
Main points

1. There are many different factors to bear in mind when drawing and interpreting time series (§11.2).
2. Single time series plots can show a lot of information (e.g., Figure 11.3).
3. Multiple series can be drawn in a single display to make comparisons easy (Figures 11.4 and 11.5).
4. Dates and times have tricky properties and it is best to take advantage of packages that can deal with them (§11.6).
5. Time series are a different kind of data and need to be treated specially. The data are not i.i.d. variables. Graphics are good for irregular time series and for displaying multiple time series. Models have difficulties with both. (§11.4)

Exercises

More detailed information for the datasets is available on their help pages in R.

1. Air miles

The dataset *airmiles* is a time series of the miles flown annually by commercial airlines in the US from 1937 to 1960.

- (a) Before plotting the graph, think about what shape you would expect it to have. Plot the series and comment on the differences between what you get and your expectations.
- (b) Which aspect ratio conveys the information you find in the series best?
- (c) Do you think the graph looks better as a line graph (as suggested on the R help page for the dataset) or with points as well?
- (d) Might plotting a transformation help you to look more closely at the early years or would zooming in be sufficient?

2. Beveridge Wheat Price Index

The Beveridge index of wheat prices covers almost four hundred years of European history from 1500 to 1869 and is available in the dataset *bev* in **tsseries**.

- (a) Plot the series and explain why you have decided to plot it in that way.
- (b) Are there any particular features in the series which stand out? How would you summarise the information in the series in words?
- (c) Many important historical events took place over this time period, including the Thirty Years' War, the English Civil War, and the Napoleonic Wars. Is there any evidence of any of these having an effect on the index?

3. Goals in soccer games

The *Bundesliga* dataset was used in §11.2.

- (a) Plot graphs of the rates of home and away goals per game over the seasons in the same plot. What limits do you recommend for the vertical scale?
- (b) Other possibilities for studying the home and away goal rates per game include plotting the differences or ratios over time and drawing a scatterplot of one rate against another. Is there any information in these graphics that is shown better by one than the others?
- (c) Can you find equivalent data for the top soccer league in your own country and are there similar patterns over the years?

4. Male and female births

Important early demographic analyses were carried out on English data from the seventeenth century. The *Arbuthnot* dataset in the **HistData** package includes data on the numbers of male and female christenings in London from 1629 to 1710.

- (a) Plot the number of male christenings over time. Which features stand out?
- (b) Why do you think there was a low level of christenings from around the mid-1640's to 1660?
- (c) Two low outliers stand out, in 1666, presumably because of the Great Fire of London and the plague the previous year, and in 1704. A possible explanation for the 1704 outlier is given on the R help page for the dataset. Compare the data values for 1674 and 1704 to check the explanation.

5. Goals in soccer games (again)

Consider the numbers of goals scored by each team.

- (a) How would you plot the annual average goals per home game for each team in the Bundesliga over the 46 seasons in the dataset? Would you choose a single graphic or a trellis display? Only one team has been a member of the Bundesliga ever since it started, Hamburg. How do you think the time series of teams with incomplete records should be displayed?
- (b) You could compare the annual home and away scoring rates of particular teams by plotting the two time series on the same display or by drawing a scatterplot of one variable against the other. Using the two teams Hamburg and Bayern Munich, comment on which display you think is better. Do the displays provide different kinds of information?

6. Deaths by horsekick

Plot separate displays for each of the 14 corps in the von Bortkiewicz dataset (*VonBort* in **vcd**).

- (a) Do any of the patterns stand out as different?
- (b) 11 of the 14 corps had no deaths in the first year (1875). Could this be worth looking into?

7. Economics data

The package **ggplot2** includes a dataset of five US economic indicators recorded monthly over about 40 years, *economics*.

- (a) If you plot all five series in one display, is it better to standardise them all at a common value initially or to align them at their means and divide by their standard deviations? What information is shown in the two displays?
- (b) Alternatively you could plot each series separately with its own scale. Do these displays provide additional information and is there any information that was shown in the displays of all series together that is not so easy to see here?

8. Australian rain

The dataset *bomregions* in the **DAAG** package includes seven regional time series of annual rain in Australia and one time series averaged over the country.

- (a) Can all seven regional series be plotted in one display or are individual displays more informative?
- (b) Are there any outliers in the series and do they affect the scales used adversely?
- (c) Is there any evidence of trend in the series? Are there cyclical effects?

9. Tree rings

The package **dplR** includes several tree ring datasets, including *ca533*. There are 34 series of measurements covering 1358 years in all from 626 to 1983. Note that no time variable is given, just the information that the data were recorded annually. The actual time range can be found from NOAA's tree ring database website.

- (a) Plot all 34 series in separate displays. Are there any common features?
- (b) There are at least two series with much higher maxima than the others. Compare a display excluding these series, but still retaining the same scaling for all the plots, with a display where each series is plotted with its own scale. What are the advantages and disadvantages of the two approaches?

10. Intermission

Salvador Dali's painting *The Persistence of Memory* is in the New York Museum of Modern Art. Do you think the distorted clocks could be interpreted as alternative models of time series?