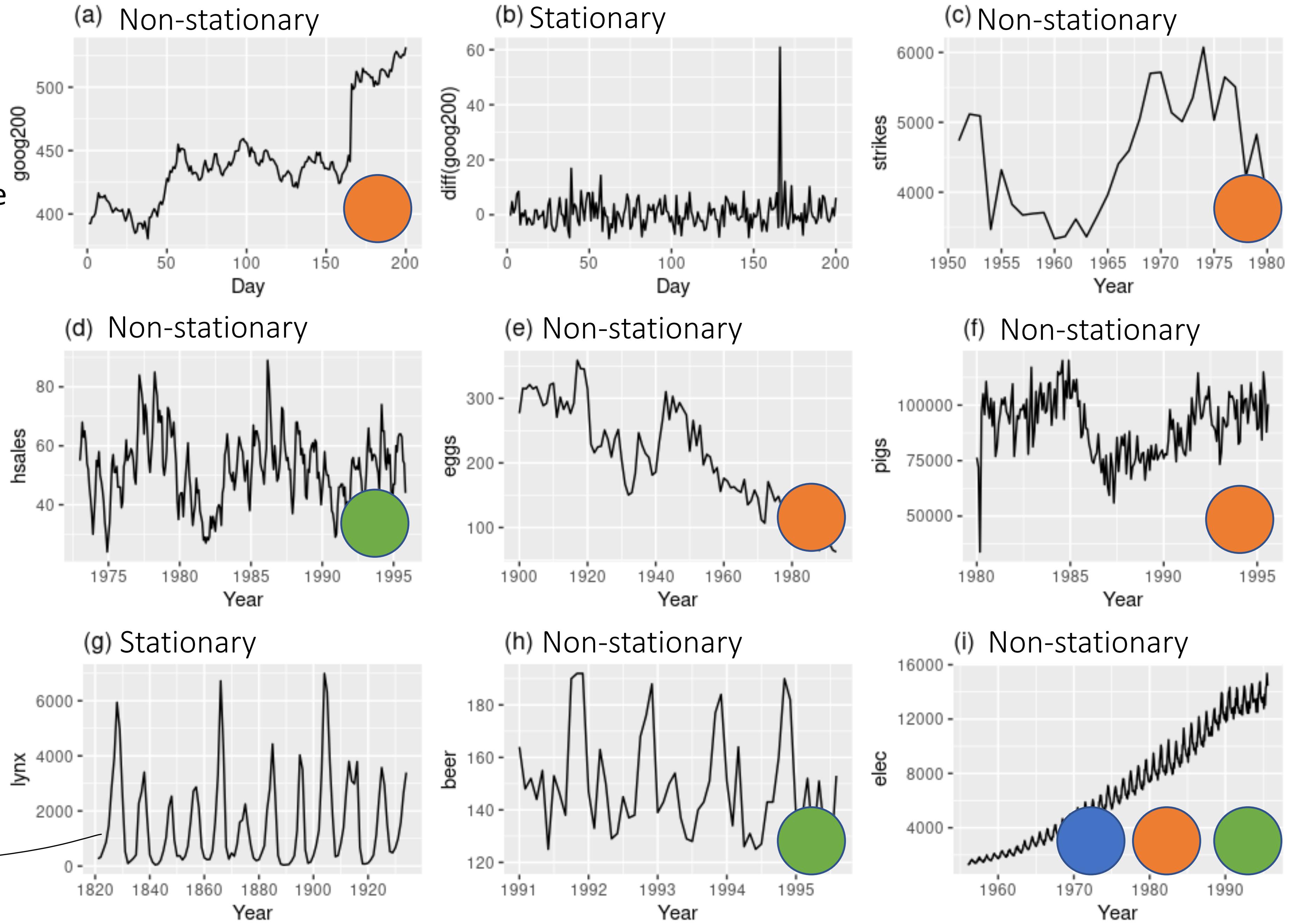


Non-constant mean

Non-constant variance

With seasonality

Irregular cycles
At first glance, the strong cycles might appear to make it non-stationary. But these cycles are aperiodic — they are caused when the lynx population becomes too large for the available feed, so that they stop breeding and the population falls to low numbers, then the regeneration of their food sources allows the population to grow again, and so on. In the long-term, the timing of these cycles is not predictable. Hence the series is stationary.



Formally…

Photo by **George Becker** from **Pexels**

Stationary
Formal definition

- Let $\{ X_t \}$ be a stochastic process and let $F_X(t_{1+\tau}, \ldots, t_{k+\tau})$ represent the cumulative distribution function of the unconditional (i.e., with no reference to any particular starting value) joint distribution of $\{ X_t \}$ at times $t_{1+\tau}, \ldots, t_{k+\tau}$.
- Then, $\{ X_t \}$ is said to be **strictly stationary, strongly stationary or strict-sense stationary** if

$$F_X(t_1, \ldots, t_k) = F_X(t_{1+\tau}, \ldots, t_{k+\tau}) \text{ for any } \tau\, e\, k.$$

How to **test** for **stationarity**


Photo by Chokniti Khongchum from Pexels

Stationary

By hand …

1. Load a time series
2. Split it two parts
3. Compute mean and variance of the two parts
4. Compare them

Stationary
Statistical tests

- We can test for stationarity using statistical tests.
- ADF test: Augmented Dickey Fuller test
- KPSS test: Kwiatkowski-Phillips-Schmidt-Shin test

Time-series Analytics

Giacomo Ziffer

giacomo.ziffer@polimi.it

POLITECNICO MILANO 1863