

## 10.1 Hierarchical time series

Figure 10.1 shows a  $K = 2$ -level hierarchical structure. At the top of the hierarchy (which we call level 0) is the “Total,” the most aggregate level of the data. The  $t$ th observation of the Total series is denoted by  $y_t$  for  $t = 1, \dots, T$ . The Total is disaggregated into two series at level 1, which in turn are divided into three and two series respectively at the bottom-level of the hierarchy. Below the top level, we use  $y_{j,t}$  to denote the  $t$ th observation of the series corresponding to node  $j$ . For example,  $y_{A,t}$  denotes the  $t$ th observation of the series corresponding to node A at level 1,  $y_{AB,t}$  denotes the  $t$ th observation of the series corresponding to node AB at level 2, and so on.

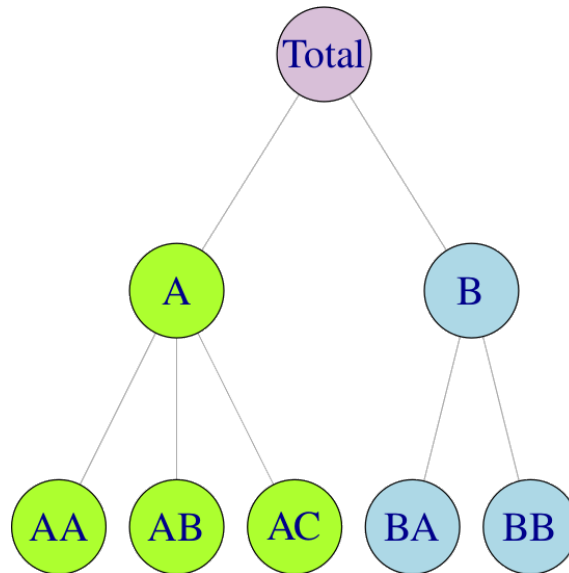


Figure 10.1: A two level hierarchical tree diagram.

In this small example, the total number of series in the hierarchy is  $n = 1 + 2 + 5 = 8$ , while the number of series at the bottom-level is  $m = 5$ . Note that  $n > m$  in all hierarchies.

For any time  $t$ , the observations at the bottom-level of the hierarchy will sum to the observations of the series above. For example,

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t} \quad (10.1)$$

and

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \quad \text{and} \quad y_{B,t} = y_{BA,t} + y_{BB,t}. \quad (10.2)$$

Substituting (10.2) into (10.1), we also get  $y_t = y_{A,t} + y_{B,t}$ . These equations can be thought of as aggregation constraints or summing equalities, and can be more efficiently represented using matrix notation. We construct an  $n \times m$  matrix  $\mathbf{S}$  (referred to as the “summing matrix”) which dictates the way in which the bottom-level series are aggregated.

For the hierarchical structure in Figure 10.1, we can write

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}$$

or in more compact notation

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t, \quad (10.3)$$

where  $\mathbf{y}_t$  is an  $n$ -dimensional vector of all the observations in the hierarchy at time  $t$ ,  $\mathbf{S}$  is the summing matrix, and  $\mathbf{b}_t$  is an  $m$ -dimensional vector of all the observations in the bottom-level of the hierarchy at time  $t$ . Note that the first row in the summing matrix  $\mathbf{S}$  represents Equation (10.1) above, the second and third rows represent (10.2). The rows below these comprise an  $m$ -dimensional identity matrix  $\mathbf{I}_m$  so that each bottom-level observation on the right hand side of the equation is equal to itself on the left hand side.

## Example: Australian tourism hierarchy

Australia is divided into eight geographic areas (some called states and others called territories) with each one having its own government and some economic and administrative autonomy. Each of these can be further subdivided into smaller areas of interest, referred to as zones. Business planners and tourism authorities are interested in forecasts for the whole of Australia, for the states and the territories, and also for the zones. In this example we concentrate on quarterly domestic tourism demand,

measured as the number of visitor nights Australians spend away from home. To simplify our analysis, we combine the two territories and Tasmania into an “Other” state. So we have six states: New South Wales (NSW), Queensland (QLD), South Australia (SAU), Victoria (VIC), Western Australia (WAU) and Other (OTH). For each of these we consider visitor nights within the following zones.

State	Zones
NSW	Metro (NSWMetro), North Coast (NSWNthCo), South Coast (NSWStHCo), South Inner (NSWStHIn), North Inner (NSWNthIn)
QLD	Metro (QLDMetro), Central (QLDCntrl), North Coast (QLDNthCo)
SAU	Metro (SAUMetro), Coastal (SAUCoast), Inner (SAUInner)
VIC	Metro (VICMetro), West Coast (VICWstCo), East Coast (VICEstCo), Inner (VICInner)
WAU	Metro (WAUMetro), Coastal (WAUCoast), Inner (WAUInner)
OTH	Metro (OTHMetro), Non-Metro (OTHNoMet)

We consider five zones for NSW, four zones for VIC, and three zones each for QLD, SAU and WAU. Note that Metro zones contain the capital cities and surrounding areas. For further details on these geographic areas, please refer to Appendix C in [Wickramasuriya, Athanasopoulos, & Hyndman \(2019\)](#).

To create a hierarchical time series, we use the `hts()` function as shown in the code below. The function requires two inputs: the bottom-level time series and information about the hierarchical structure. `visnights` is a time series matrix containing the bottom-level series. There are several ways to input the structure of the hierarchy. In this case we are using the `characters` argument. The first three characters of each column name of `visnights` capture the categories at the first level of the hierarchy (States). The following five characters capture the bottom-level categories (Zones).

```
library(hts)
tourism.hts <- hts(visnights, characters = c(3, 5))
tourism.hts %>% aggts(levels=0:1) %>%
  autoplot(facet=TRUE) +
  xlab("Year") + ylab("millions") + ggtitle("Visitor nights")
```

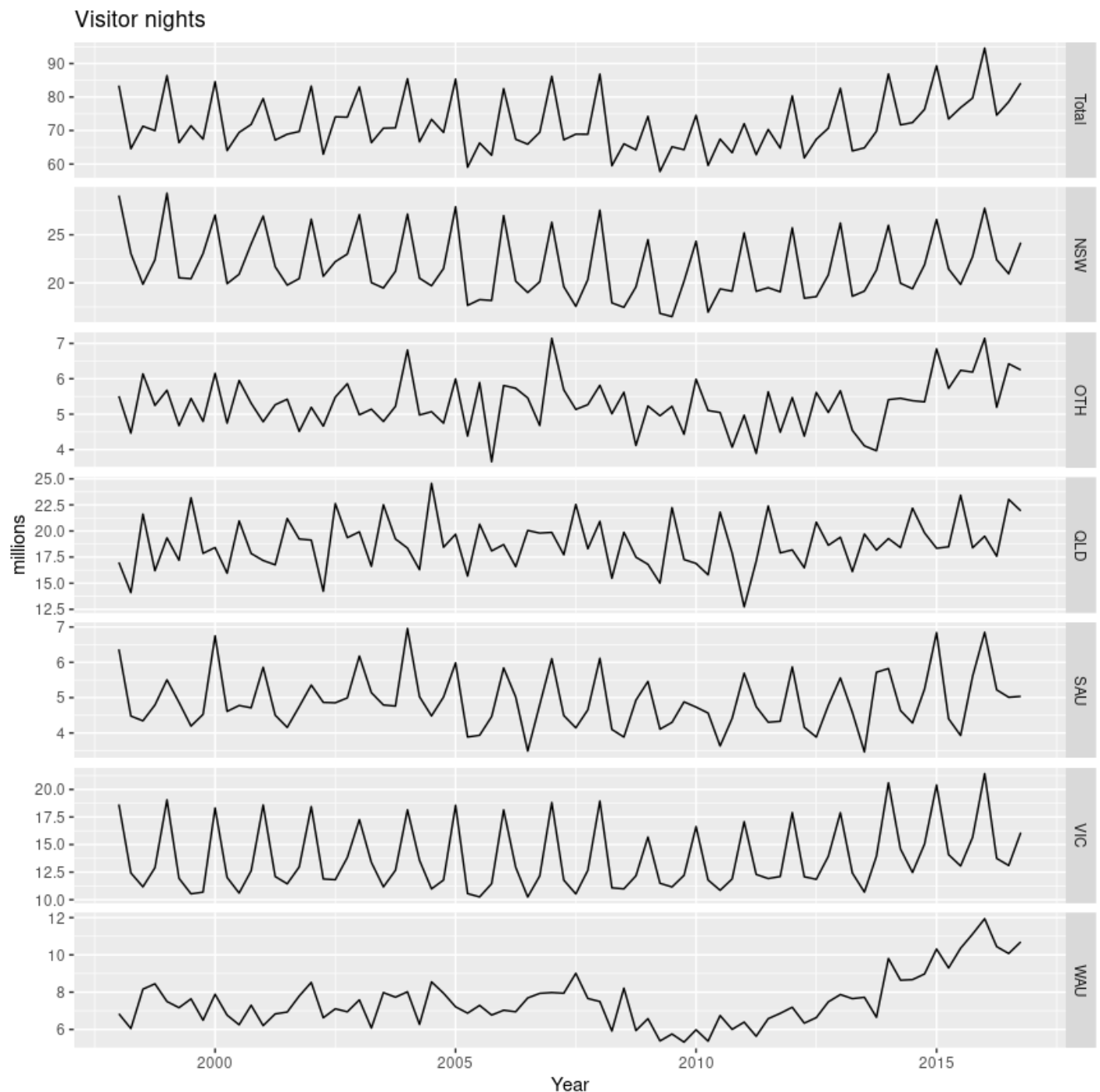


Figure 10.2: Australian domestic visitor nights over the period 1998 Q1 to 2016 Q4 disaggregated by State.

The top plot in Figure 10.2 shows the total number of visitor nights for the whole of Australia, while the plots below show the data disaggregated by state. These reveal diverse and rich dynamics at the aggregate national level, and the first level of disaggregation for each state. The `aggrts()` function extracts time series from an `hts` object for any level of aggregation.

The plots in Figure 10.3 show the bottom-level time series, namely the visitor nights for each zone. These help us visualise the diverse individual dynamics within each zone, and assist in identifying unique and important time series. Notice, for example, the coastal WAU zone which shows significant growth over the last few years.

```
library(tidyverse)
cols <- sample(scales::hue_pal(h=c(15,375),
  c=100,l=65,h.start=0,direction = 1)(NCOL(visnights)))
as_tibble(visnights) %>%
  gather(Zone) %>%
  mutate(Date = rep(time(visnights), NCOL(visnights)),
    State = str_sub(Zone,1,3)) %>%
  ggplot(aes(x=Date, y=value, group=Zone, colour=Zone)) +
    geom_line() +
    facet_grid(State~., scales="free_y") +
    xlab("Year") + ylab("millions") +
    ggtitle("Visitor nights by Zone") +
    scale_colour_manual(values = cols)
```

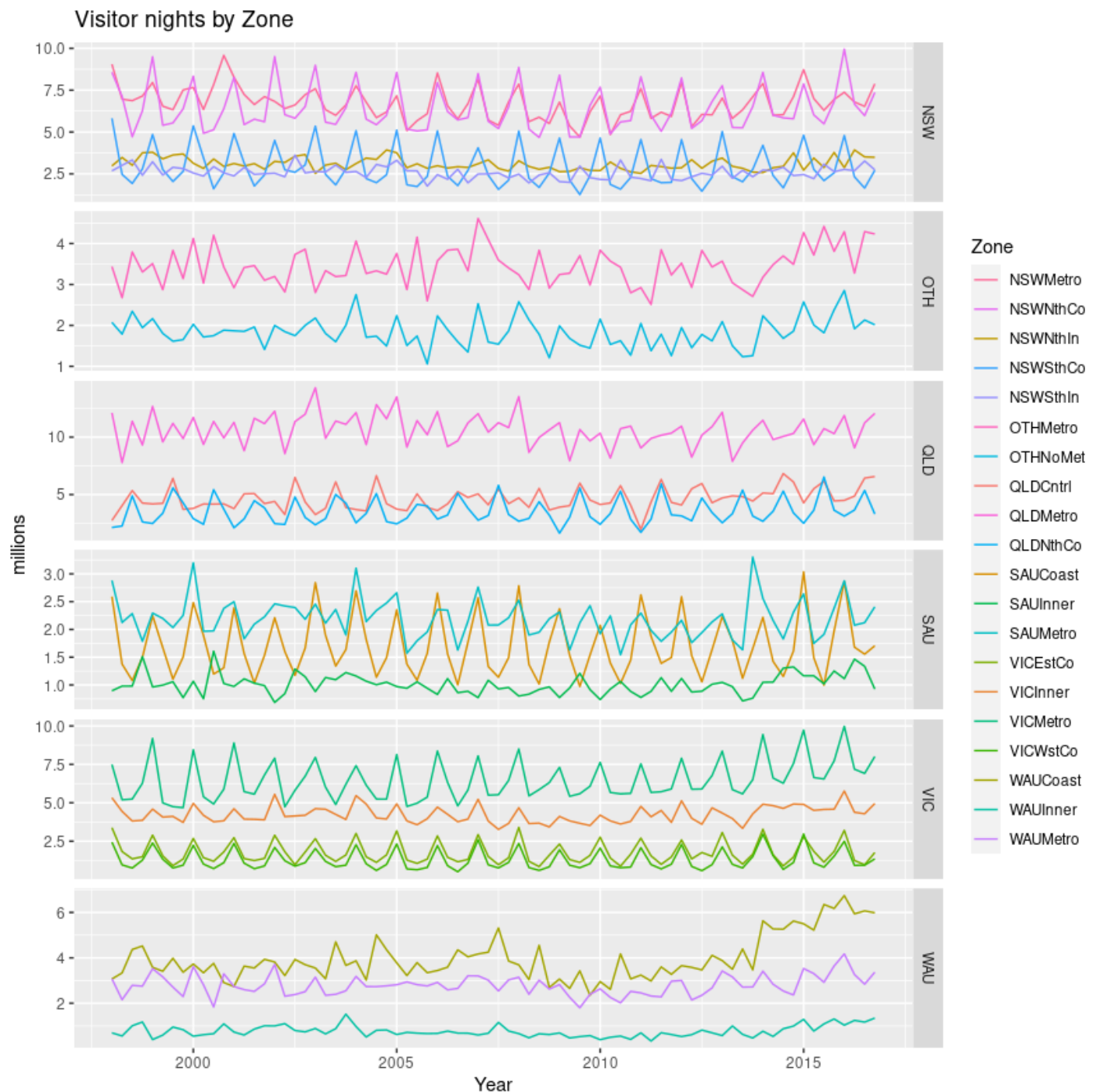


Figure 10.3: Australian domestic visitor nights over the period 1998 Q1 to 2016 Q4 disaggregated by Zones.

To produce this figure, we are using various functions from the [tidyverse collection of packages](#). The details are beyond the scope of this book, but there are many good online resources available to learn how to use these packages.