Time Series Practice

Abhishek Gupta

12/08/2020

```r
#tinytex::install_tinytex()
```

```r
library(tidyverse)
```

```
## -- Attaching packages -------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ----------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggfortify)
library(fpp3)
```

```
## -- Attaching packages ----------------------------------- fpp3 0.3 --
```

```
## v lubridate   1.7.9     v feasts      0.1.5
## v tsibble     0.9.3     v fable       0.2.1
## v tsibbledata 0.2.0
```

```
## -- Conflicts ------------------------------------------ fpp3_conflicts --
## x lubridate::date()   masks base::date()
## x dplyr::filter()     masks stats::filter()
## x tsibble::interval() masks lubridate::interval()
## x dplyr::lag()        masks stats::lag()
```

```r
library(fpp2)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method           from
##   as.zoo.data.frame zoo
```

```
## Registered S3 methods overwritten by 'forecast':
##   method                 from
##   autoplot.Arima         ggfortify
##   autoplot.acf           ggfortify
##   autoplot.ar            ggfortify
##   autoplot.bats          ggfortify
##   autoplot.decomposed.ts ggfortify
##   autoplot.ets           ggfortify
##   autoplot.forecast      ggfortify
##   autoplot.stl           ggfortify
##   autoplot.ts            ggfortify
##   fitted.ar              ggfortify
##   fortify.ts             ggfortify
##   residuals.ar           ggfortify


## -- Attaching packages --------------------------------------- fpp2 2.4 --


## v forecast  8.13     v expsmooth 2.3
## v fma       2.4


##
```

```r
library(tsibble)
knitr::opts_chunk$set(echo = TRUE)
```

Time Series Practice Session :

Let's learn time series forecasting in R.

First Time Series :

```
## 2012 2013 2014 2015 2016
##   43   54   87   21   42
```

```
## 2013 2014 2015 2016 2017 2018
##   75   78   69    1   12   39
```

Airline Data

```
## [1] "First.Class"    "Business.Class" "Economy.Class"
```
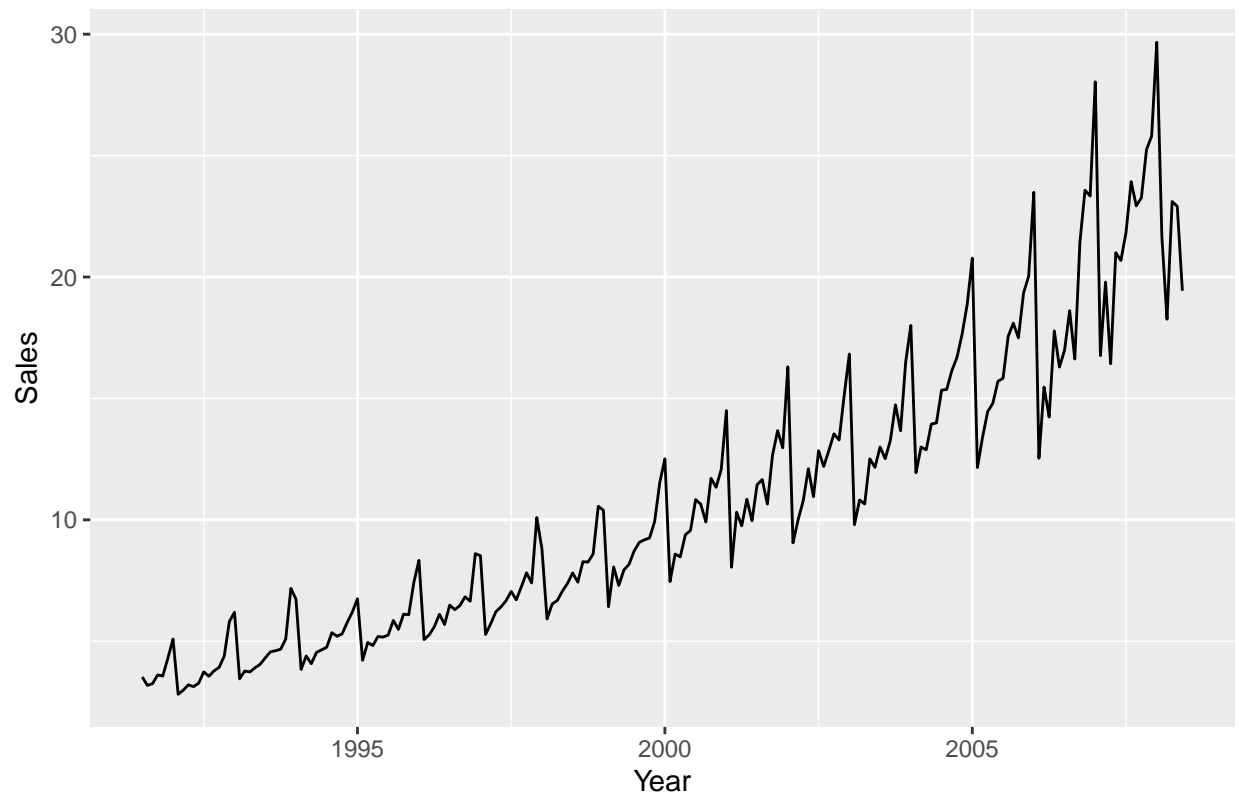
Melbourne–Sydney Economy Air Passengers

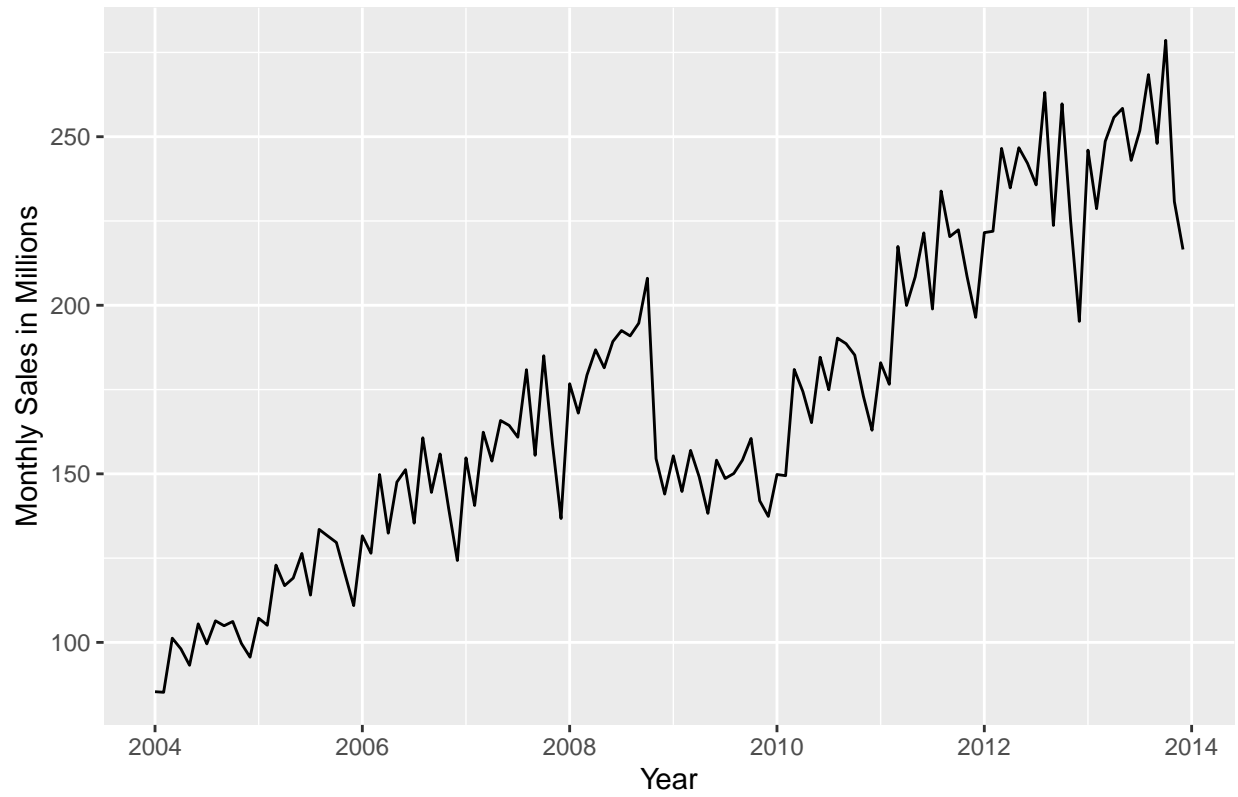**Antibody Sales Data**

```
## [1] "Date"  "Time"  "Month" "Year"  "Sales"
```

Antibiotics sales each month
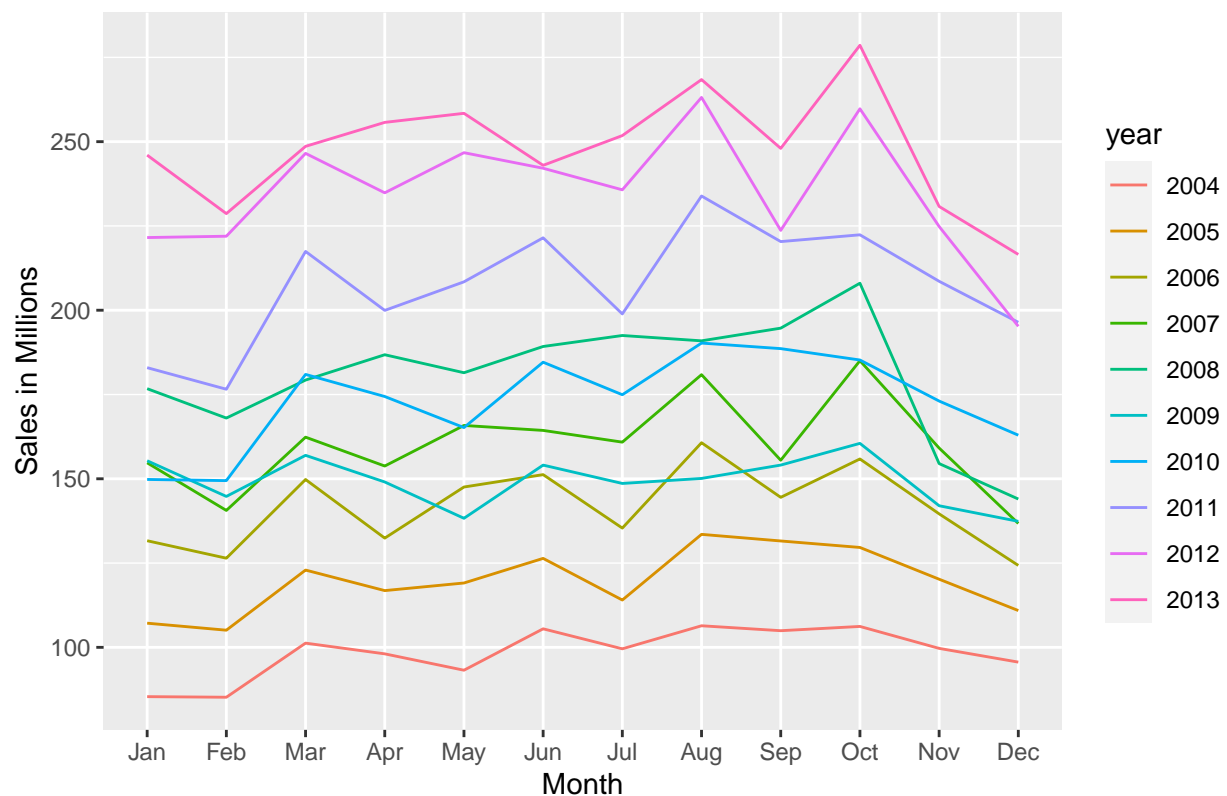


```
## [1] "Time"            "Month"          "Month.Num"       "Year"
## [5] "NumBDays"        "AvSalesPD"      "Total.Sales"     "Total.Fastner"
## [9] "Total.Nonfastner"
```

Total Monthly Sales (2004–2013)

Trends, Seasonality and Cyclicity

Seasonal plot: TotSales

Switching to the 3rd Edition

```
y <-  tsibble(Year = 2015:2019, Observation = c(123,39,78,52,110), index = Year)
y
```

```
## # A tsibble: 5 x 2 [1Y]
##     Year Observation
##    <int>        <dbl>
## 1  2015         123
## 2  2016          39
## 3  2017          78
## 4  2018          52
## 5  2019         110
```

```
 w <-  tsibble(Month=yearmonth("2015 May") + 0:4, Observation = c(123,39,78,52,110), index = Month)
w
```

```
## # A tsibble: 5 x 2 [1M]
##       Month Observation
##       <mth>        <dbl>
## 1 2015 May         123
```

```
## 2 2015 Jun          39
## 3 2015 Jul          78
## 4 2015 Aug          52
## 5 2015 Sep         110
```

olympic_running

```
## # A tsibble: 312 x 4 [4Y]
## # Key:        Length, Sex [14]
##     Year Length Sex   Time
##    <int>  <int> <chr> <dbl>
## 1  1896    100 men     12
## 2  1900    100 men     11
## 3  1904    100 men     11
## 4  1908    100 men     10.8
## 5  1912    100 men     10.8
## 6  1916    100 men     NA
## 7  1920    100 men     10.8
## 8  1924    100 men     10.6
## 9  1928    100 men     10.8
## 10 1932    100 men     10.3
## # ... with 302 more rows
```

PBS

```
## # A tsibble: 65,219 x 9 [1M]
## # Key:        Concession, Type, ATC1, ATC2 [336]
##       Month Concession  Type    ATC1  ATC1_desc   ATC2  ATC2_desc    Scripts  Cost
##       <mth> <chr>       <chr>   <chr> <chr>       <chr> <chr>          <dbl> <dbl>
## 1  1991 Jul Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    18228 67877
## 2  1991 Aug Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    15327 57011
## 3  1991 Sep Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    14775 55020
## 4  1991 Oct Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    15380 57222
## 5  1991 Nov Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    14371 52120
## 6  1991 Dec Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    15028 54299
## 7  1992 Jan Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    11040 39753
## 8  1992 Feb Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    15165 54405
## 9  1992 Mar Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    16898 61108
## 10 1992 Apr Concession~ Co-pa~  A     Alimentary~ A01   STOMATOLOG~    18141 65356
## # ... with 65,209 more rows
```

PBS **%>%** **filter**(ATC2 **==** 'A10')

```
## # A tsibble: 816 x 9 [1M]
## # Key:        Concession, Type, ATC1, ATC2 [4]
##       Month Concession  Type    ATC1  ATC1_desc   ATC2  ATC2_desc  Scripts   Cost
##       <mth> <chr>       <chr>   <chr> <chr>       <chr> <chr>        <dbl>  <dbl>
## 1  1991 Jul Concession~ Co-pa~  A     Alimentary~ A10   ANTIDIABE~   89733 2.09e6
## 2  1991 Aug Concession~ Co-pa~  A     Alimentary~ A10   ANTIDIABE~   77101 1.80e6
## 3  1991 Sep Concession~ Co-pa~  A     Alimentary~ A10   ANTIDIABE~   76255 1.78e6
## 4  1991 Oct Concession~ Co-pa~  A     Alimentary~ A10   ANTIDIABE~   78681 1.85e6
## 5  1991 Nov Concession~ Co-pa~  A     Alimentary~ A10   ANTIDIABE~   70554 1.69e6
```

```
##  6 1991 Dec Concession~ Co-pa~ A      Alimentary~ A10   ANTIDIABE~   75814 1.84e6
##  7 1992 Jan Concession~ Co-pa~ A      Alimentary~ A10   ANTIDIABE~   64186 1.56e6
##  8 1992 Feb Concession~ Co-pa~ A      Alimentary~ A10   ANTIDIABE~   75899 1.73e6
##  9 1992 Mar Concession~ Co-pa~ A      Alimentary~ A10   ANTIDIABE~   89445 2.05e6
## 10 1992 Apr Concession~ Co-pa~ A      Alimentary~ A10   ANTIDIABE~   97315 2.23e6
## # ... with 806 more rows
```

```
PBS %>%
  filter(ATC2=="A10") %>%
  select(Month, Concession, Type, Cost)
```

```
## # A tsibble: 816 x 4 [1M]
## # Key:       Concession, Type [4]
##        Month Concession   Type            Cost
##        <mth> <chr>        <chr>          <dbl>
##  1 1991 Jul Concessional Co-payments 2092878
##  2 1991 Aug Concessional Co-payments 1795733
##  3 1991 Sep Concessional Co-payments 1777231
##  4 1991 Oct Concessional Co-payments 1848507
##  5 1991 Nov Concessional Co-payments 1686458
##  6 1991 Dec Concessional Co-payments 1843079
##  7 1992 Jan Concessional Co-payments 1564702
##  8 1992 Feb Concessional Co-payments 1732508
##  9 1992 Mar Concessional Co-payments 2046102
## 10 1992 Apr Concessional Co-payments 2225977
## # ... with 806 more rows
```

```
PBS %>%
  filter(ATC2=="A10") %>%
  select(Month, Concession, Type, Cost) %>%
  summarise(TotalC = sum(Cost))
```

```
## # A tsibble: 204 x 2 [1M]
##        Month  TotalC
##        <mth>   <dbl>
##  1 1991 Jul 3526591
##  2 1991 Aug 3180891
##  3 1991 Sep 3252221
##  4 1991 Oct 3611003
##  5 1991 Nov 3565869
##  6 1991 Dec 4306371
##  7 1992 Jan 5088335
##  8 1992 Feb 2814520
##  9 1992 Mar 2985811
## 10 1992 Apr 3204780
## # ... with 194 more rows
```

```
PBS %>%
  filter(ATC2=="A10") %>%
  select(Month, Concession, Type, Cost) %>%
  summarise(TotalC = sum(Cost)) %>%
  mutate(Cost = TotalC/1e6)
```

```
## # A tsibble: 204 x 3 [1M]
##        Month  TotalC   Cost
##        <mth>   <dbl>  <dbl>
##  1 1991 Jul 3526591   3.53
##  2 1991 Aug 3180891   3.18
##  3 1991 Sep 3252221   3.25
##  4 1991 Oct 3611003   3.61
##  5 1991 Nov 3565869   3.57
##  6 1991 Dec 4306371   4.31
##  7 1992 Jan 5088335   5.09
##  8 1992 Feb 2814520   2.81
##  9 1992 Mar 2985811   2.99
## 10 1992 Apr 3204780   3.20
## # ... with 194 more rows
```

```r
PBS %>%
  filter(ATC2=="A10") %>%
  select(Month, Concession, Type, Cost) %>%
  summarise(TotalC = sum(Cost)) %>%
  mutate(Cost = TotalC/1e6) -> a10
```

```r
prison <- readr::read_csv("https://OTexts.com/fpp3/extrafiles/prison_population.csv")
```

```
##
## -- Column specification -------------------------------------------------
## cols(
##   Date = col_date(format = ""),
##   State = col_character(),
##   Gender = col_character(),
##   Legal = col_character(),
##   Indigenous = col_character(),
##   Count = col_double()
## )
```

```r
prison
```

```
## # A tibble: 3,072 x 6
##    Date       State Gender Legal     Indigenous Count
##    <date>     <chr> <chr>  <chr>     <chr>      <dbl>
##  1 2005-03-01 ACT   Female Remanded  ATSI           0
##  2 2005-03-01 ACT   Female Remanded  Non-ATSI       2
##  3 2005-03-01 ACT   Female Sentenced ATSI           0
##  4 2005-03-01 ACT   Female Sentenced Non-ATSI       5
##  5 2005-03-01 ACT   Male   Remanded  ATSI           7
##  6 2005-03-01 ACT   Male   Remanded  Non-ATSI      58
##  7 2005-03-01 ACT   Male   Sentenced ATSI           5
##  8 2005-03-01 ACT   Male   Sentenced Non-ATSI     101
##  9 2005-03-01 NSW   Female Remanded  ATSI          51
## 10 2005-03-01 NSW   Female Remanded  Non-ATSI     131
## # ... with 3,062 more rows
```

```
prison <- prison %>%
  mutate(Quarter = yearquarter(Date)) %>%
  select(-Date) %>%
  as_tsibble(key = c(State, Gender, Legal, Indigenous), index = Quarter)

prison
```

```
## # A tsibble: 3,072 x 6 [1Q]
## # Key:        State, Gender, Legal, Indigenous [64]
##    State Gender Legal     Indigenous Count Quarter
##    <chr> <chr> <chr>      <chr>      <dbl>   <qtr>
##  1 ACT   Female Remanded  ATSI           0 2005 Q1
##  2 ACT   Female Remanded  ATSI           1 2005 Q2
##  3 ACT   Female Remanded  ATSI           0 2005 Q3
##  4 ACT   Female Remanded  ATSI           0 2005 Q4
##  5 ACT   Female Remanded  ATSI           1 2006 Q1
##  6 ACT   Female Remanded  ATSI           1 2006 Q2
##  7 ACT   Female Remanded  ATSI           1 2006 Q3
##  8 ACT   Female Remanded  ATSI           0 2006 Q4
##  9 ACT   Female Remanded  ATSI           0 2007 Q1
## 10 ACT   Female Remanded  ATSI           1 2007 Q2
## # ... with 3,062 more rows
```

Time Plots

```
melsyd_economy <- ansett %>%
  filter(Airports == "MEL-SYD", Class=="Economy")
melsyd_economy %>%
  autoplot(Passengers) +
    labs(title = "Ansett economy class passengers", subtitle = "Melbourne-Sydney") +
    xlab("Year")
```

Ansett economy class passengers
Melbourne−Sydney
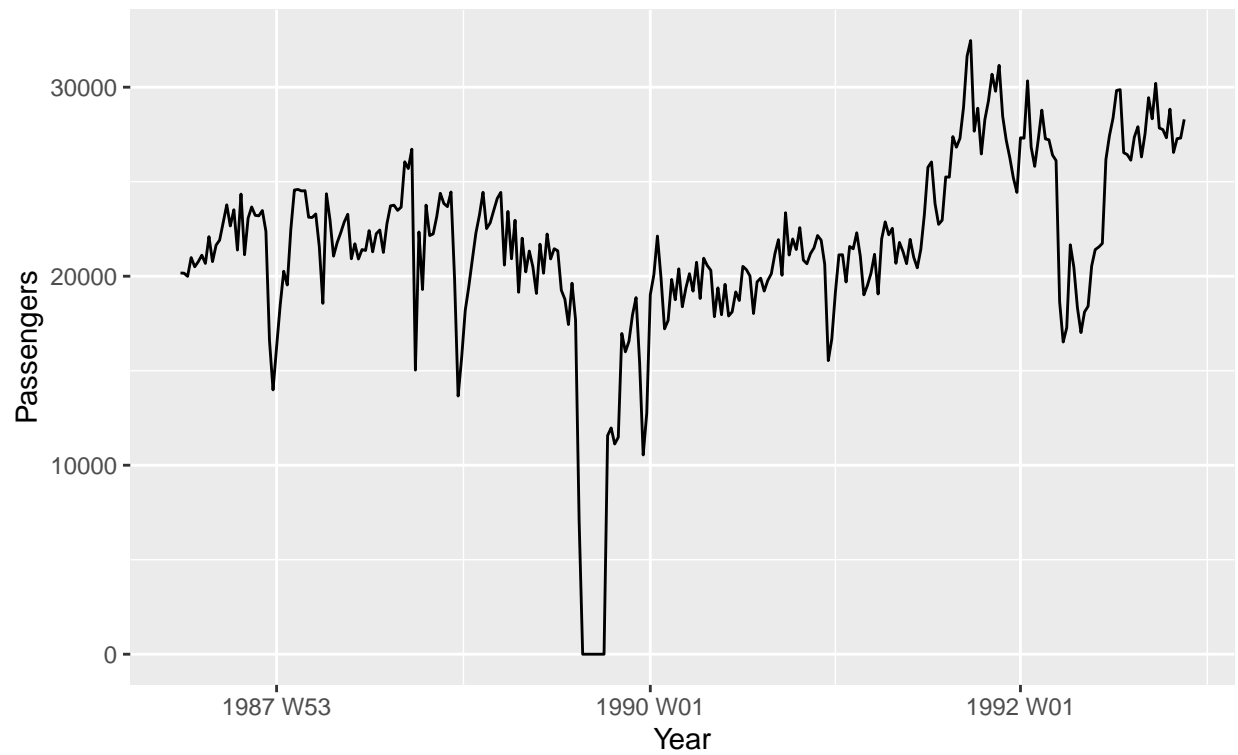

```r
a10 %>% autoplot(Cost) +
  ggtitle("Antidiabetic drug sales") +
  ylab("$ million") + xlab("Year")
```
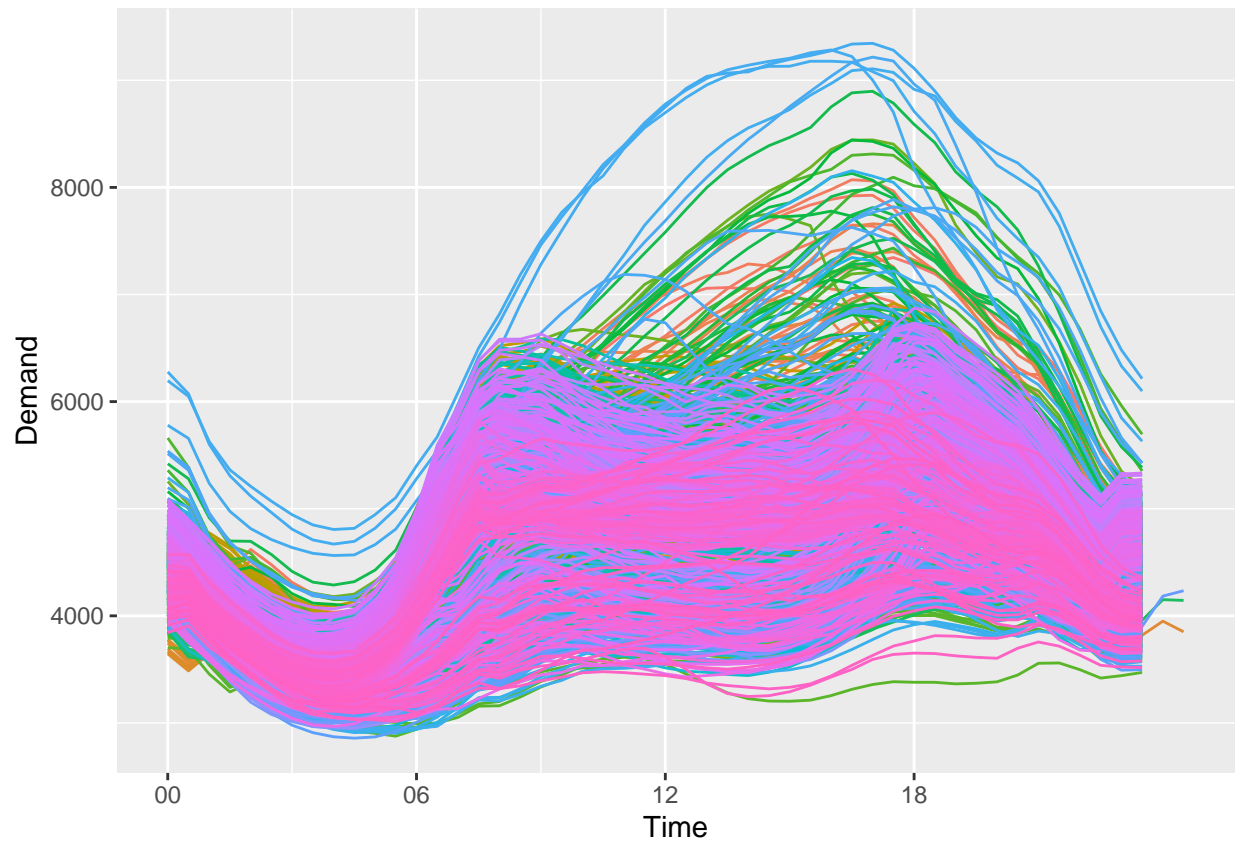
Antidiabetic drug sales
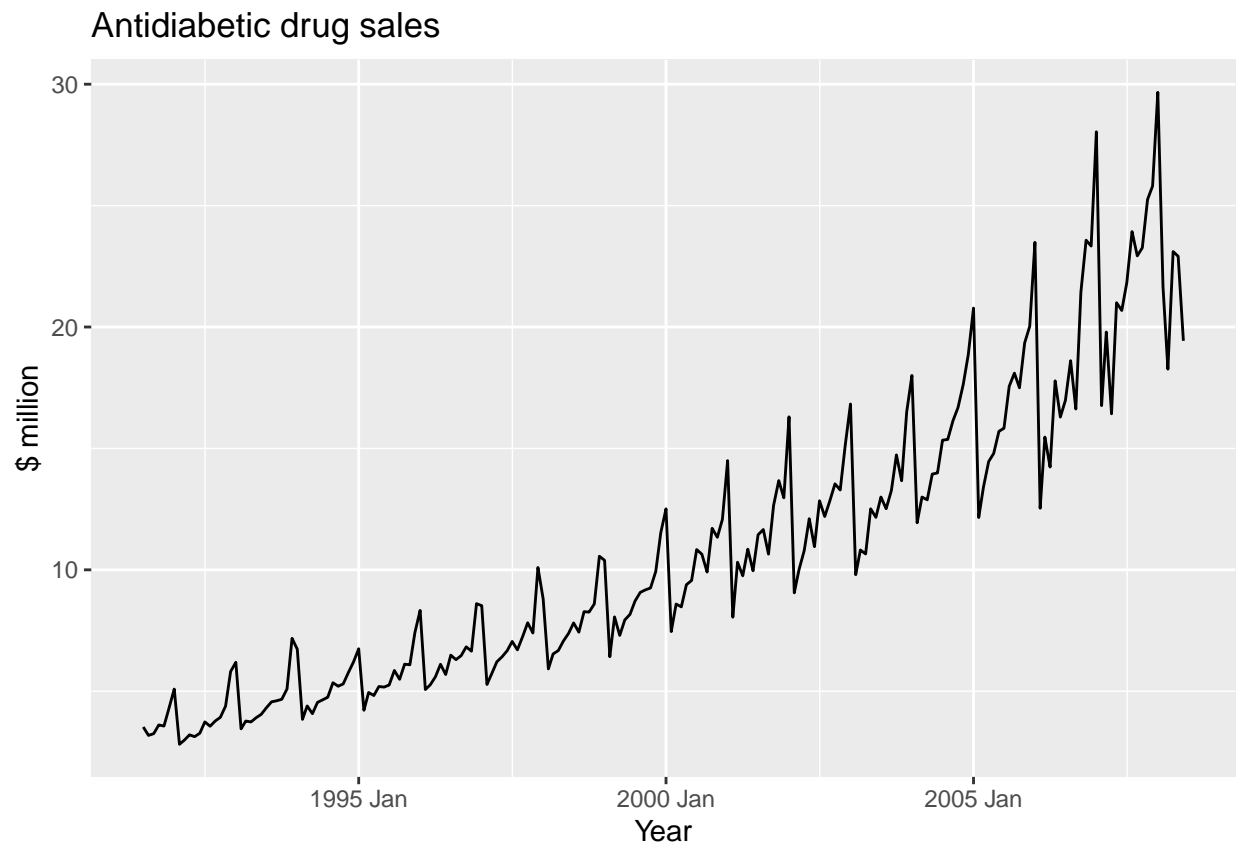

```
a10 %>% gg_season(Cost, labels = "both") +
  ylab("$ million") +
  ggtitle("Seasonal plot: antidiabetic drug sales")
```
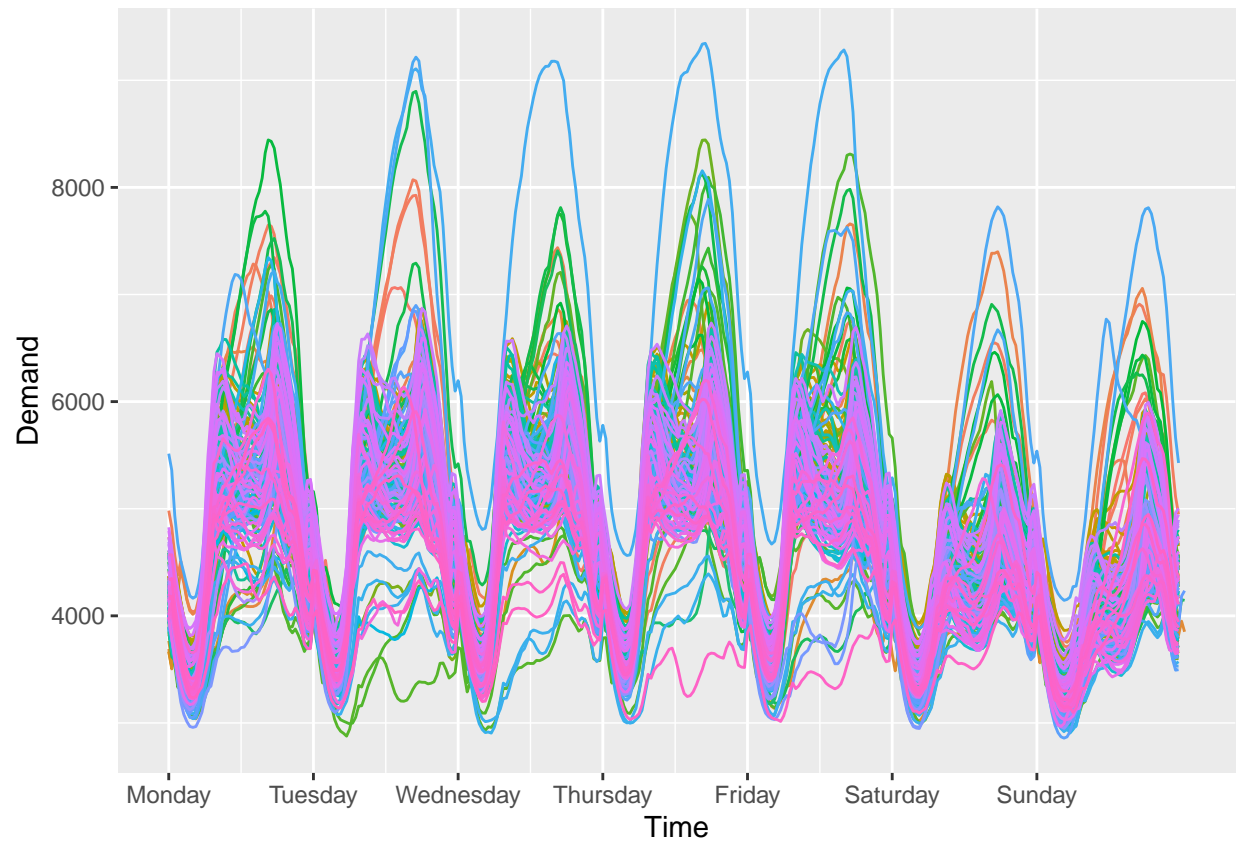
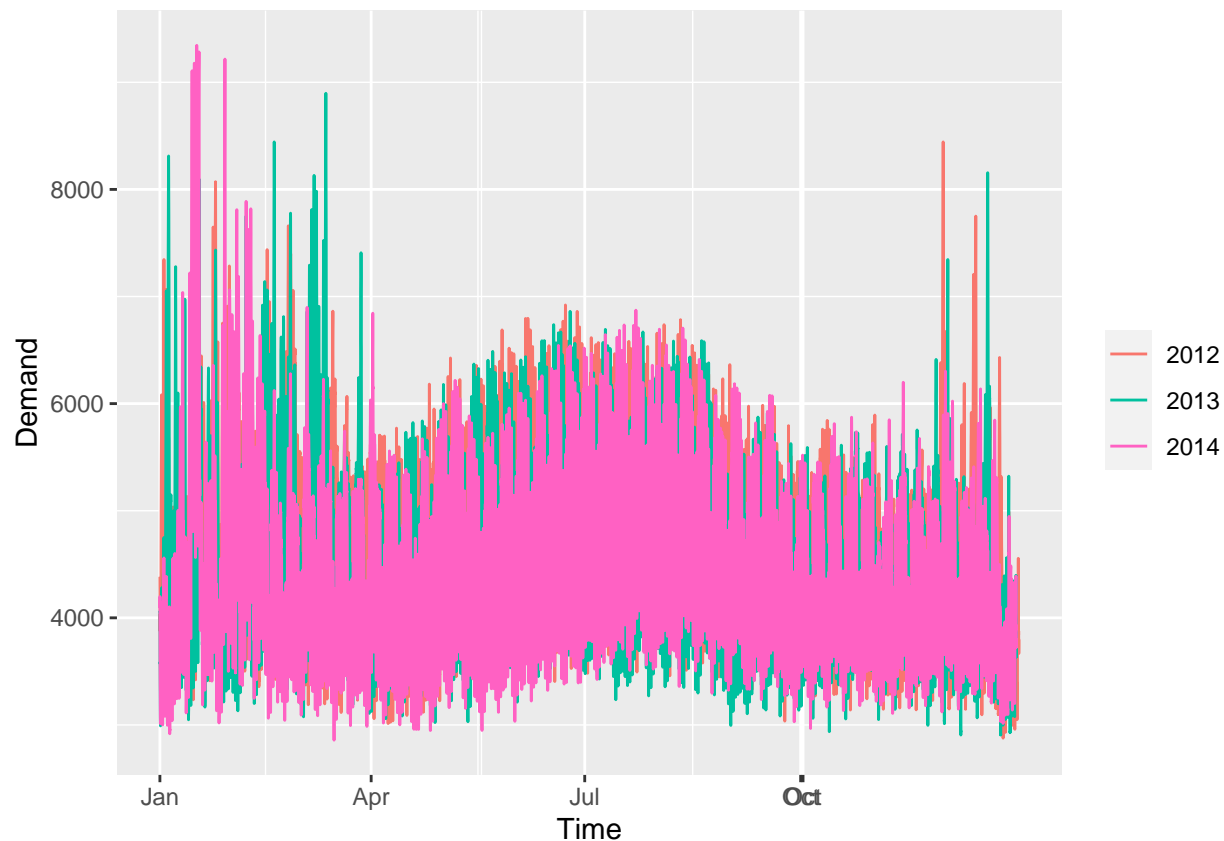Seasonal plot: antidiabetic drug sales
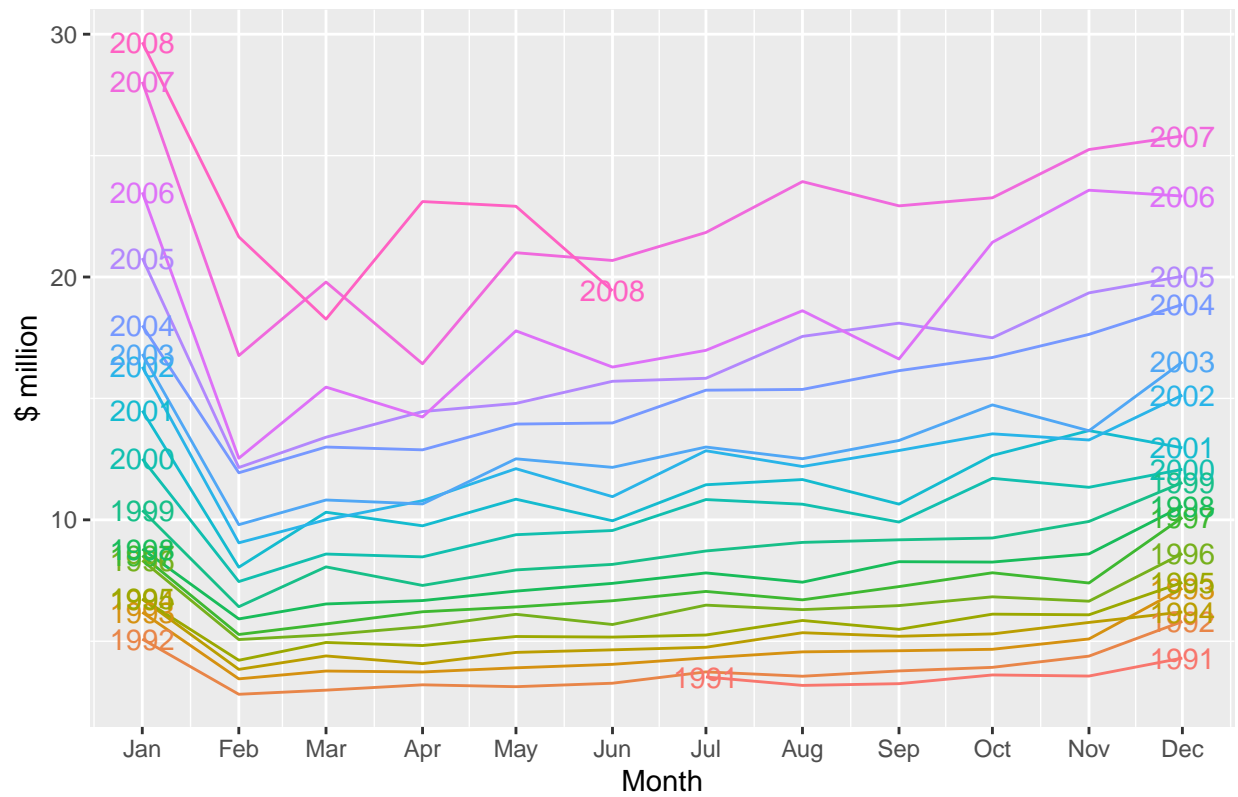
```
vic_elec %>% gg_season(Demand, period="day") + theme(legend.position = "none")
```

```
vic_elec %>% gg_season(Demand, period="week") + theme(legend.position = "none")
```

```
vic_elec %>% gg_season(Demand, period="year")
```

```
vic_elec
```

```
## # A tsibble: 52,608 x 5 [30m] <Australia/Melbourne>
##    Time                Demand Temperature Date       Holiday
##    <dttm>               <dbl>       <dbl> <date>     <lgl>
##  1 2012-01-01 00:00:00  4383.        21.4 2012-01-01 TRUE
##  2 2012-01-01 00:30:00  4263.        21.0 2012-01-01 TRUE
##  3 2012-01-01 01:00:00  4049.        20.7 2012-01-01 TRUE
##  4 2012-01-01 01:30:00  3878.        20.6 2012-01-01 TRUE
##  5 2012-01-01 02:00:00  4036.        20.4 2012-01-01 TRUE
##  6 2012-01-01 02:30:00  3866.        20.2 2012-01-01 TRUE
##  7 2012-01-01 03:00:00  3694.        20.1 2012-01-01 TRUE
##  8 2012-01-01 03:30:00  3562.        19.6 2012-01-01 TRUE
##  9 2012-01-01 04:00:00  3433.        19.1 2012-01-01 TRUE
## 10 2012-01-01 04:30:00  3359.        19.0 2012-01-01 TRUE
## # ... with 52,598 more rows
```

```
a10 %>%
  gg_subseries(Cost) +
    ylab("$ million") +
    xlab("Year") +
    ggtitle("Seasonal subseries plot: antidiabetic drug sales")
```

16

Seasonal subseries plot: antidiabetic drug sales


```
tourism
```

```
## # A tsibble: 24,320 x 5 [1Q]
## # Key:       Region, State, Purpose [304]
##      Quarter Region    State            Purpose  Trips
##        <qtr> <chr>     <chr>            <chr>    <dbl>
##  1 1998 Q1 Adelaide South Australia Business   135.
##  2 1998 Q2 Adelaide South Australia Business   110.
##  3 1998 Q3 Adelaide South Australia Business   166.
##  4 1998 Q4 Adelaide South Australia Business   127.
##  5 1999 Q1 Adelaide South Australia Business   137.
##  6 1999 Q2 Adelaide South Australia Business   200.
##  7 1999 Q3 Adelaide South Australia Business   169.
##  8 1999 Q4 Adelaide South Australia Business   134.
##  9 2000 Q1 Adelaide South Australia Business   154.
## 10 2000 Q2 Adelaide South Australia Business   169.
## # ... with 24,310 more rows
```

```r
holidays <- tourism %>%
  filter(Purpose == "Holiday") %>%
  group_by(State) %>%
  summarise(Trips = sum(Trips))

holidays
```

```
## # A tsibble: 640 x 3 [1Q]
```

```
## # Key:       State [8]
##    State Quarter Trips
##    <chr>   <qtr> <dbl>
##  1 ACT   1998 Q1  196.
##  2 ACT   1998 Q2  127.
##  3 ACT   1998 Q3  111.
##  4 ACT   1998 Q4  170.
##  5 ACT   1999 Q1  108.
##  6 ACT   1999 Q2  125.
##  7 ACT   1999 Q3  178.
##  8 ACT   1999 Q4  218.
##  9 ACT   2000 Q1  158.
## 10 ACT   2000 Q2  155.
## # ... with 630 more rows
```

```r
holidays %>% autoplot(Trips) +
ylab("thousands of trips") + xlab("Year") +
ggtitle("Australian domestic holiday nights")
```
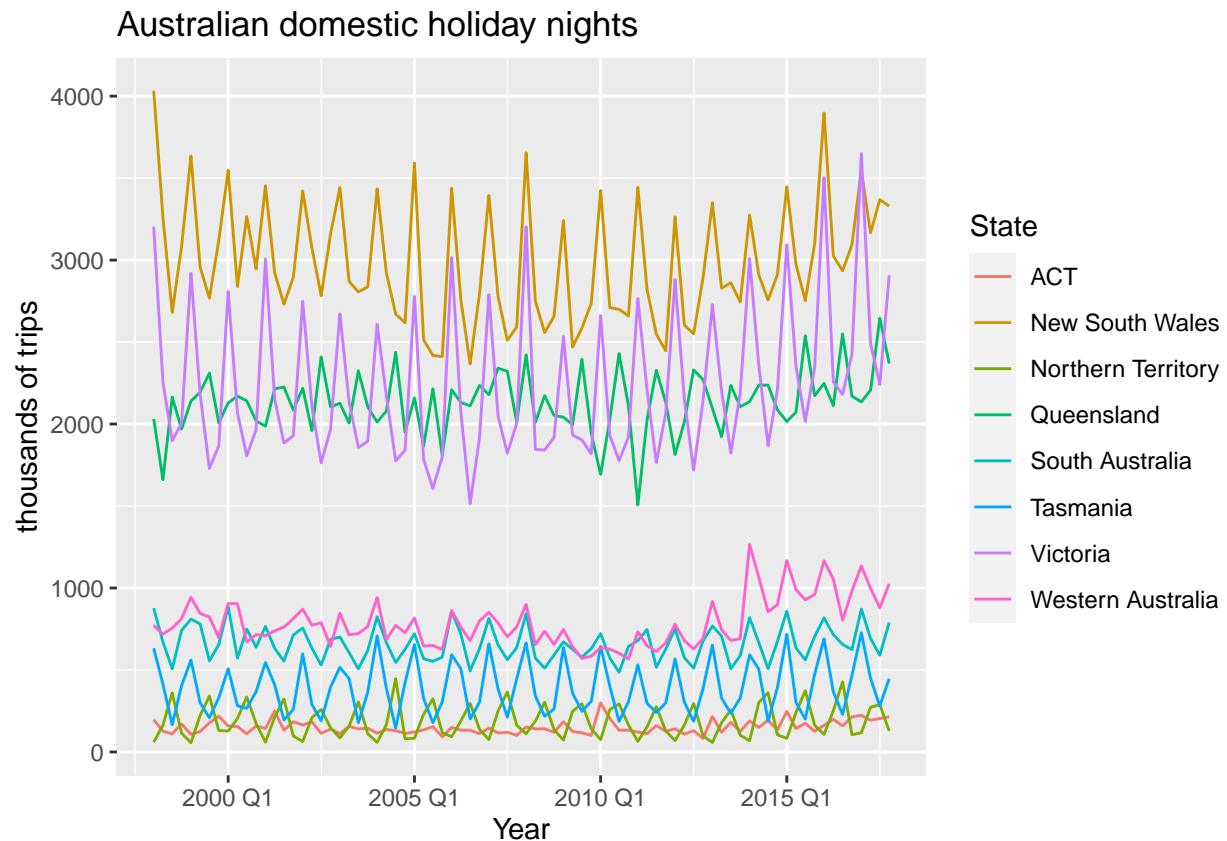


```r
holidays %>% gg_season(Trips) +
  ylab("thousands of trips") +
  ggtitle("Australian domestic holiday nights")
```

Australian domestic holiday nights
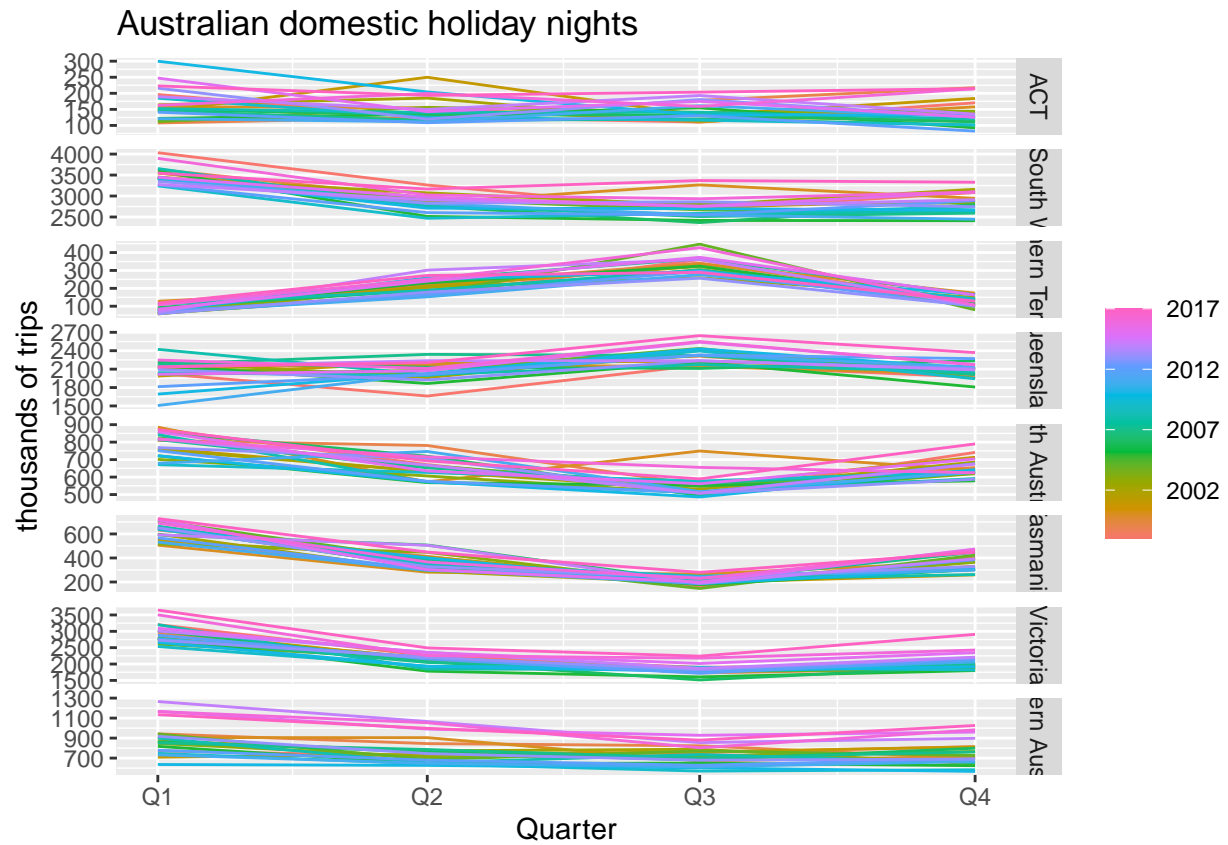

```
holidays %>%
  gg_subseries(Trips) + ylab("thousands of trips") +
  ggtitle("Australian domestic holiday nights")
```

Australian domestic holiday nights
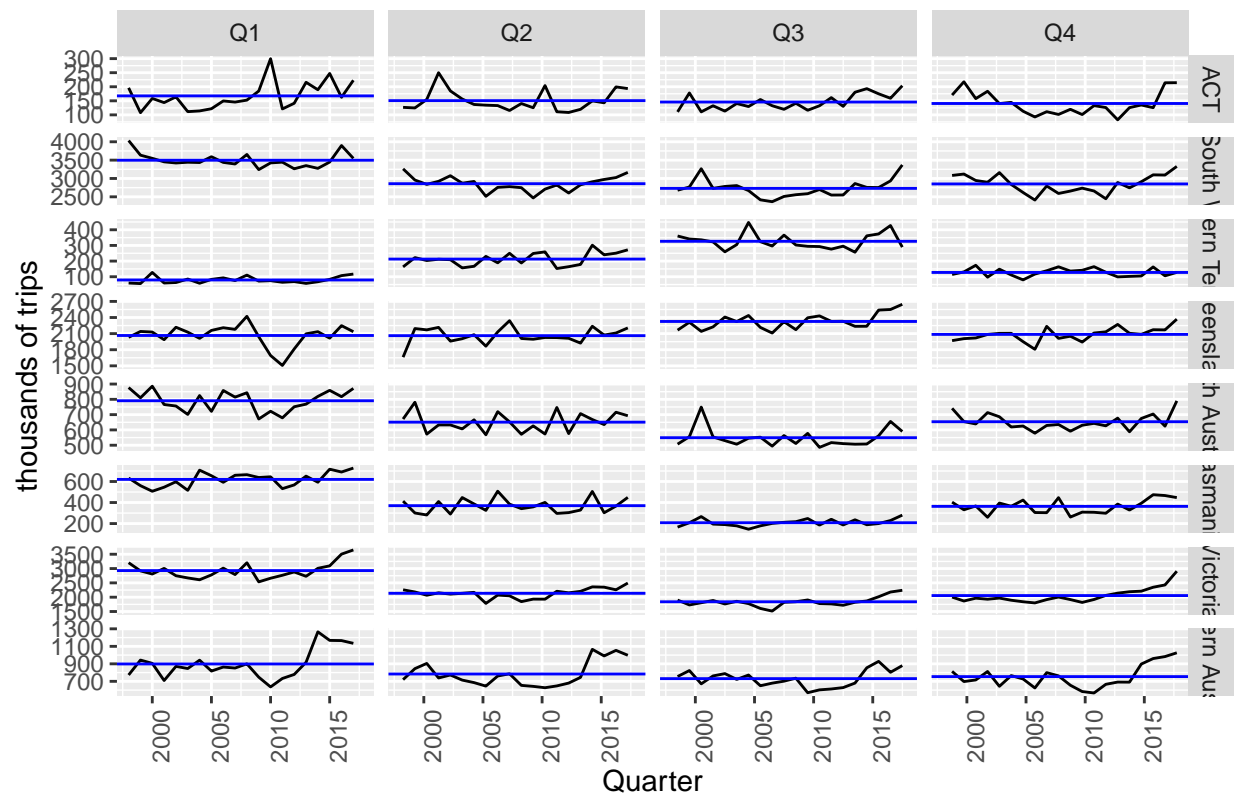
Scatter Plot

```
vic_elec %>%
  filter(year(Time) == 2014) %>%
  autoplot(Demand) +
    xlab("Year: 2014") + ylab(NULL) +
    ggtitle("Half-hourly electricity demand: Victoria, Australia")
```

Half−hourly electricity demand: Victoria, Australia
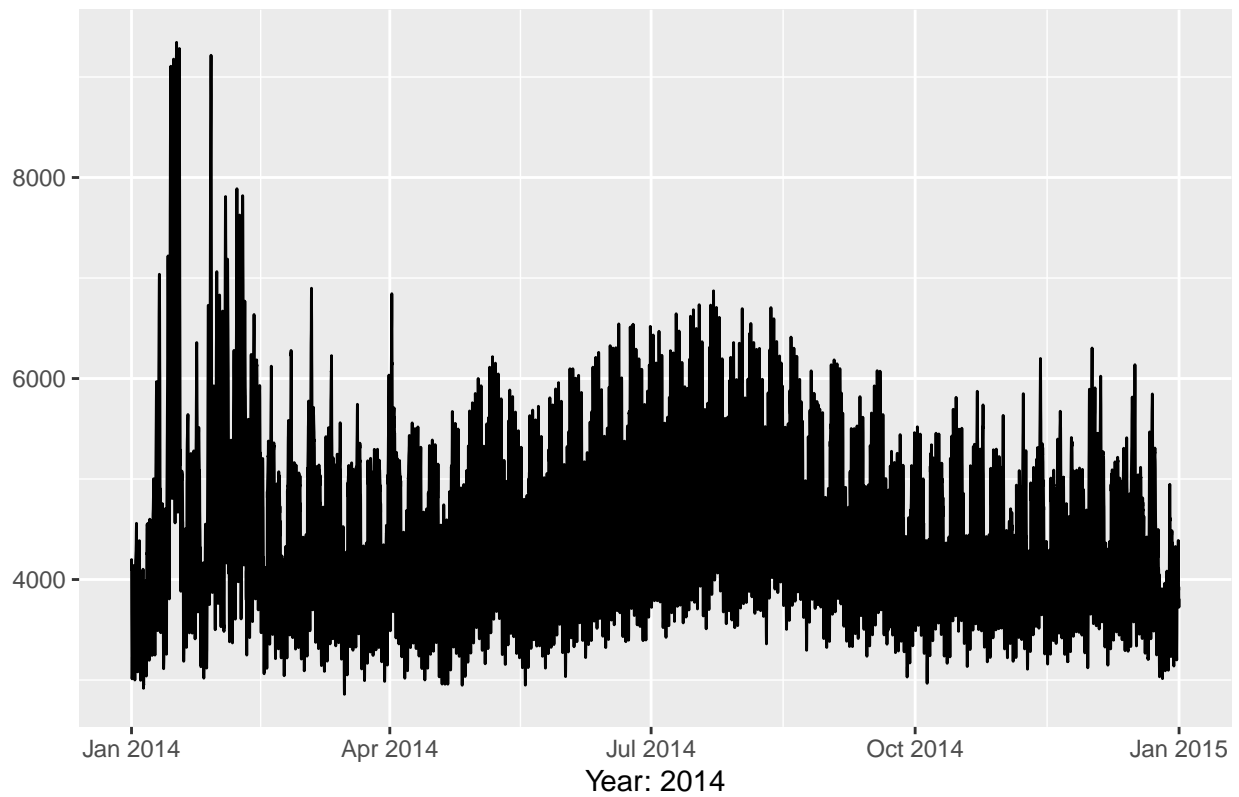
Year: 2014

```r
vic_elec %>%
  filter(year(Time) == 2014) %>%
  autoplot(Temperature) +
    xlab("Year: 2014") + ylab(NULL) +
    ggtitle("Half-hourly temperatures: Melbourne, Australia")
```
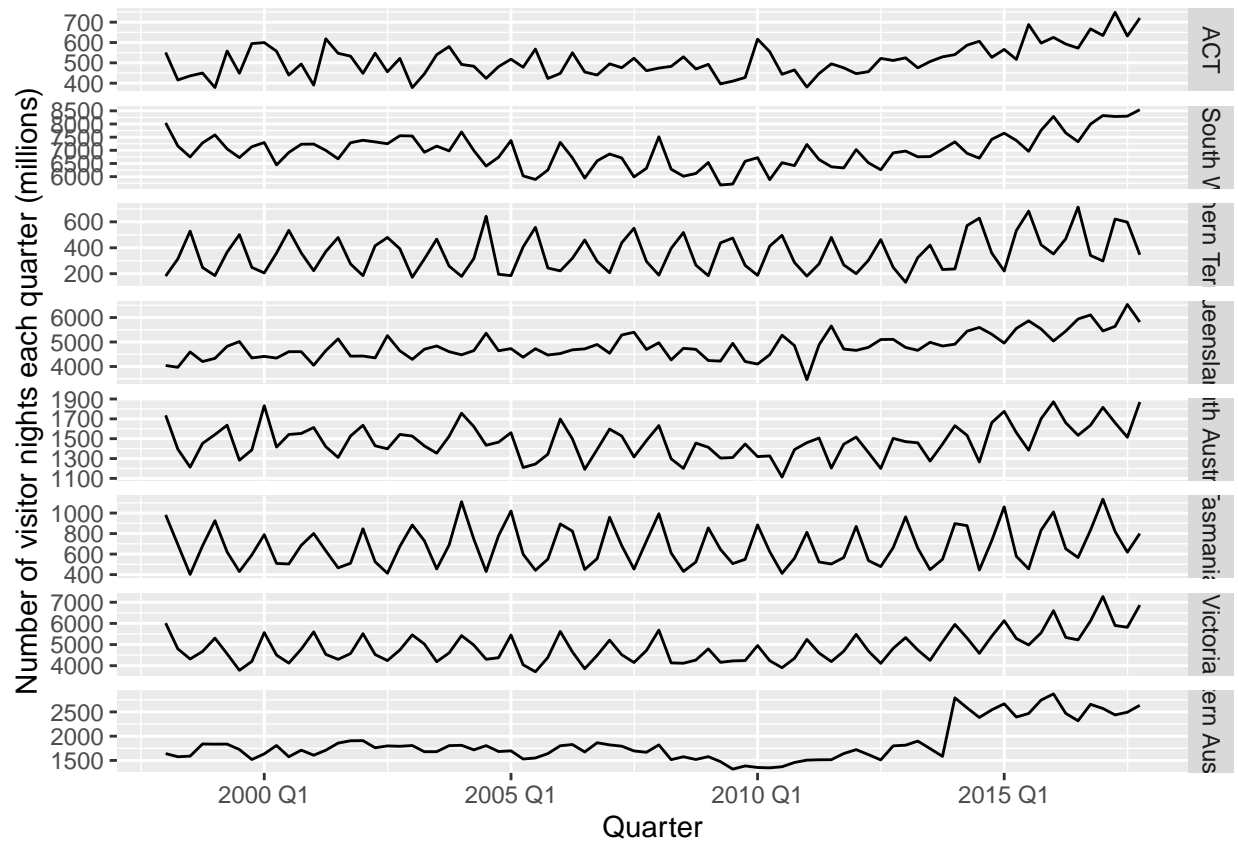
Half−hourly temperatures: Melbourne, Australia
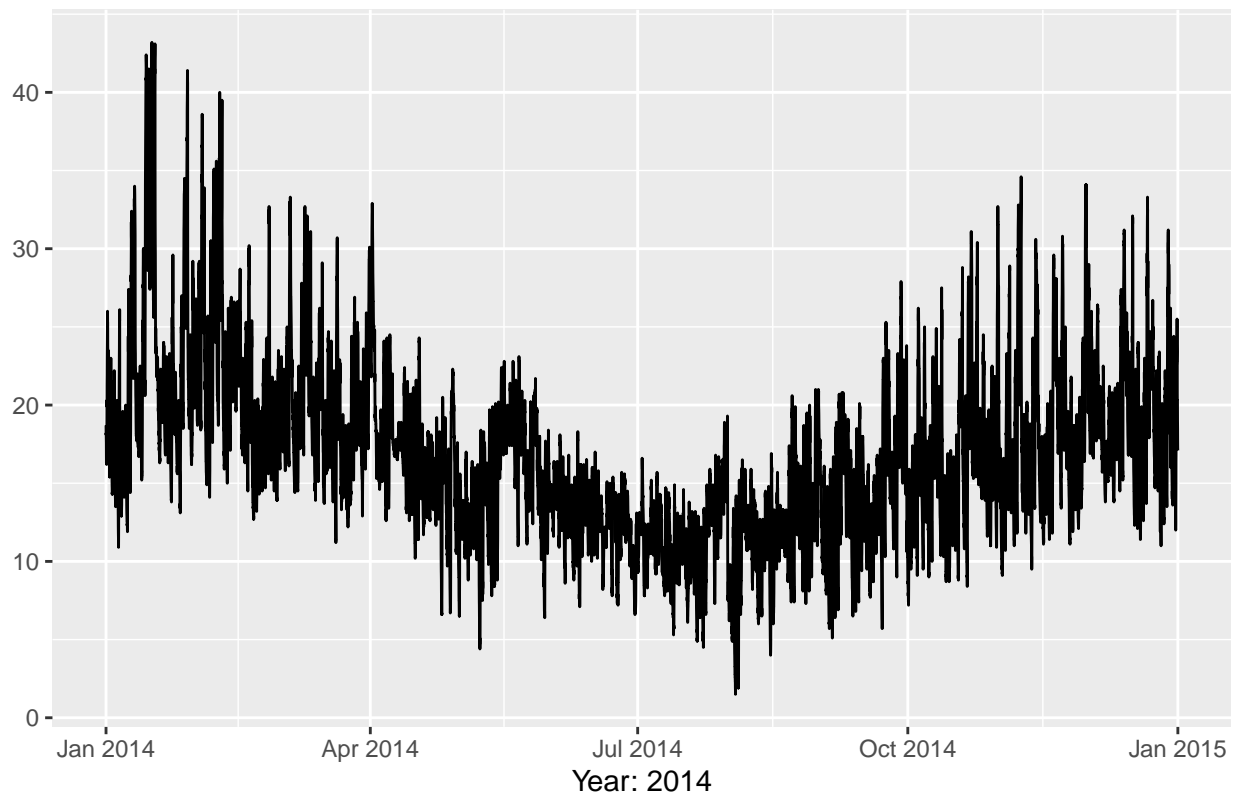

```r
vic_elec %>%
  filter(year(Time) == 2014) %>%
  ggplot(aes(x = Temperature, y = Demand)) +
    geom_point() +
    ylab("Demand (GW)") + xlab("Temperature (Celsius)")
```
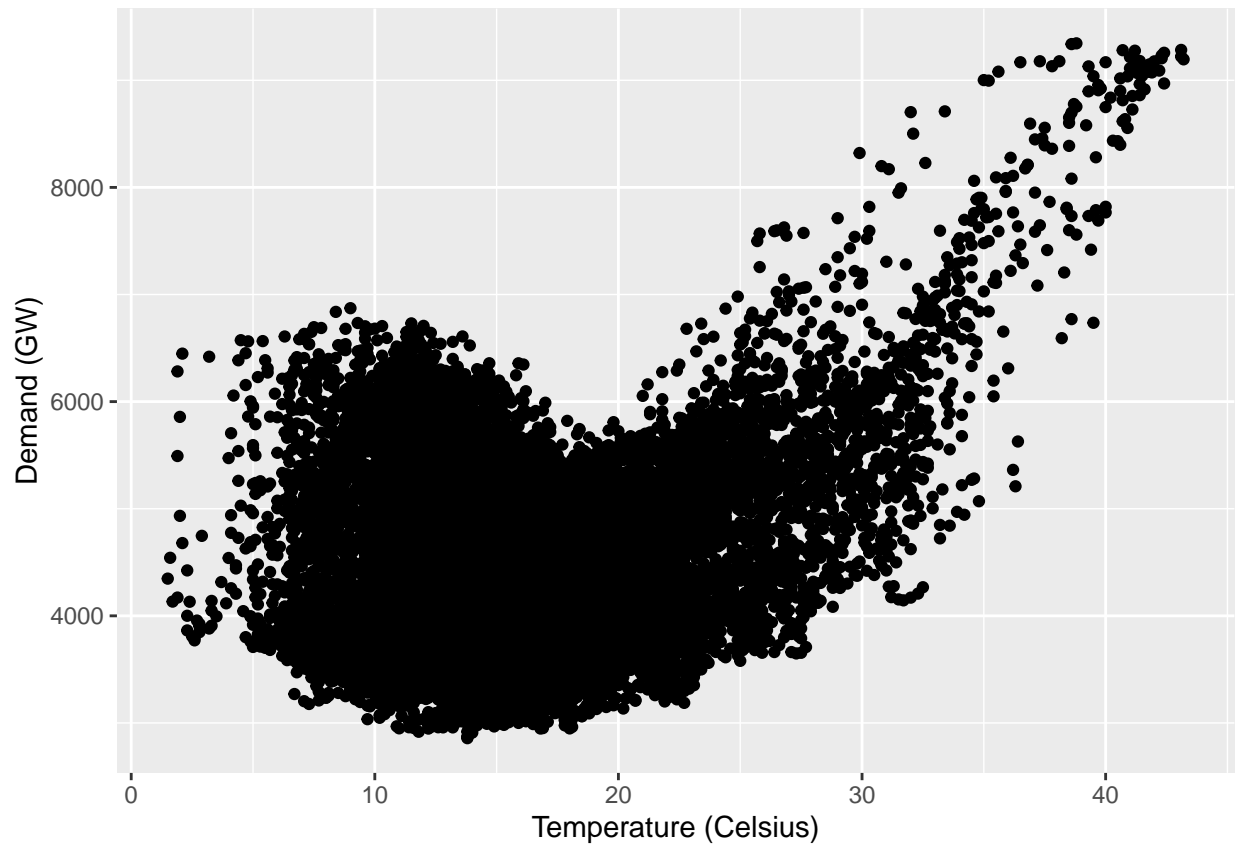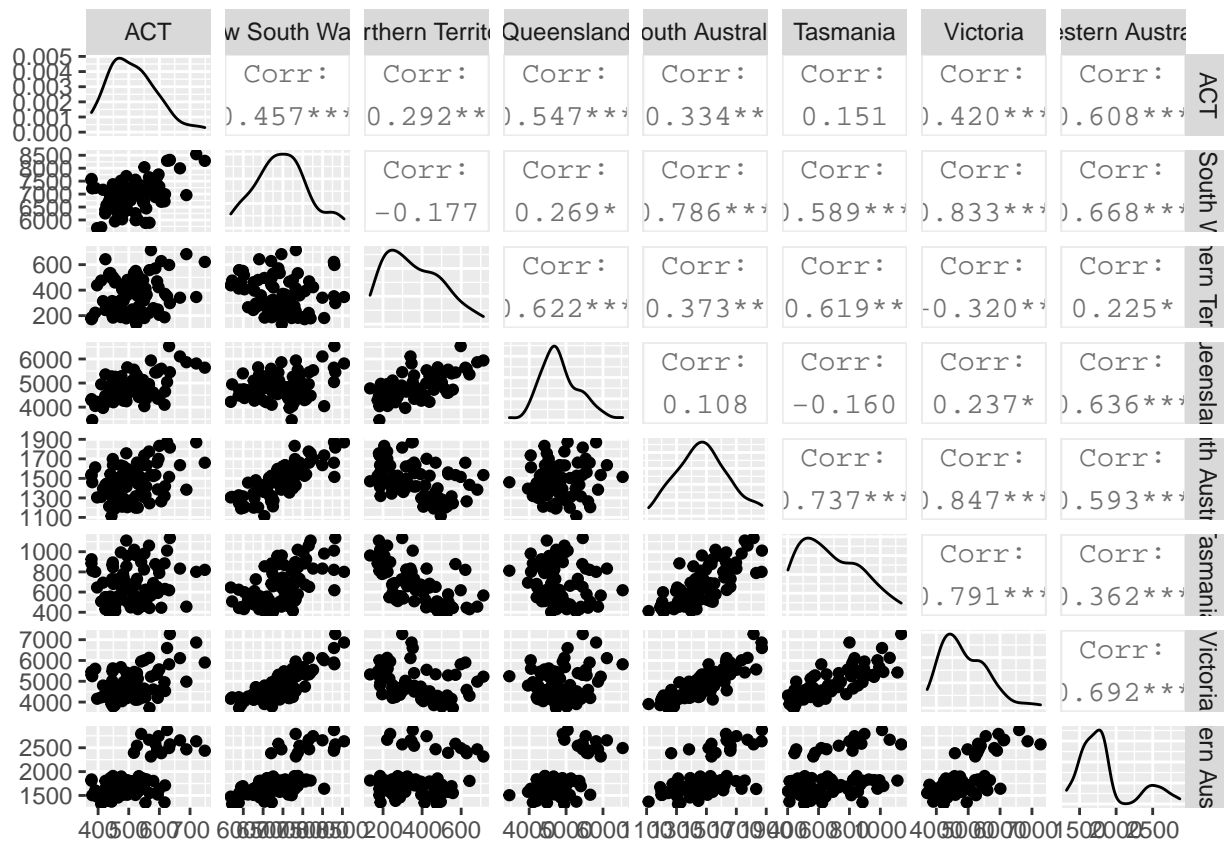
```
visitors <- tourism %>%
  group_by(State) %>%
  summarise(Trips = sum(Trips))
visitors %>%
  ggplot(aes(x = Quarter, y = Trips)) +
    geom_line() +
    facet_grid(vars(State), scales = "free_y") +
    ylab("Number of visitor nights each quarter (millions)")
```
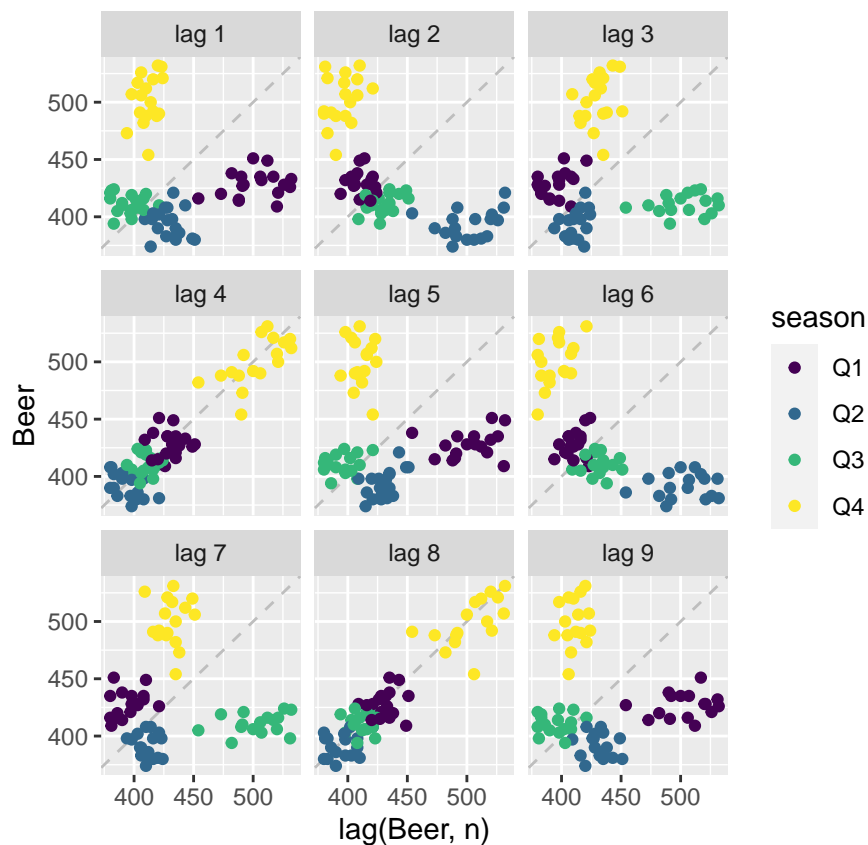
```
visitors %>%
  spread(State, Trips) %>%
  GGally::ggpairs(columns = 2:9)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
recent_production <- aus_production %>%
  filter(year(Quarter) >= 1992)
recent_production %>% gg_lag(Beer, geom="point")
```

```
recent_production
```

```
## # A tsibble: 74 x 7 [1Q]
##     Quarter  Beer Tobacco Bricks Cement Electricity   Gas
##       <qtr> <dbl>   <dbl>  <dbl>  <dbl>       <dbl> <dbl>
##  1 1992 Q1    443    5777    383   1289       38332   117
##  2 1992 Q2    410    5853    404   1501       39774   151
##  3 1992 Q3    420    6416    446   1539       42246   175
##  4 1992 Q4    532    5825    420   1568       38498   129
##  5 1993 Q1    433    5724    394   1450       39460   116
##  6 1993 Q2    421    6036    462   1668       41356   149
##  7 1993 Q3    410    6570    475   1648       42949   163
##  8 1993 Q4    512    5675    443   1863       40974   138
##  9 1994 Q1    449    5311    421   1468       40162   127
## 10 1994 Q2    381    5717    475   1755       41199   159
## # ... with 64 more rows
```
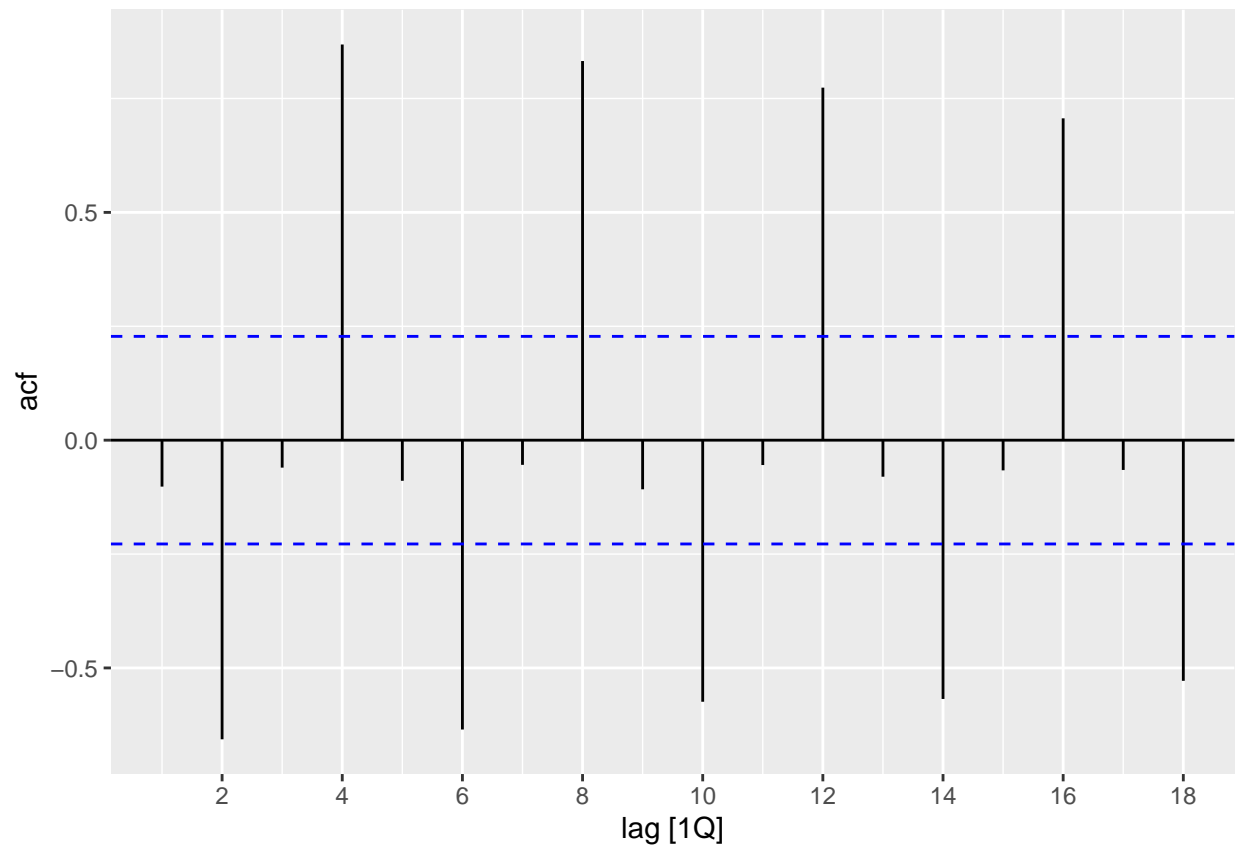
Autocorrelation Function

```
recent_production %>% ACF(Beer, lag_max = 9)
```
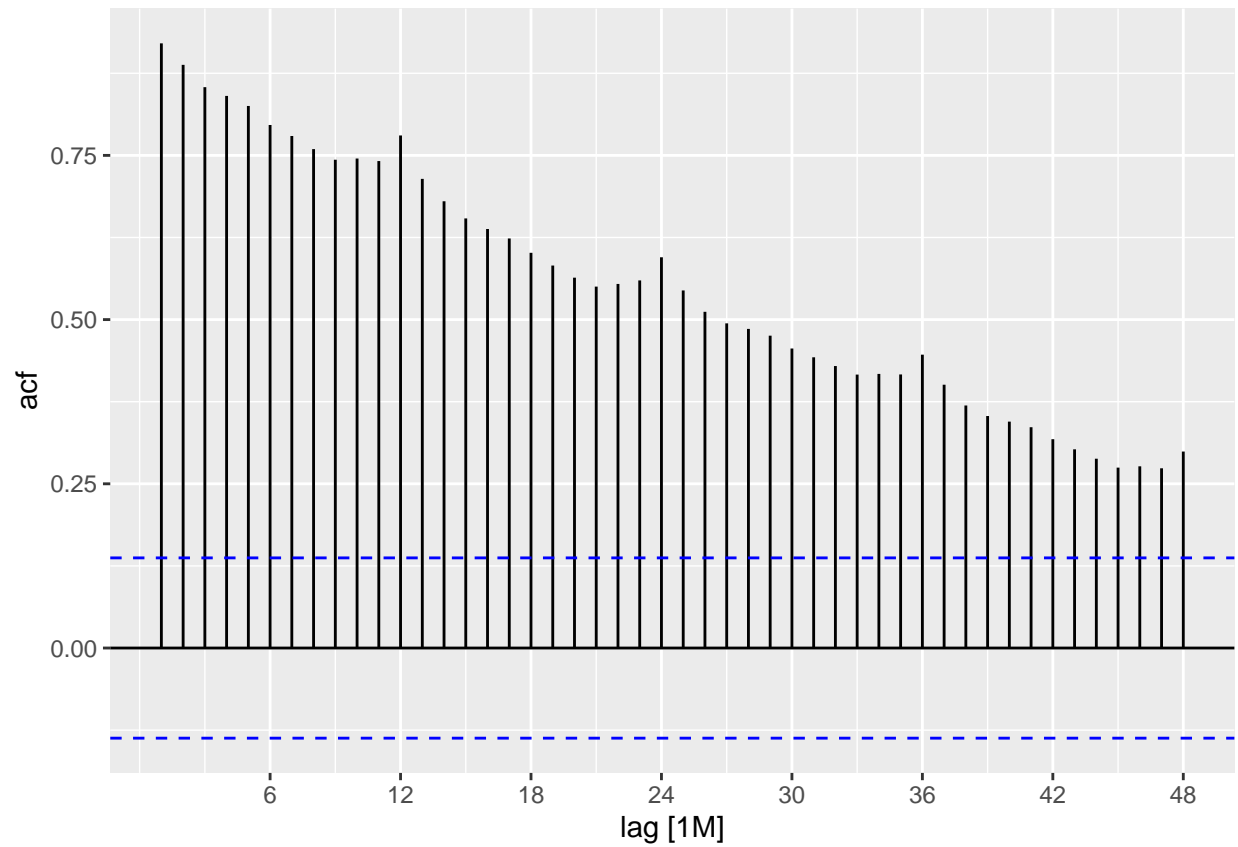
```
## # A tsibble: 9 x 2 [1Q]
##     lag    acf
```

```
##    <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
## 9    9Q  -0.108
```
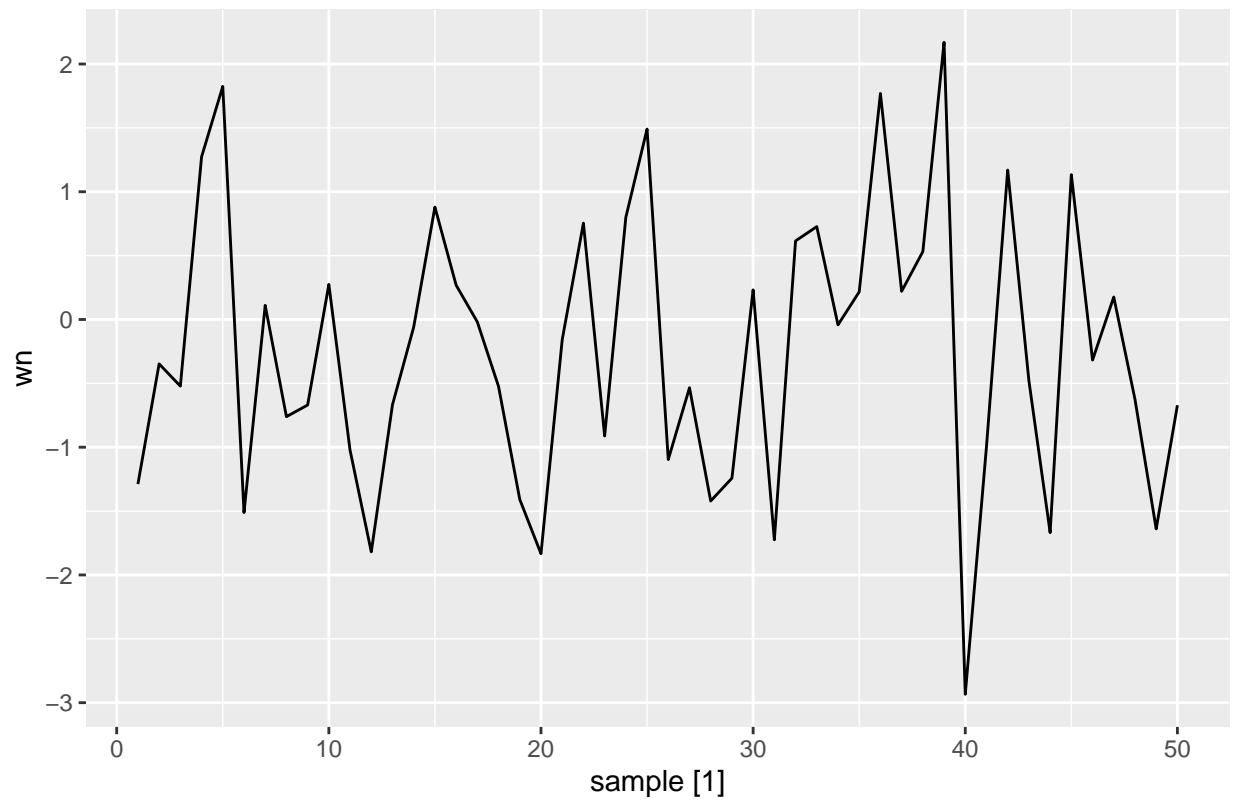
```
recent_production %>% ACF(Beer) %>% autoplot()
```



```
a10 %>% ACF(Cost, lag_max = 48) %>% autoplot()
```

White Noise

```r
set.seed(30)
y <- tsibble(sample = 1:50, wn = rnorm(50), index = sample)
y %>% autoplot(wn) + ggtitle("White noise")
```

White noise


```
y %>% ACF(wn) %>% autoplot()
```