Time series visualization techniques I

EC 361–001

Prof. Santetti
Spring 2024

Materials

Required readings:

- `Hyndman & Athanasopoulos, ch. 2`

Motivation

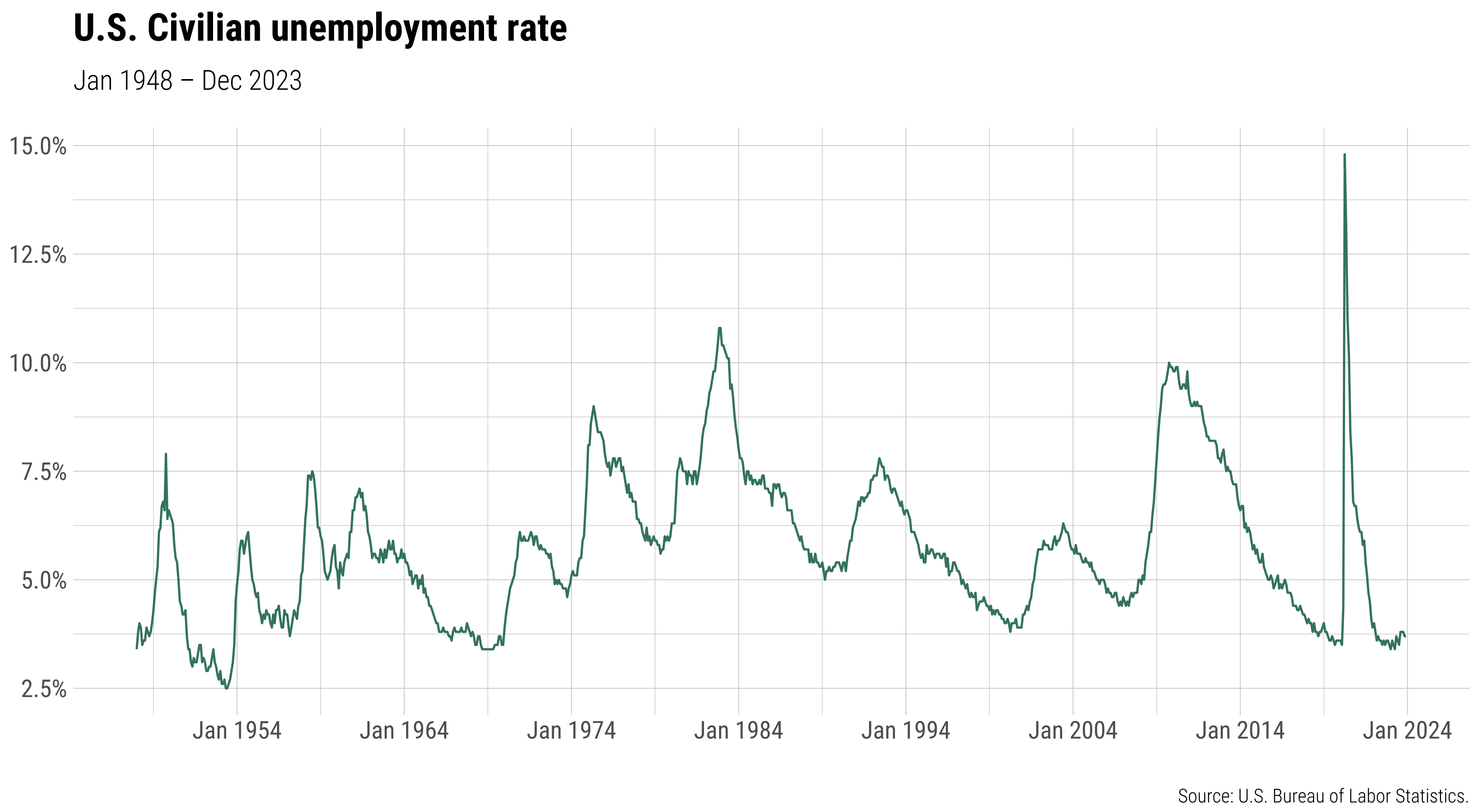
Motivation

Many times, a single **picture** informs more than *1,000 words*.

For instance, we can spend several minutes discussing about the ups and downs of the U.S. **unemployment rate** over time.

While words are indeed *needed* and *informative*, a **graph** can do the same in a more **efficient** way.

Motivation



U.S. Civilian unemployment rate
Jan 1948 – Dec 2023
15.0%
12.5%
10.0%
7.5%
5.0%
2.5%
Jan 1954
Jan 1964
Jan 1974
Jan 1984
Jan 1994
Jan 2004
Jan 2014
Jan 2024
Source: U.S. Bureau of Labor Statistics.

Motivation

The **first** thing to do in any data analysis task is to **plot the data**.

Graphs enable many **features** of the data to be visualized:

- *Patterns*;
- Unusual observations (i.e., *outliers*);
- *Changes* over time;
- *Relationships* between variables.

The reason **why** we should graph our data is that these features must then be **incorporated**, as much as possible, into the forecasting methods to be used.

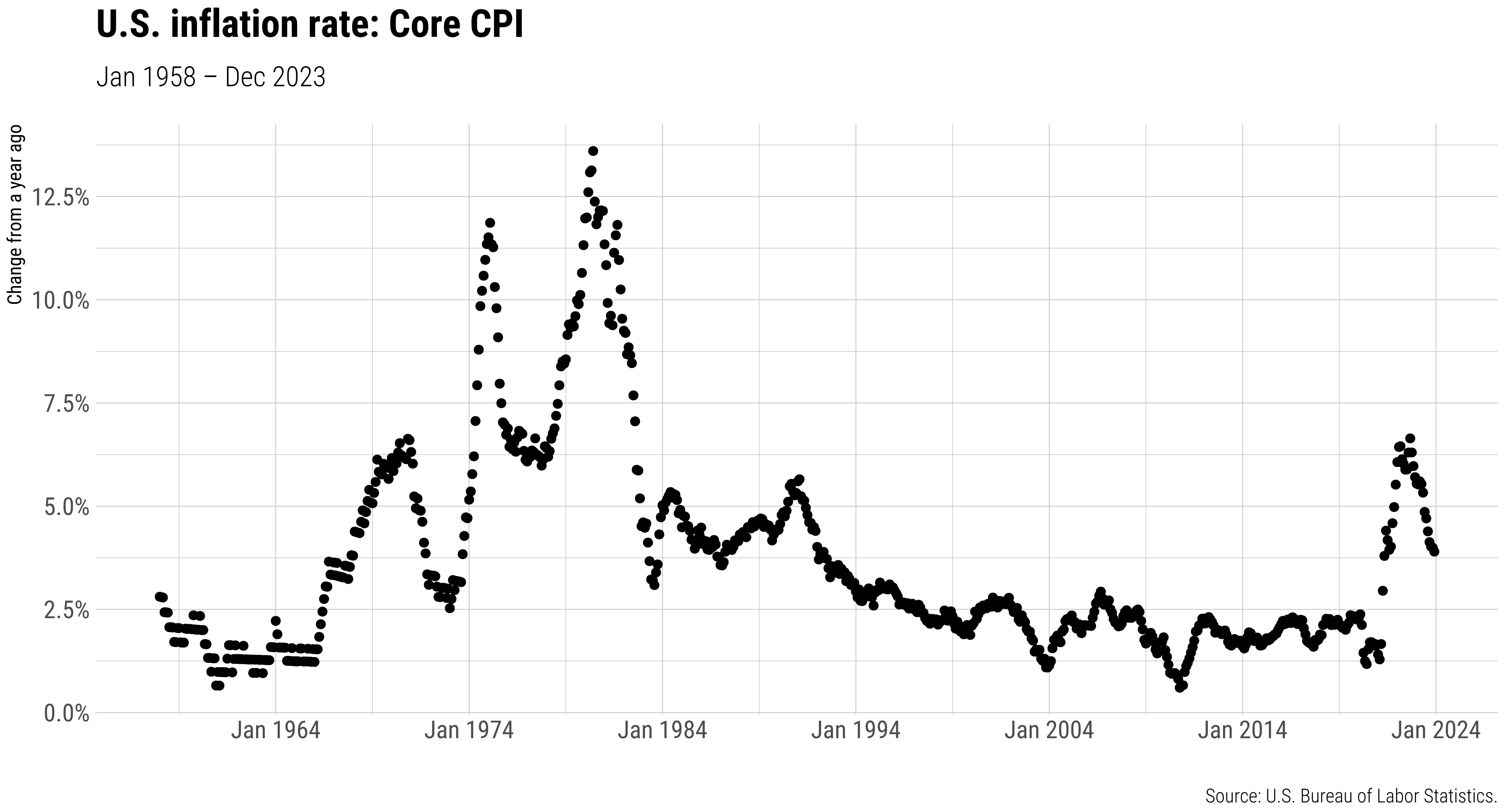
Time plots

Time plots

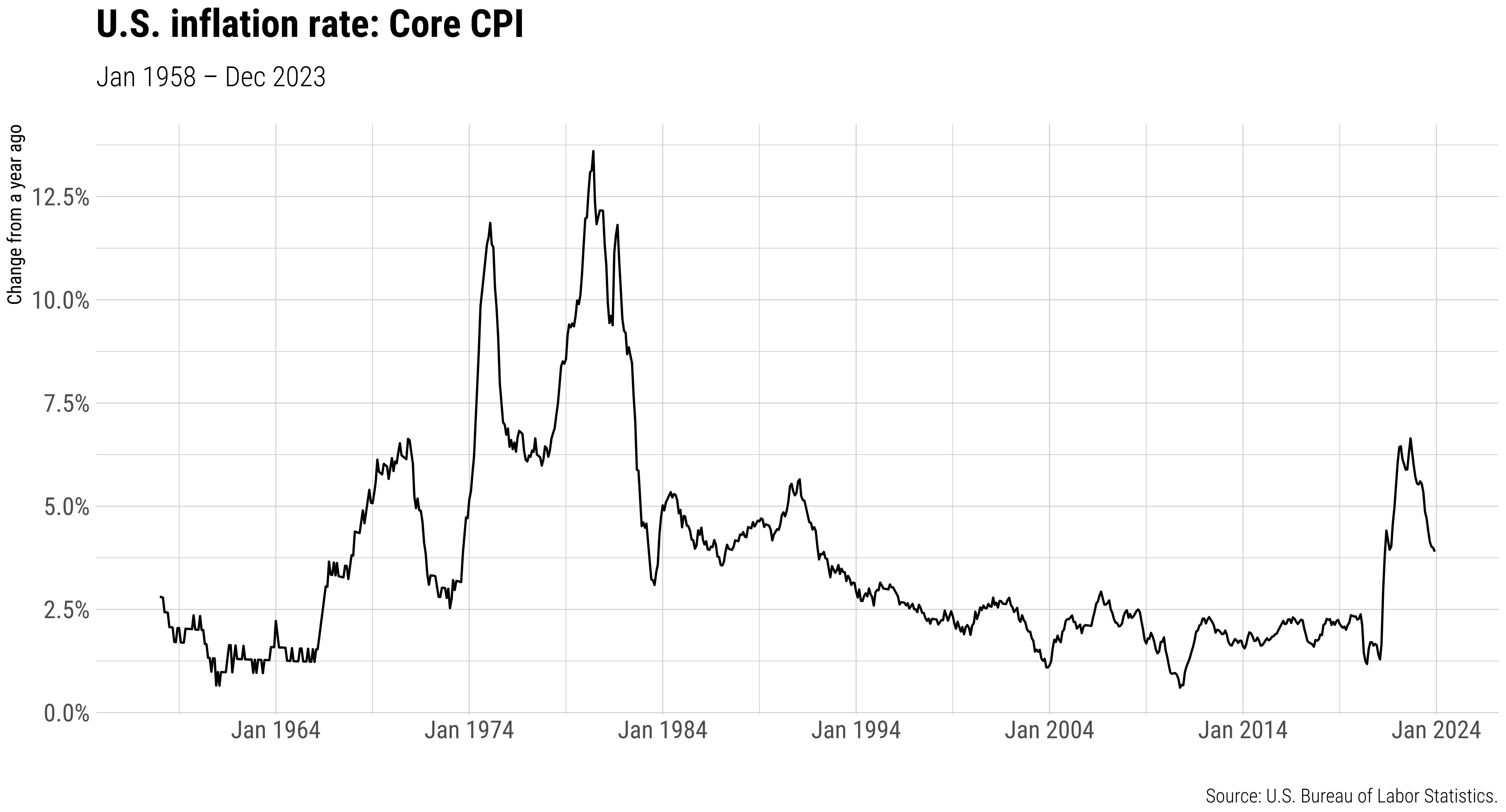
The most **obvious** place to start is by graphing the variable(s) *over time*.

Time plots display data points against the time of observation, with consecutive observations joined by *straight lines*.

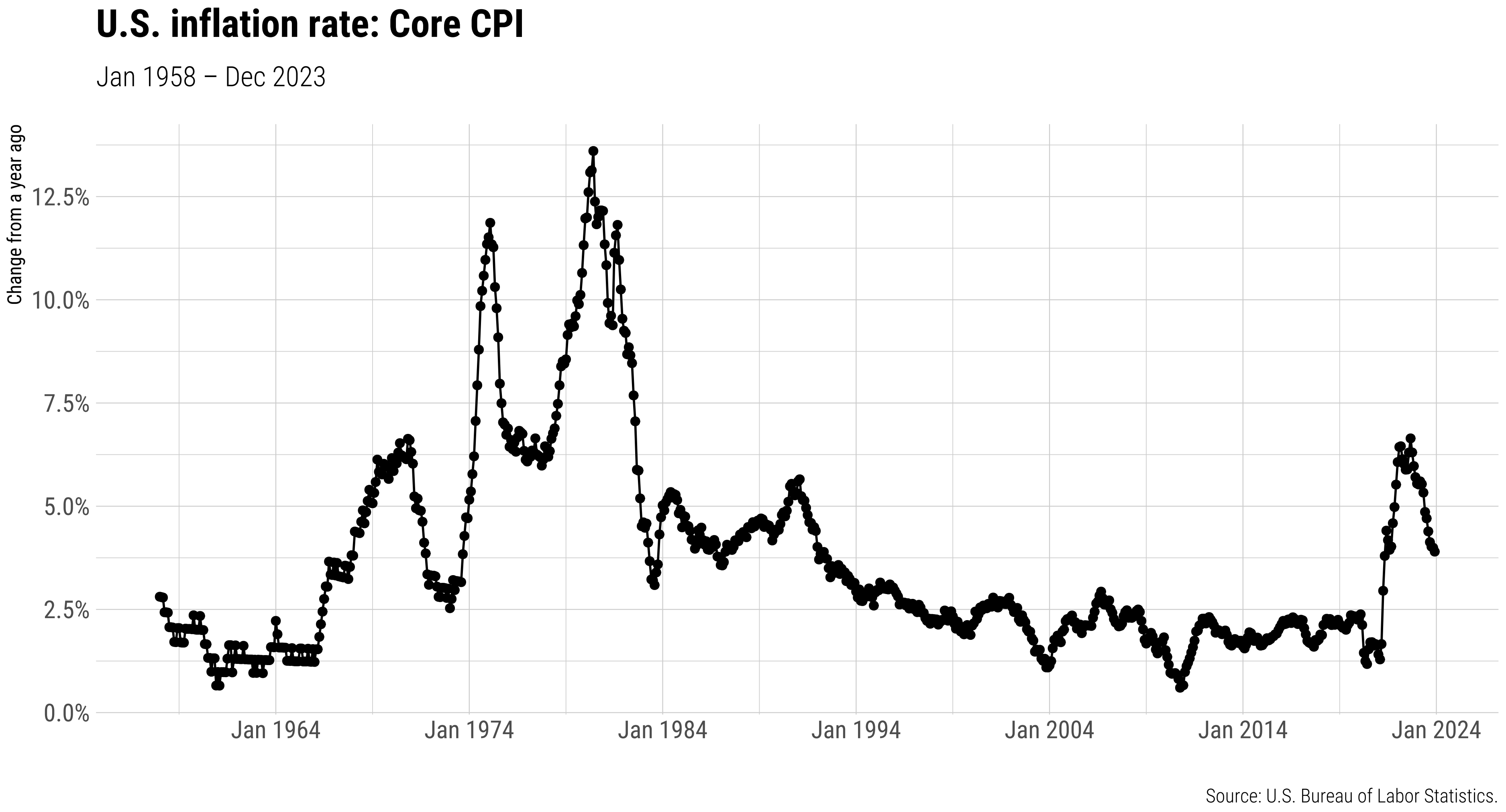
Time plots


U.S. inflation rate: Core CPI
Jan 1958 – Dec 2023
Change from a year ago
12.5%
10.0%
7.5%
5.0%
2.5%
0.0%
Jan 1964
Jan 1974
Jan 1984
Jan 1994
Jan 2004
Jan 2014
Jan 2024
Source: U.S. Bureau of Labor Statistics.

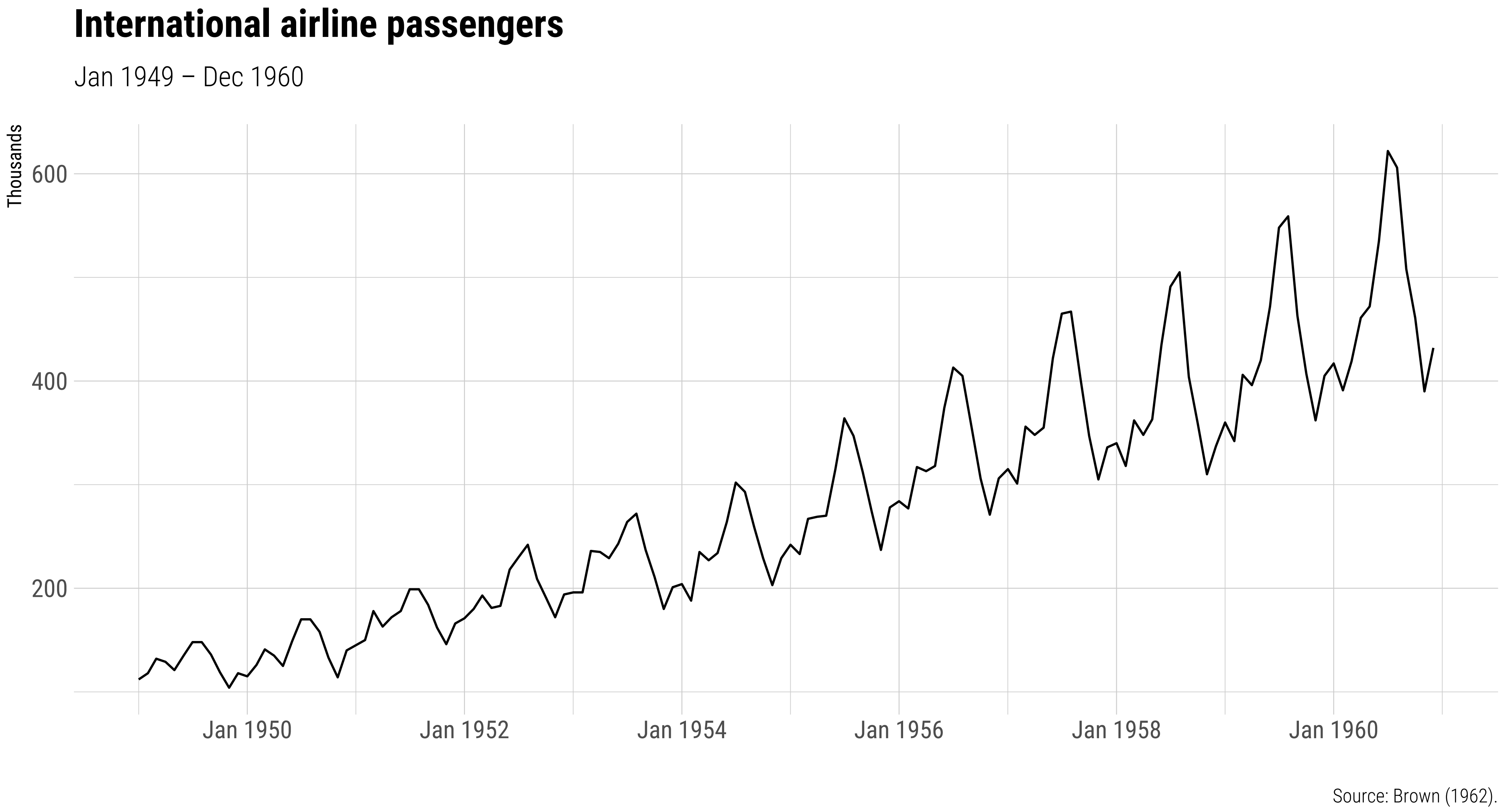
Time plots


U.S. inflation rate: Core CPI
Jan 1958 – Dec 2023
Change from a year ago
12.5%
10.0%
7.5%
5.0%
2.5%
0.0%
Jan 1964
Jan 1974
Jan 1984
Jan 1994
Jan 2004
Jan 2014
Jan 2024
Source: U.S. Bureau of Labor Statistics.

Time plots


U.S. inflation rate: Core CPI
Jan 1958 – Dec 2023
Change from a year ago
12.5%
10.0%
7.5%
5.0%
2.5%
0.0%
Jan 1964
Jan 1974
Jan 1984
Jan 1994
Jan 2004
Jan 2014
Jan 2024
Source: U.S. Bureau of Labor Statistics.

Time plots


International airline passengers
Jan 1949 – Dec 1960
Thousands
600
400
200
Jan 1950
Jan 1952
Jan 1954
Jan 1956
Jan 1958
Jan 1960
Source: Brown (1962).

Time series patterns

Time series patterns

Many time series have important **features**, which should be taken into account when the aim is to *predict* future values.

The first is the **trend**.

- Is there a *long-run* increase/decrease in the data?
- If so, such behavior might be **linear** or not.

The second is a series' **seasonal** component.

- Is the series affected by *seasonal factors* (e.g., time of the year, day of the week, hour of the day?)

Lastly, many time series show **cyclic** behavior.

- Cycles occur when the data exhibit *rises and falls* that are not of a *fixed frequency*.

Time series patterns

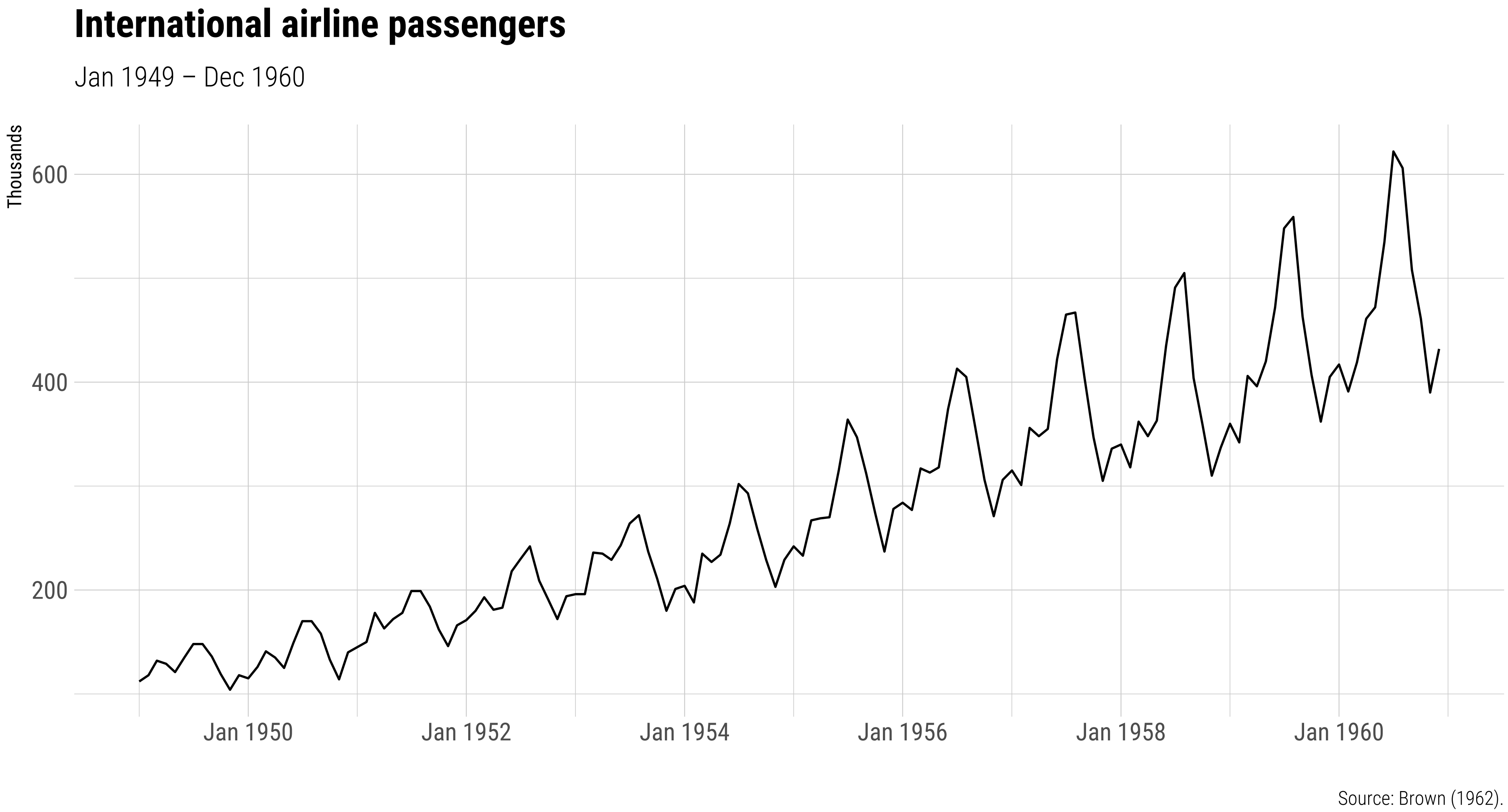
We should clearly **distinguish** between **seasonality** and **cycles**.

While **seasonal** patterns are *unchanging* and associated with some aspect of the *calendar*, **cycles** do not have a *fixed* frequency and tend to have a longer average *length* than seasonality.

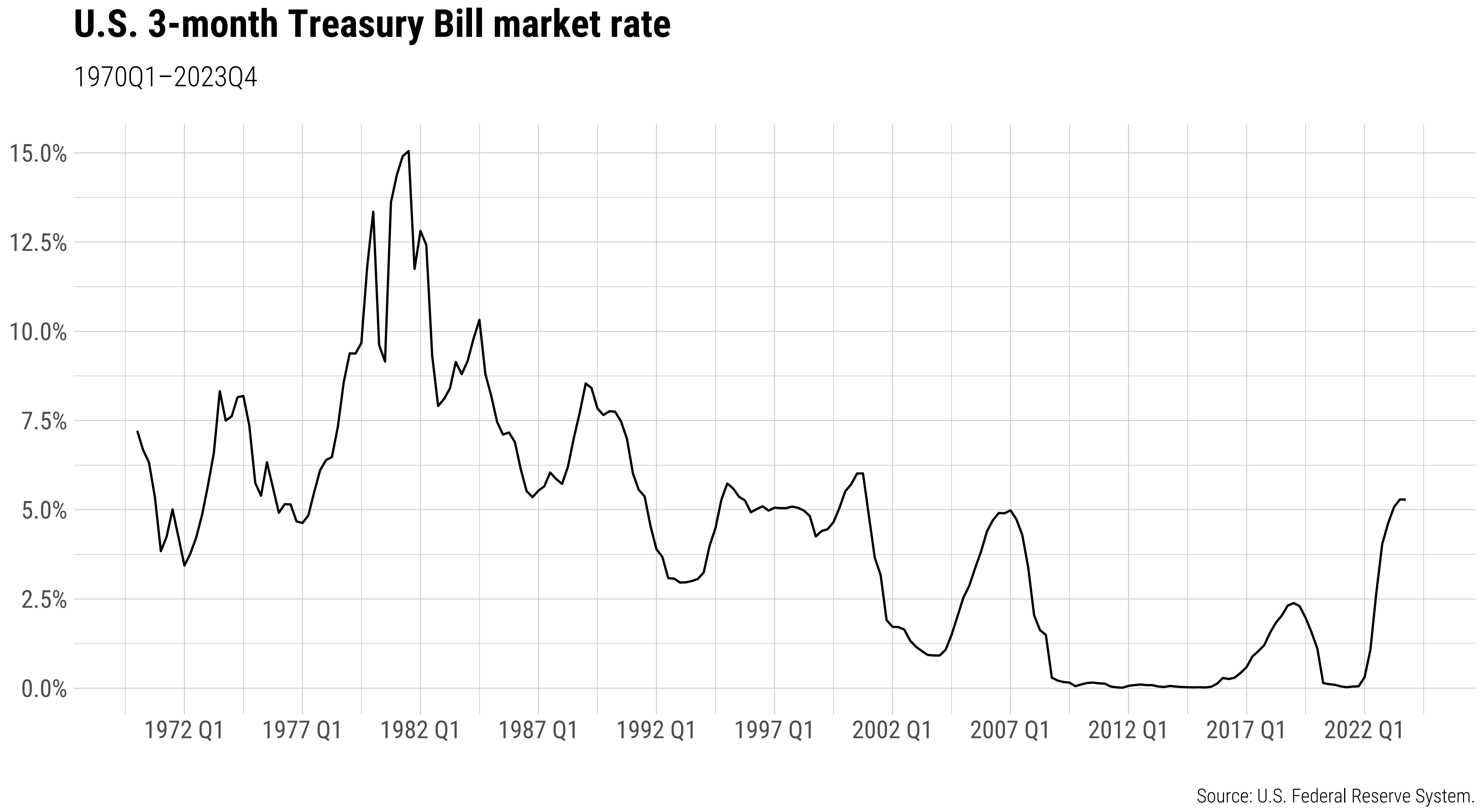
A time series may have *all* three components, while others might have *two*, just *one*, or no visible pattern *at all*.

When **choosing** a forecasting method, we will first need to *identify* the time series patterns in the data, and then choose a method that is able to capture the patterns properly.

Time series patterns



Time series patterns



Seasonal plots

Seasonal plots

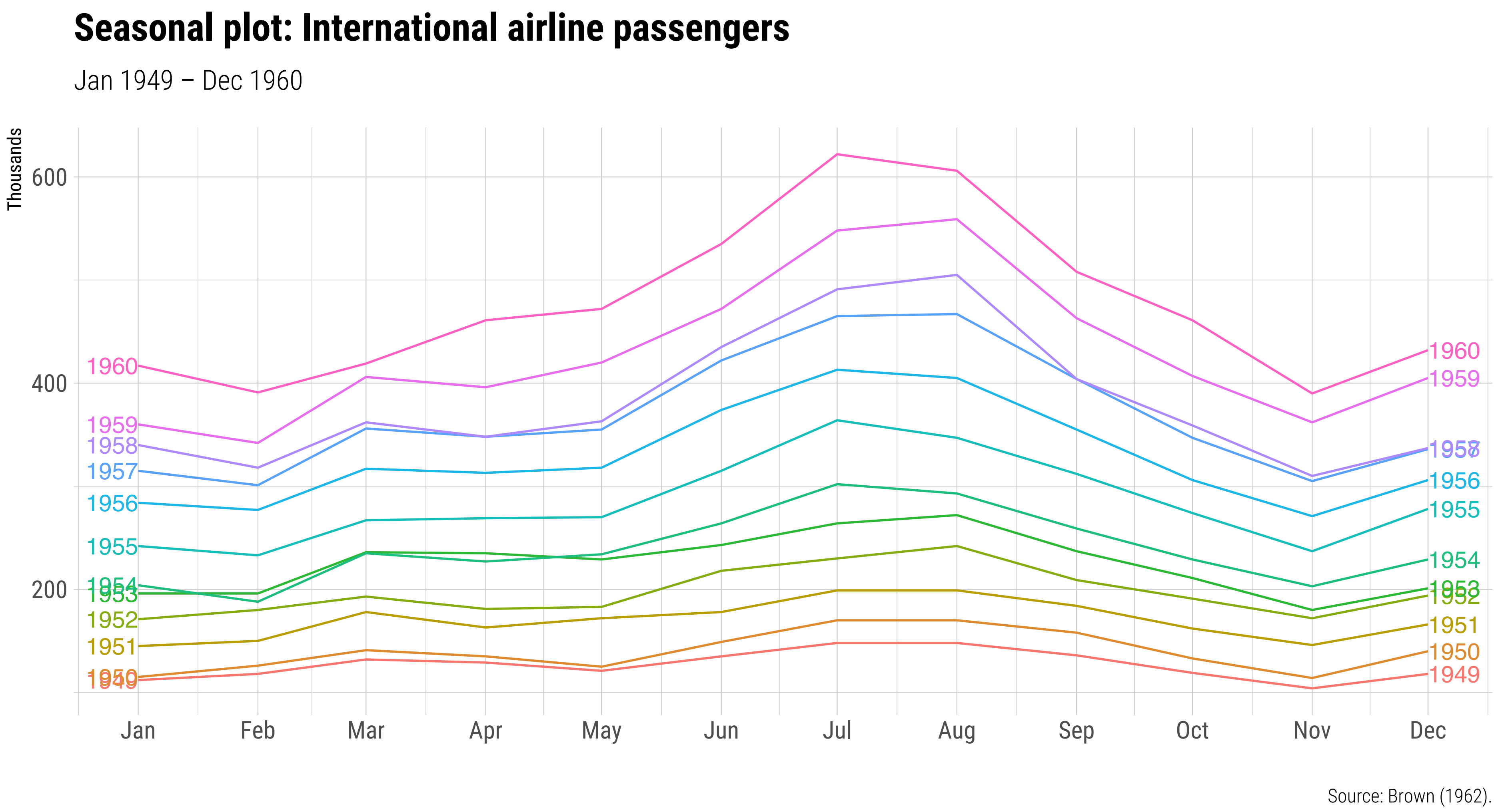
For time series with strong **seasonal** behavior, **seasonal plots** allow for better visualizing the underlying *seasonal pattern*.

In this kind of plot, data points are plotted against the *season* in which they are observed.

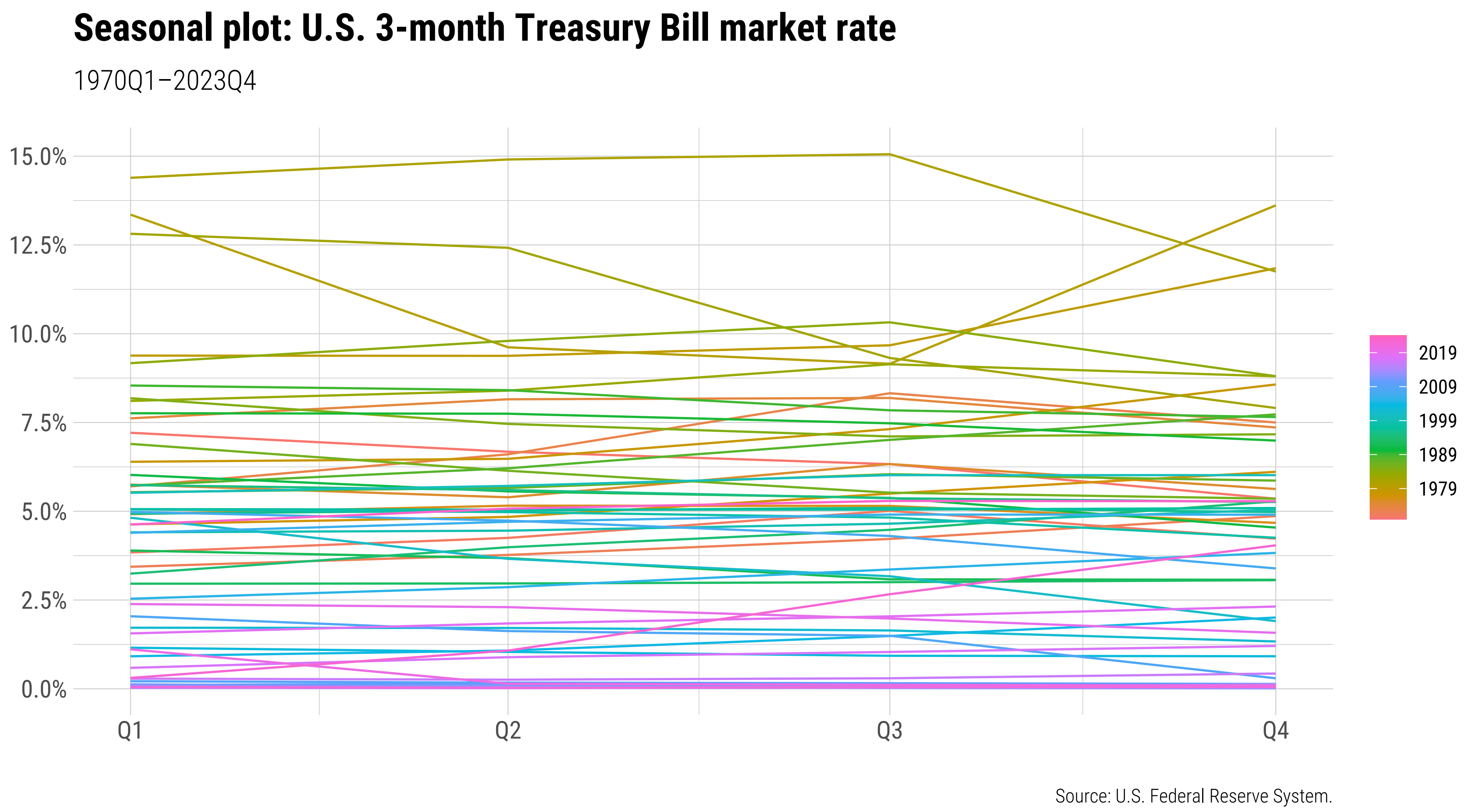
This season might be a *day of the week*, a *month*, a *year*, and so on.

- It all depends on the **frequency** of the time series we are working with!

Seasonal plots



Seasonal plots



Scatter plots

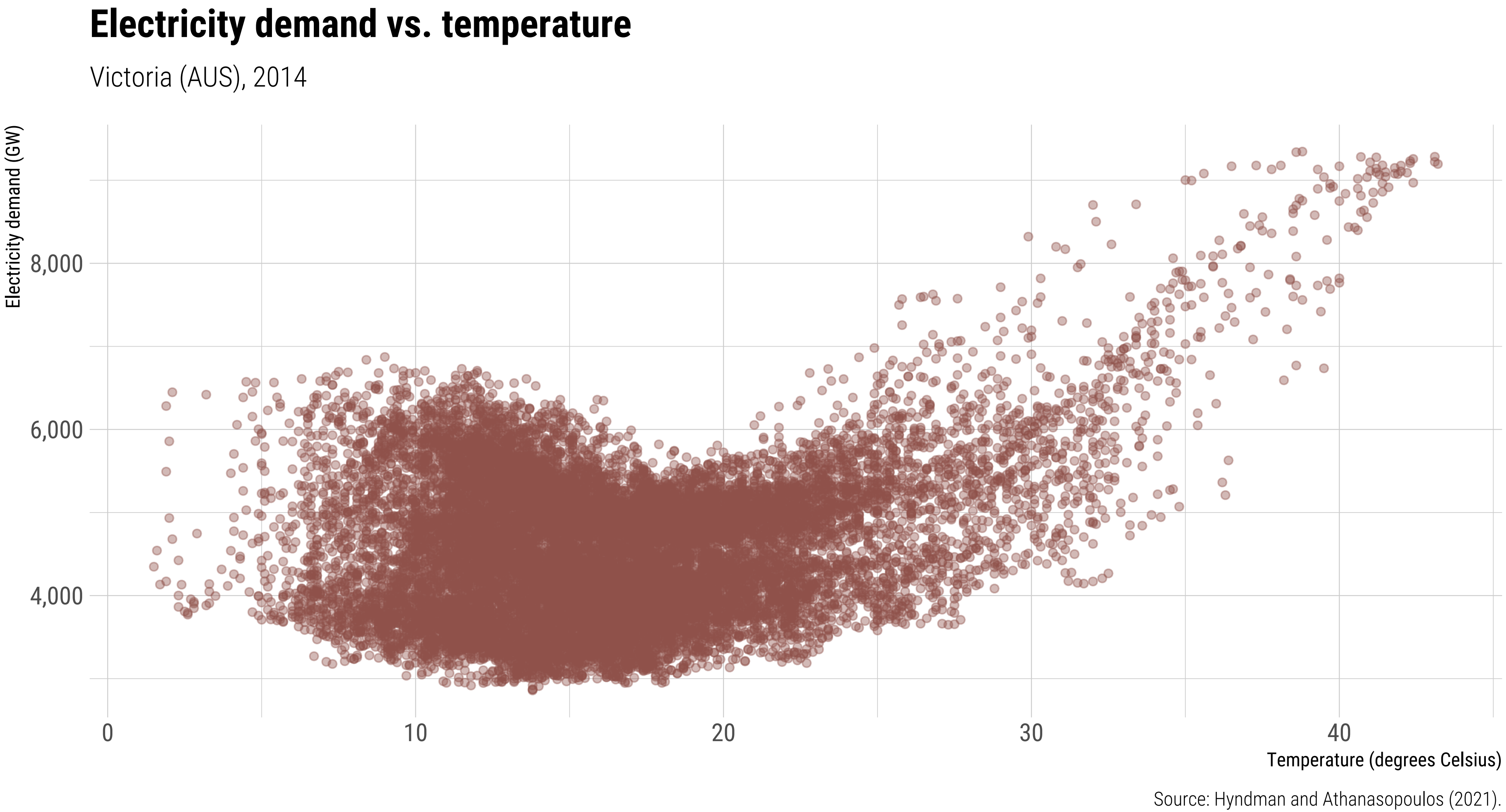
Scatter plots

The techniques we've studied so far *only* capture the behavior of **one** specific time series over time.

In case we are interested in looking at the **relationship(s)** between time series, we need to move on to other techniques.

The most appropriate one are **scatter plots**.

Scatter plots



Scatter plots

When analyzing **bivariate** relationships for time series, one important *descriptive* measure is the **correlation coefficient**.

> Correlation measures the **linear** association between two variables.

For two time series, x and y, their correlation coefficient (r) is given by:

$$r = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2}\sqrt{\sum(y_t - \bar{y})^2}}$$

And its value lies between **-1** (perfect, *negative* linear association) and **+1** (perfect, *positive* linear association).

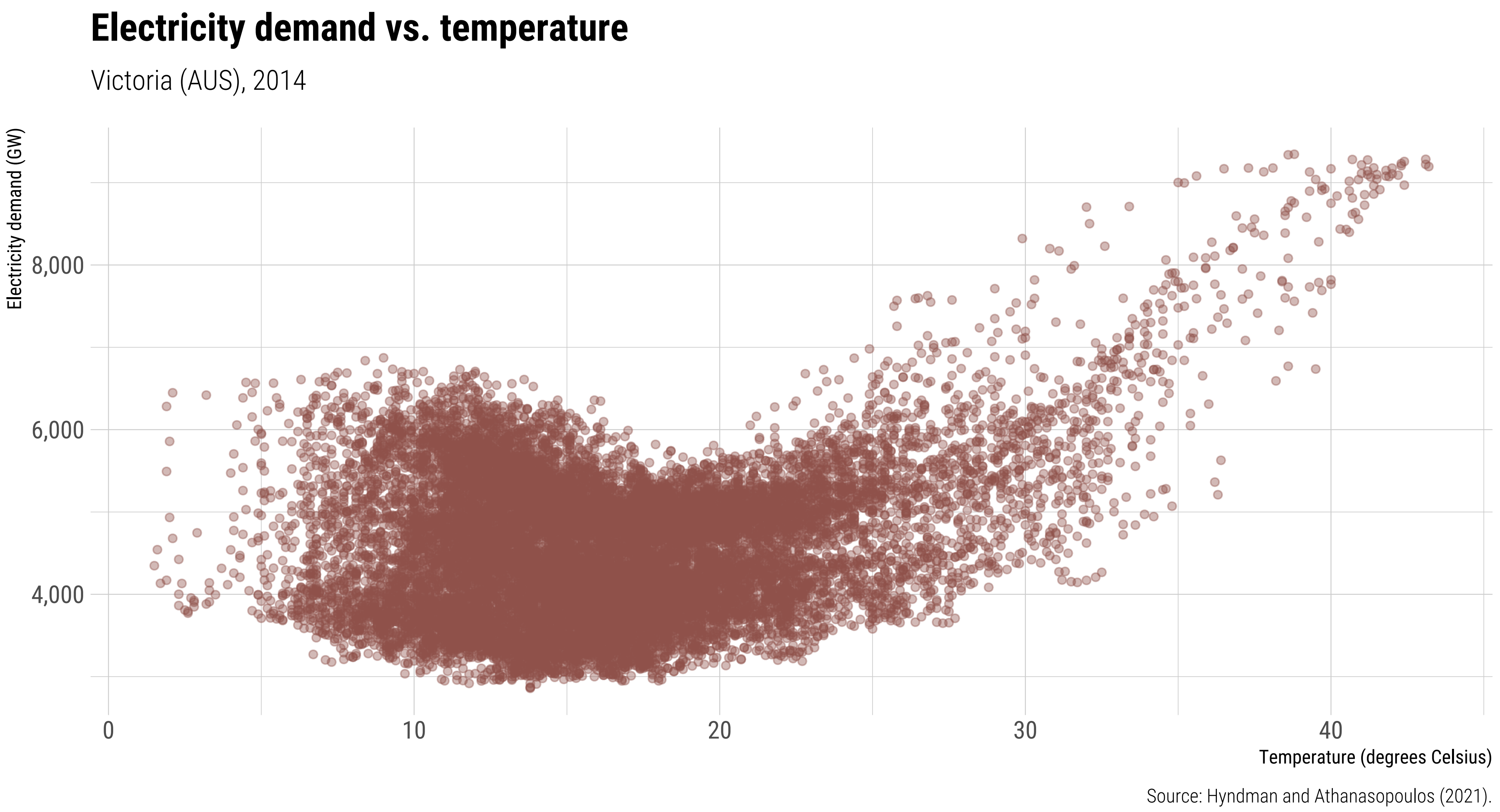
Scatter plots

When using correlation coefficients, **two** things to keep in mind:

1. ***Correlation is not causation***!

2. Looking for *linear* relationships may be **misleading**.

Scatter plots



Next time: Time series graphics II