Time series visualization techniques II

EC 361–001

Prof. Santetti

Spring 2024

Materials

Required readings:

- `Hyndman & Athanasopoulos, ch. 2`

Motivation

Motivation

So far, we have studied *three* types of time-series **visualization** techniques:

- *Time* plots;
- *Seasonal* plots;
- *Scatter* plots.

There is *more* to explore, and we will study further techniques that can give important **insights** about the relevant **features** of our time series.

Autocorrelation

Autocorrelation

Before the next plots, we must clarify what a **lag** means:

> In time series jargon, the number of **time steps** between a series' observations is known as a **lag**.

In terms of **notation**, we will use the subscript t to denote a **time index**.

- For instance, given a time series called y, its value at time t can be denoted by y_t.

In case we want to denote the value of time series y **one step in the past**, we will write this as y_{t-1}.

Similarly, for any step k in the past, we can denote such value by y_{t-k}.

Such steps must be consistent with the **frequency** of the time series (e.g., monthly, quarterly, daily, yearly, etc.)

Autocorrelation

One key feature of time-series data is that **past** values usually help to explain **present** and/or **future** values.

Therefore, values of a time series in the *past* may be **correlated** with more *recent* observations.

More formally, for a given k value, $y_{t\text{-}k}$ and y_t may share a *nonzero* correlation coefficient.

Just as **correlation** measures the extent of a *linear* relationship between two variables, **autocorrelation** measures the linear relationship between *lagged values of a time series*.

Autocorrelation

The sample **autocovariance** at lag k of a time series y is given by

$$c_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{T}$$

The **variance** of y is given by

$$\mathrm{Var}(y) = \frac{\sum_{t=1}^{T} (y_t - \bar{y})(y_t - \bar{y})}{T}$$

And the **autocorrelation coefficient** at lag k is given by

$$r_k = \frac{c_k}{\mathrm{Var}(y)}$$

Lag plots

Lag plots

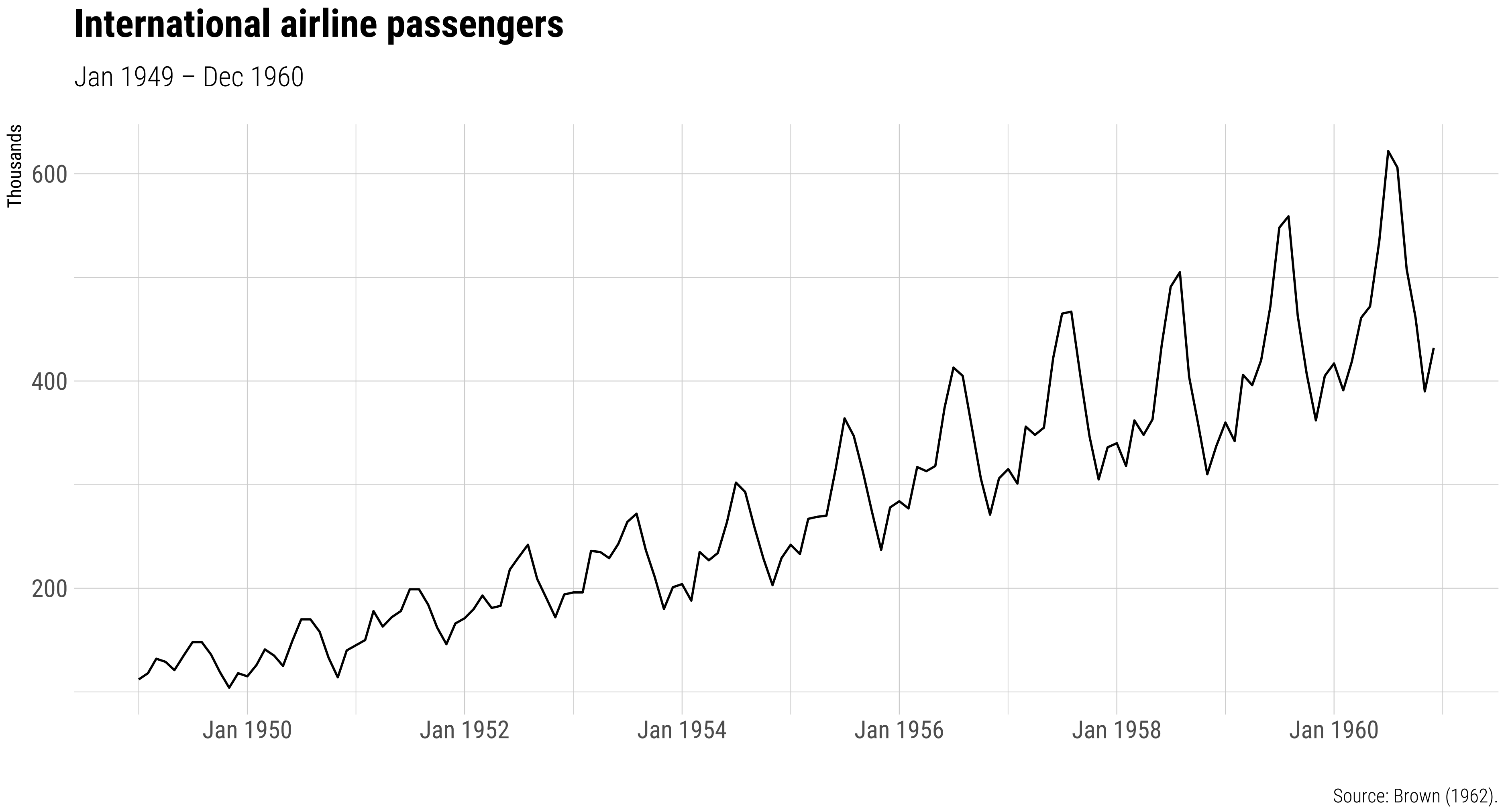
Lag plots display the data plotted against its different **lags** in a scatter plot.

Different **colors** correspond to each period (*vertical axis*) against lagged values (*horizontal axis*).

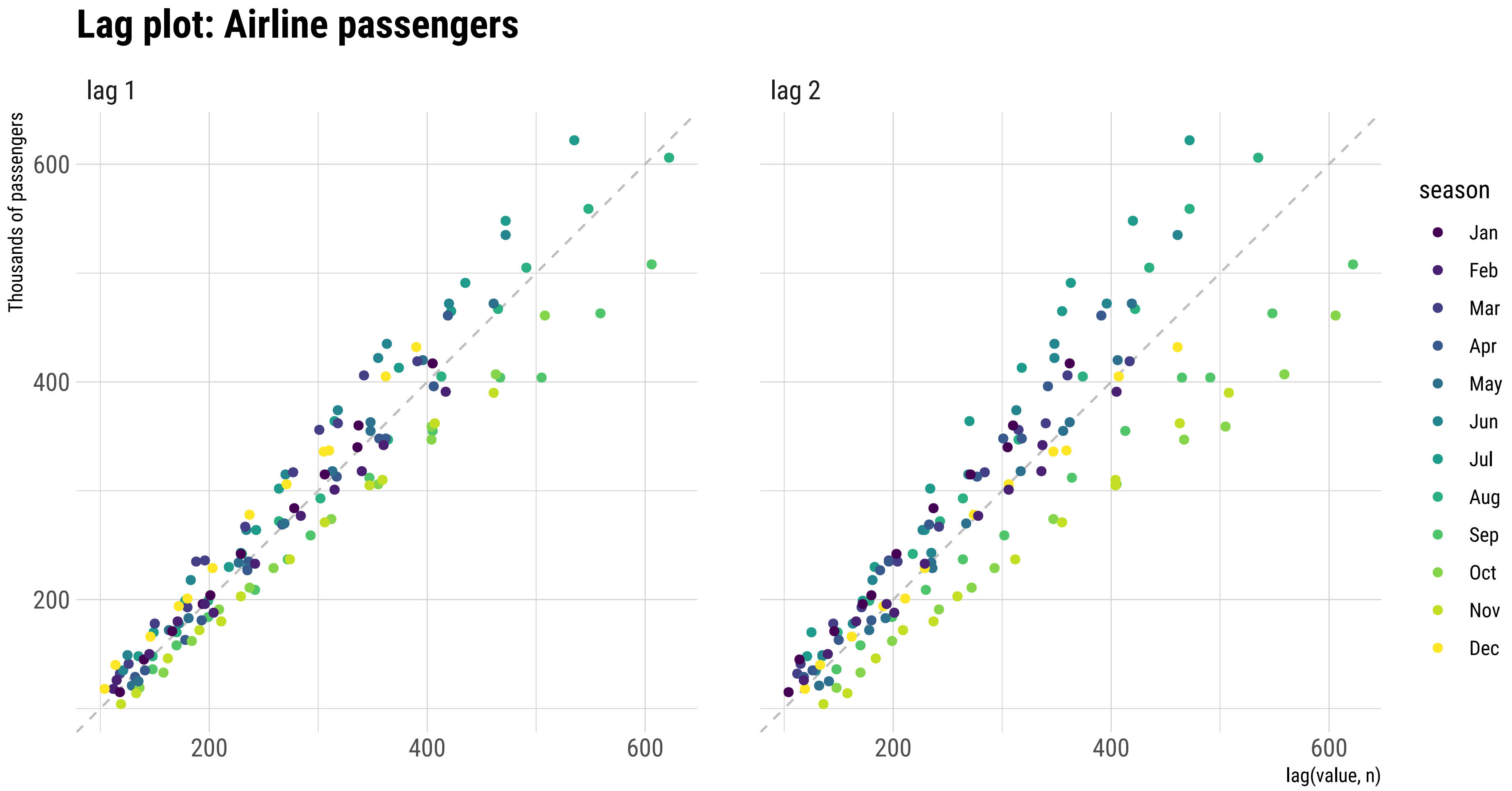
In case the time series shows **seasonality**, the autocorrelation coefficient will be **large** and **positive** at the **multiples** of the seasonal period.

- e.g., 12 lags for *monthly* data; 4 lags for *quarterly* data, and so on.

Lag plots

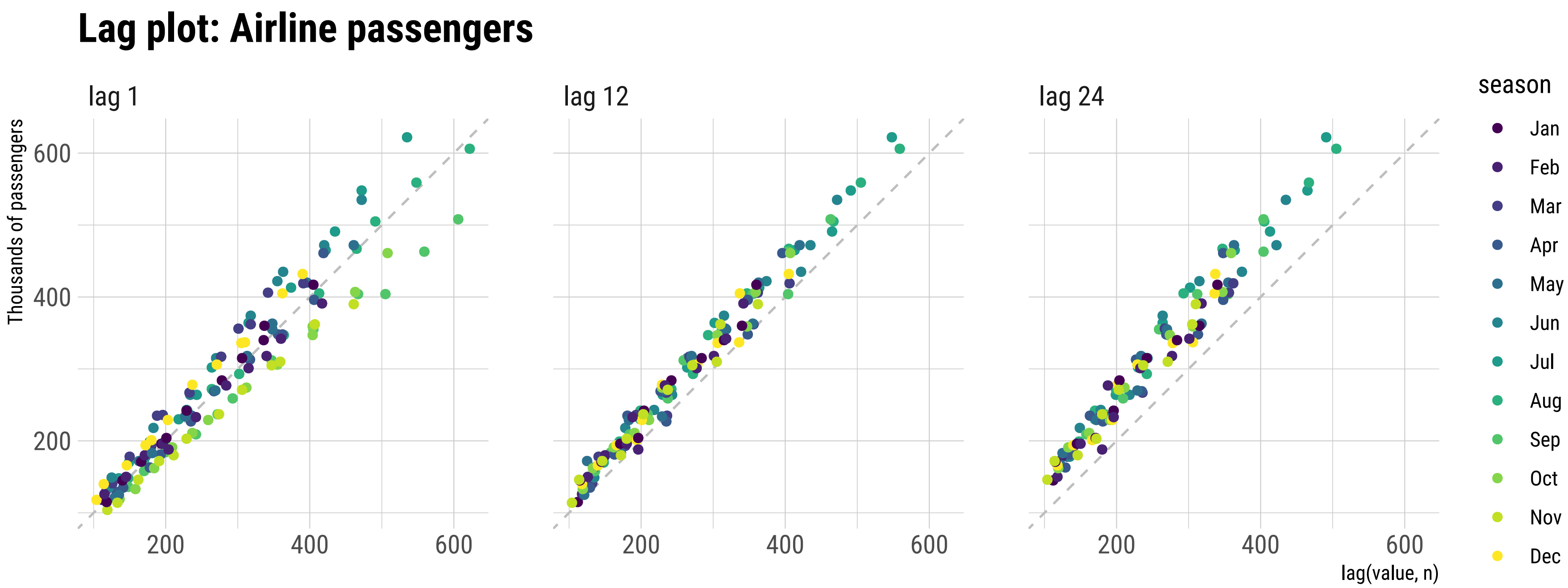

International airline passengers
Jan 1949 – Dec 1960
Thousands
600
400
200
Jan 1950
Jan 1952
Jan 1954
Jan 1956
Jan 1958
Jan 1960
Source: Brown (1962).

Lag plots



Lag plot: Airline passengers
lag 1
lag 2
Thousands of passengers
600
400
200
200
400
600
200
400
600
lag(value, n)
season
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Lag plots



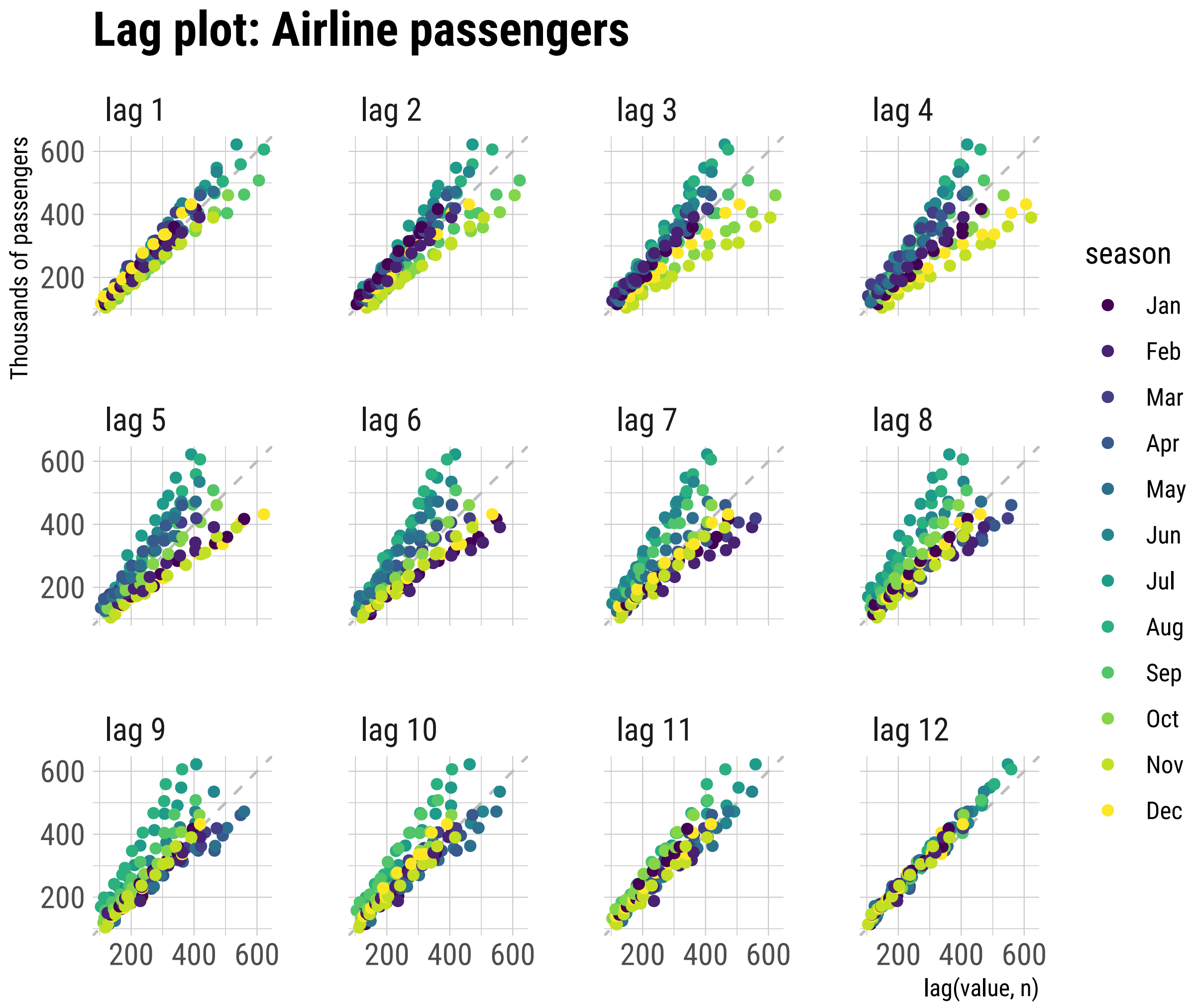
Lag plot: Airline passengers
lag 1
lag 12
lag 24
Thousands of passengers
600
400
200
200
400
600
lag(value, n)
season
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Lag plots

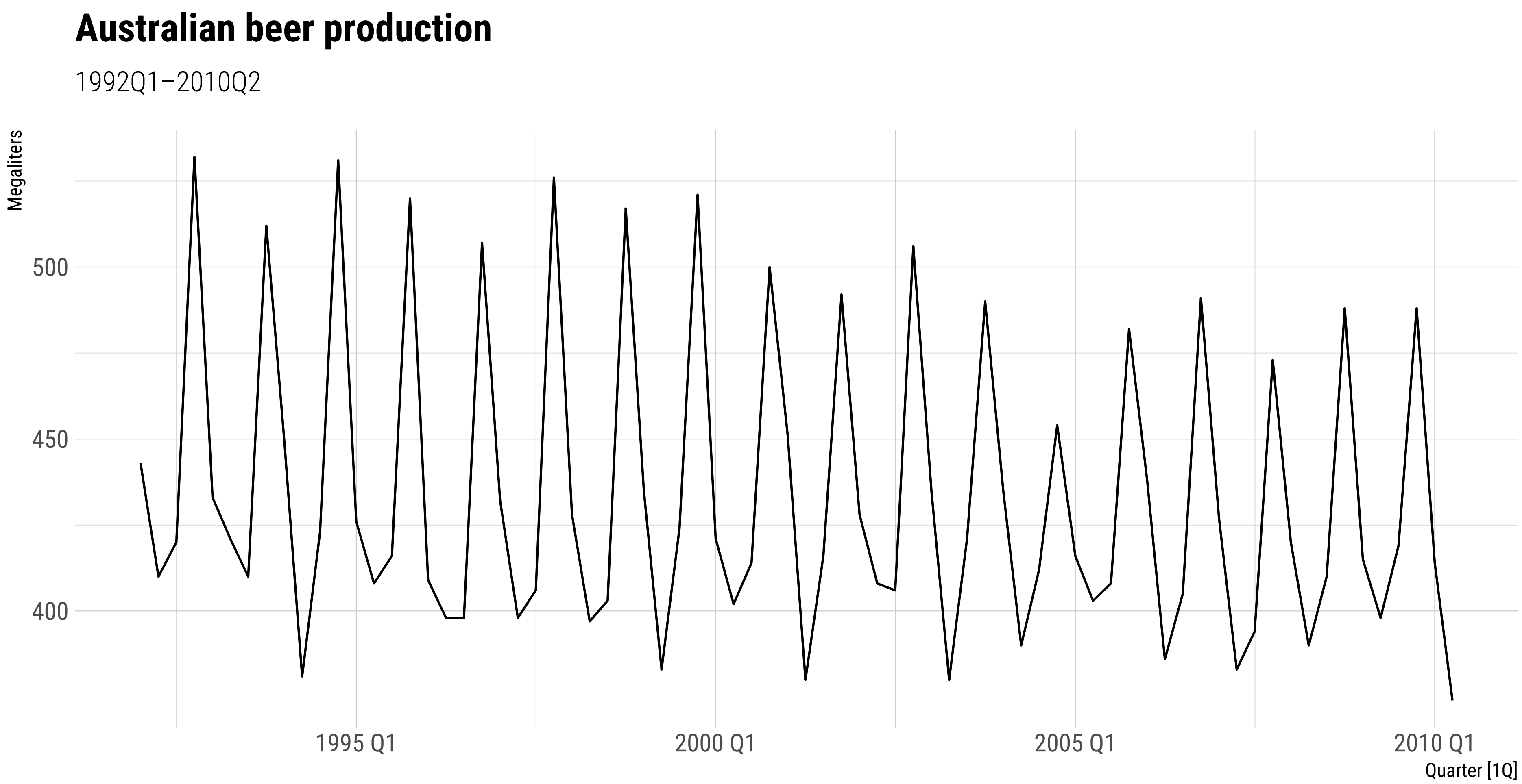
```
air_passengers ▷
  ACF(lag_max = 12)
```

```
#> # A tsibble: 12 x 2 [1M]
#>          lag   acf
#>    <cf_lag> <dbl>
#>  1        1M 0.948
#>  2        2M 0.876
#>  3        3M 0.807
#>  4        4M 0.753
#>  5        5M 0.714
#>  6        6M 0.682
#>  7        7M 0.663
#>  8        8M 0.656
#>  9        9M 0.671
#> 10       10M 0.703
#> 11       11M 0.743
#> 12       12M 0.760
```

Lag plots


Lag plot: Airline passengers
lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12
Thousands of passengers
200
400
600
lag(value, n)
season
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

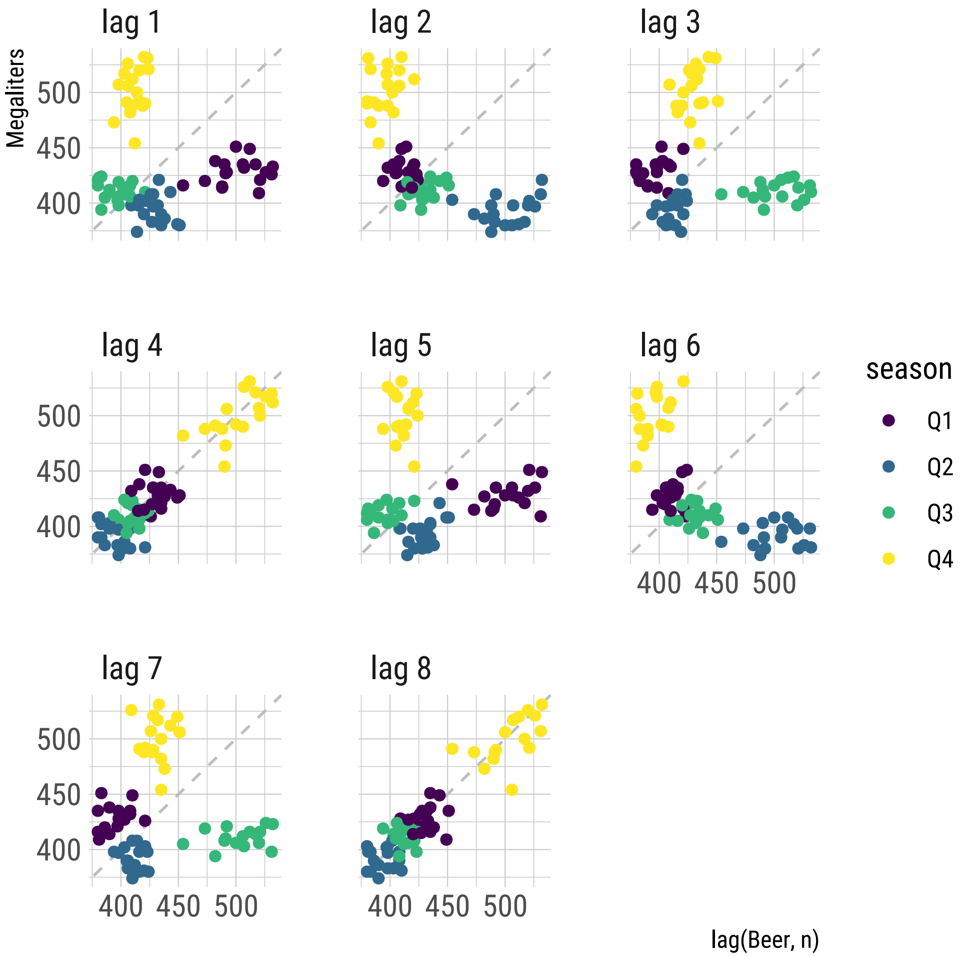
Lag plots



Source: Hyndman and Athanasopoulos (2021).

Lag plot: Australian beer production

1992Q1–2010Q2



Source: Hyndman and Athanasopoulos (2021).

Lag plots

```
beer |>
  ACF(lag_max = 8)
```

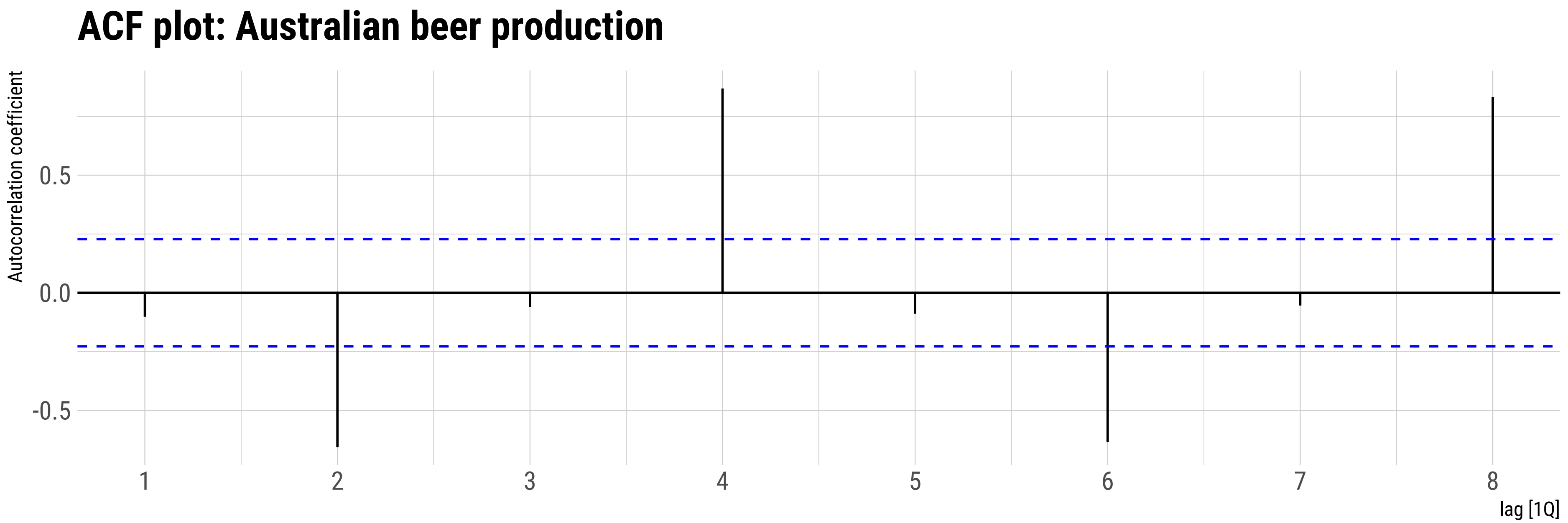
```
#> # A tsibble: 8 x 2 [1Q]
#>         lag     acf
#>   <cf_lag>   <dbl>
#> 1        1Q -0.102
#> 2        2Q -0.657
#> 3        3Q -0.0603
#> 4        4Q  0.869
#> 5        5Q -0.0892
#> 6        6Q -0.635
#> 7        7Q -0.0542
#> 8        8Q  0.832
```

The autocorrelation function

The autocorrelation function

A more **insightful** way to observe the autocorrelation coefficient is through the **Autocorrelation Function (ACF) plot**.

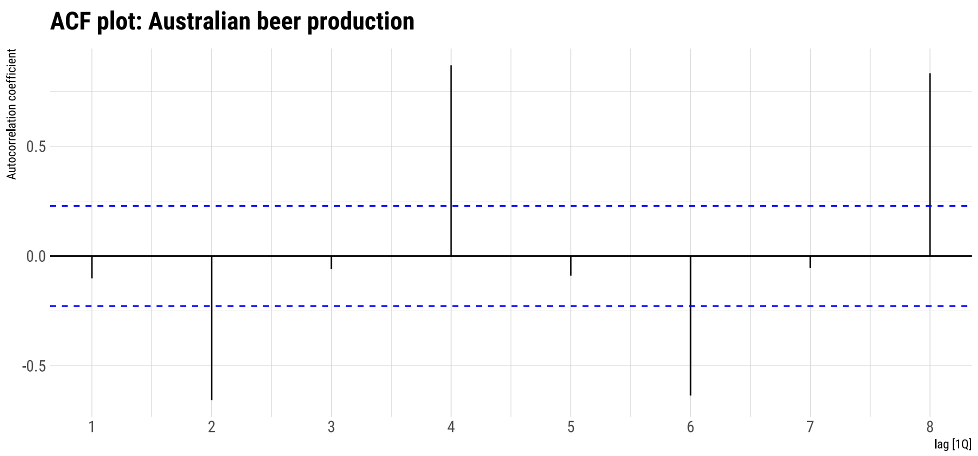
It simply plots together the *values* of the **autocorrelation coefficient** against different *lags*.



The autocorrelation function

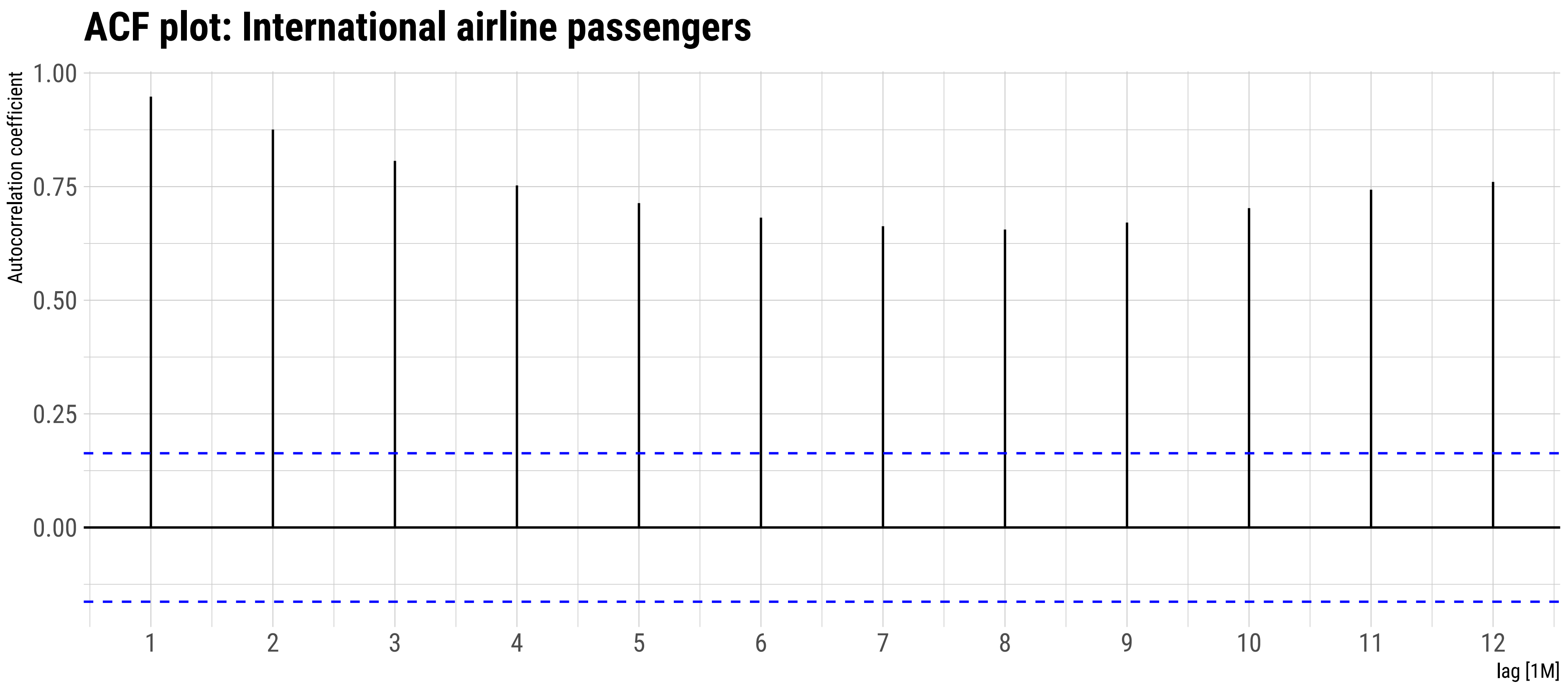
```
beer |>
  ACF(lag_max = 8)
```

```
#> # A tsibble: 8 x 2 [1Q]
#>         lag     acf
#>   <cf_lag>   <dbl>
#> 1        1Q -0.102
#> 2        2Q -0.657
#> 3        3Q -0.0603
#> 4        4Q  0.869
#> 5        5Q -0.0892
#> 6        6Q -0.635
#> 7        7Q -0.0542
#> 8        8Q  0.832
```



The autocorrelation function

An autocorrelation function plot is also known as **correlogram**.



The autocorrelation function

```
air_passengers ▷
  ACF(lag_max = 12)
```

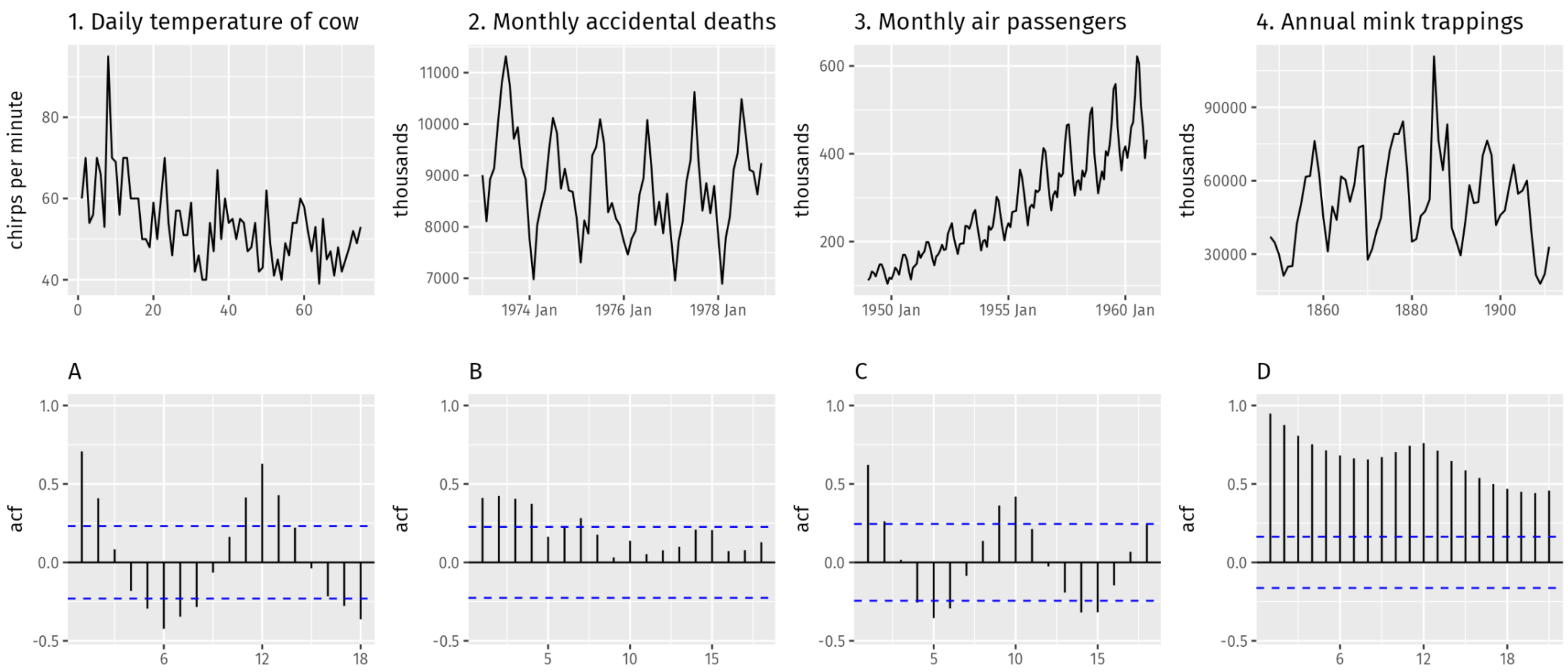
```
#> # A tsibble: 12 x 2 [1M]
#>          lag   acf
#>    <cf_lag> <dbl>
#>  1        1M 0.948
#>  2        2M 0.876
#>  3        3M 0.807
#>  4        4M 0.753
#>  5        5M 0.714
#>  6        6M 0.682
#>  7        7M 0.663
#>  8        8M 0.656
#>  9        9M 0.671
#> 10       10M 0.703
#> 11       11M 0.743
#> 12       12M 0.760
```

The autocorrelation function

The ACF plot may show interesting patterns, depending on the **features** of the time series.

- When data have a **trend**, the autocorrelations for *small* lags tend to be **large** and **positive**.
    - Observations nearby in time are *also* nearby in value!
- When data are **seasonal**, the autocorrelations will be *larger* for the *seasonal* lags (at multiples of the seasonal period) than for other lags.
- When data are both **trended** and **seasonal**, the ACF plot usually shows a **combination** of the above effects.

The autocorrelation function



The autocorrelation function

`A different look at the autocorrelation function, by Alison Horst`

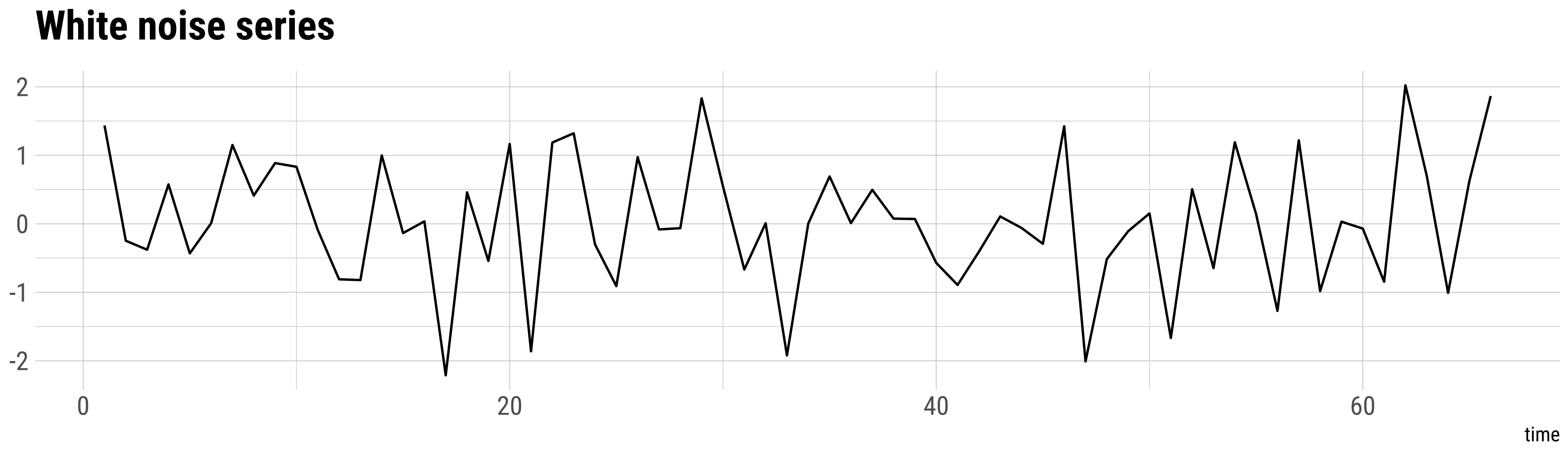
White noise

White noise

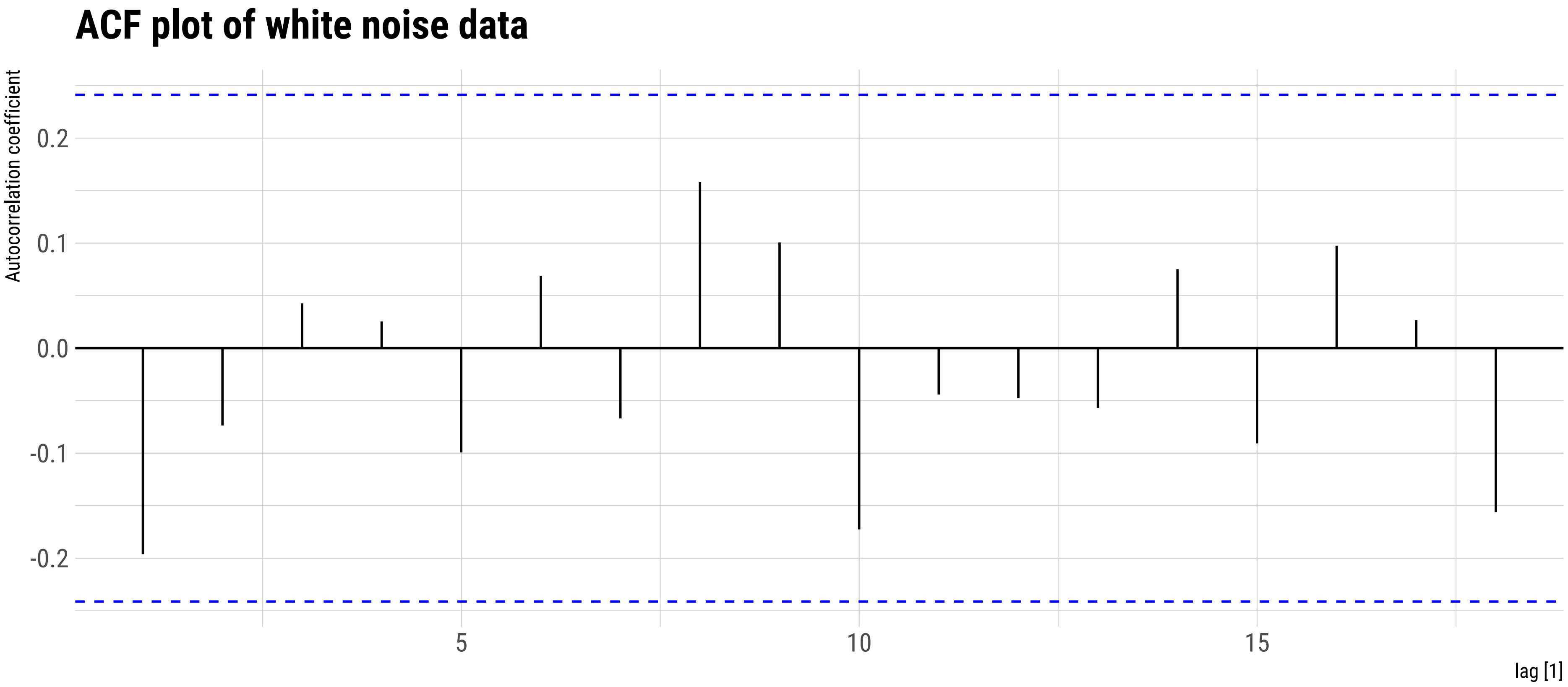
Some time series may *not* show autocorrelation **at all**.

When this is the case, they are called **white noise**.

> A time series w_t is defined to be **white noise** if its observations are *identically and independently distributed* (i.i.d.) with a *mean* of zero and constant *variance*.



White noise


ACF plot of white noise data
Autocorrelation coefficient
0.2
0.1
0.0
-0.1
-0.2
5
10
15
lag [1]

Next time: Time series decomposition