Applied Economic Forecasting

2. Exploring & Visualizing Time series

Spring 2020

Section 1

Time series in R

`ts` objects and `ts` function

A time series is stored in a `ts` object in R:

- a list of numbers
- information about times those numbers were recorded.

Example

Year	Obs.
2012	123
2013	39
2014	78
2015	52
2016	110

```
y <- ts(c(123,39,78,52,110), start=2012)
```

`ts` objects and `ts` function

For observations that are more frequent than once per year, we need to add a `frequency` argument.

E.g., monthly data stored as a numerical vector `z`:

```
y <- ts(z, frequency=12, start=c(2003, 1))
```

`ts` objects and `ts` function

`ts(data, frequency, start)`

Type of data	frequency	start example
Annual	1	1995
Quarterly	4	c(1995,2)
Monthly	12	c(1995,9)
Daily	7 *or* 365.25	*or* c(1995,234)
Weekly	52.18	c(1995,23)
Hourly	24 *or* 168 *or* 8,766	1
Half-hourly	48 *or* 336 *or* 17,532	1

Let's Practice!!!

1. Set a seed as 10ˆ3
2. Generate a random normal variable (`x`) with 200 observations, mean=75, and sd= 5
3. Declare x as a *quarterly* ts object ending December 2018 (`x.ts`)
4. Now repeat step 3 with weekly, monthly and annual frequencies (*How about using a loop?*)

Australian GDP

```
ausgdp <- ts(x, frequency=4, start=c(1971,3))
```

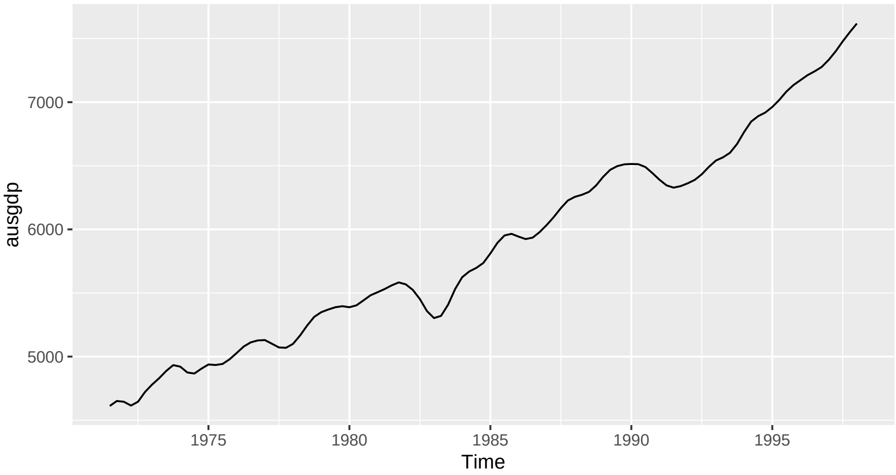
- Class: “ts”
- Print and plotting methods available.

```
head(ausgdp, 30)
```

```
##      Qtr1 Qtr2 Qtr3 Qtr4
## 1971           4612 4651
## 1972 4645 4615 4645 4722
## 1973 4780 4830 4887 4933
## 1974 4921 4875 4867 4905
## 1975 4938 4934 4942 4979
## 1976 5028 5079 5112 5127
## 1977 5130 5101 5072 5069
## 1978 5100 5166 5244 5312
```

Australian GDP

```r
autoplot(ausgdp)
```



Residential electricity sales

```
elecsales
```

```
## Time Series:
## Start = 1989
## End = 2008
## Frequency = 1
##  [1] 2354.34 2379.71 2318.52 2468.99 2386.09 2569.47 25
## [10] 3000.70 3108.10 3357.50 3075.70 3180.60 3221.60 3
## [19] 3637.89 3655.00
```

Class package

```
> library(fpp2)
```

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **ggplot2** package (for graphics functions)
- **fma** package (for lots of time series data)
- **expsmooth** package (for more time series data)

Section 2

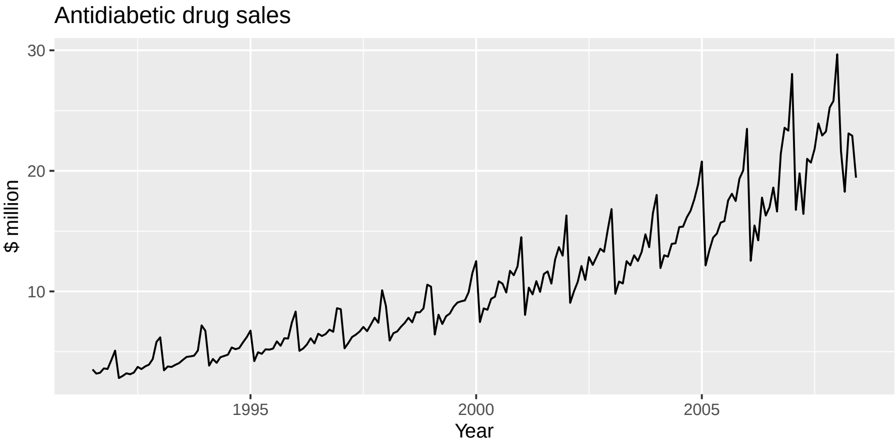
Time plots

Time plots

```r
autoplot(melsyd[, "Economy.Class"])
```

Time plots

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



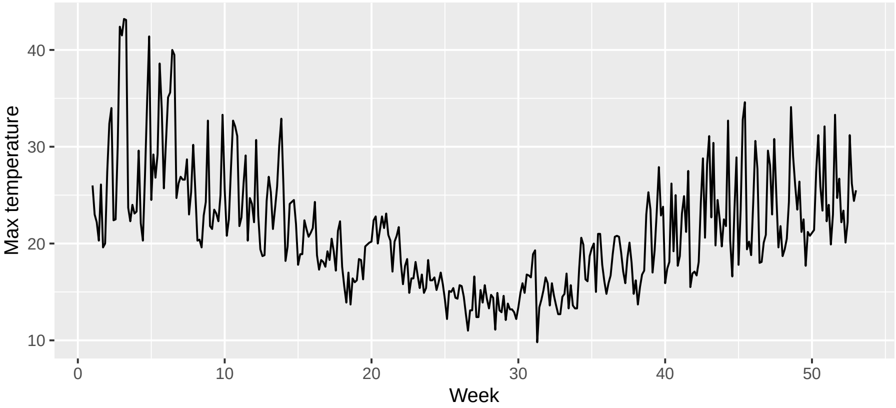
Let's Practice

1. Using the data from our earlier exercise: Plot `x.ts` at the monthly, quarterly and yearly frequencies.
2. Create plots of the following time series: `dole`, `bricksq`, `lynx`, `goog`

- Use `help()` to find out about the data in each series.
- For each of the plots, be sure to modify the axis labels and title.

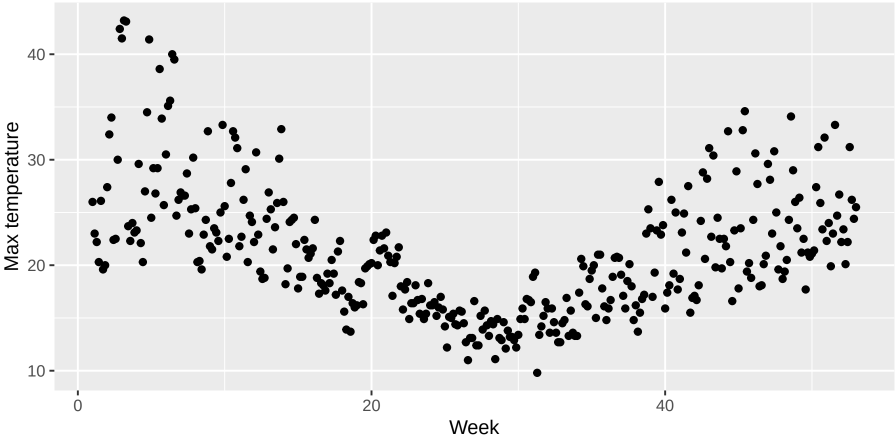
Are time plots best?

```r
autoplot(elecdaily[, "Temperature"]) + labs(x = "Week",
    y = "Max temperature")
```



Are time plots best?

```
qplot(time(elecdaily), elecdaily[, "Temperature"]) +
    labs(x = "Week", y = "Max temperature")
```

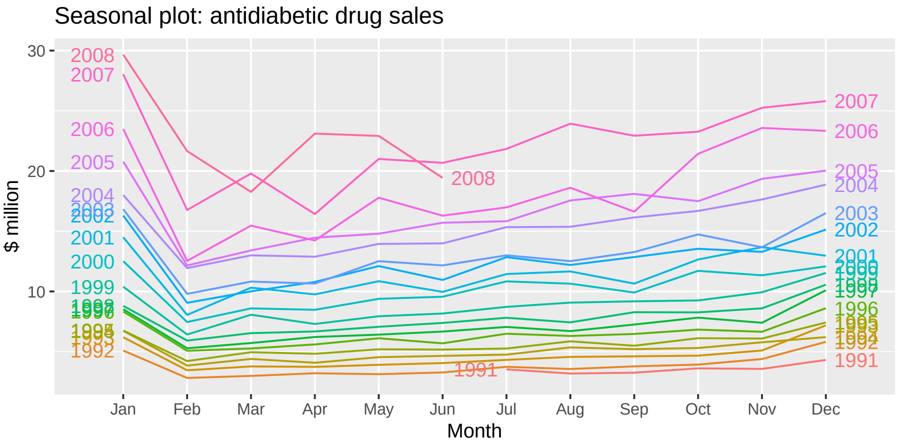


Section 3

Seasonal plots

Seasonal plots

```r
ggseasonplot(a10, year.labels = TRUE, year.labels.left = TRUE) +
    ylab("$ million") + ggtitle("Seasonal plot: antidiabetic drug sales")
```

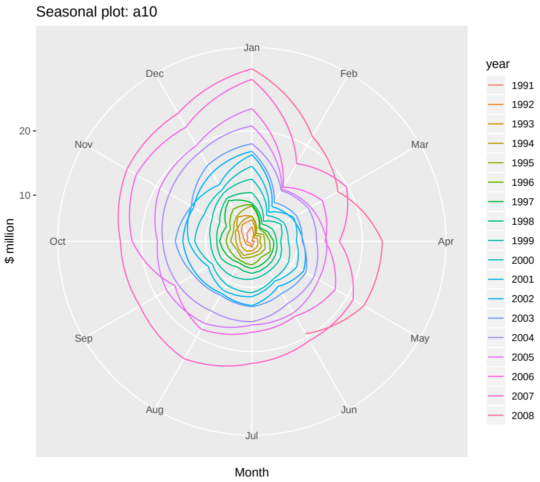


Seasonal plots

- Data plotted against the individual "seasons" in which the data were observed. (In this case a "season" is a month.)
- Something like a time plot except that the data from each season are overlapped.
- Enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily identified.
- In R: `ggseasonplot()`

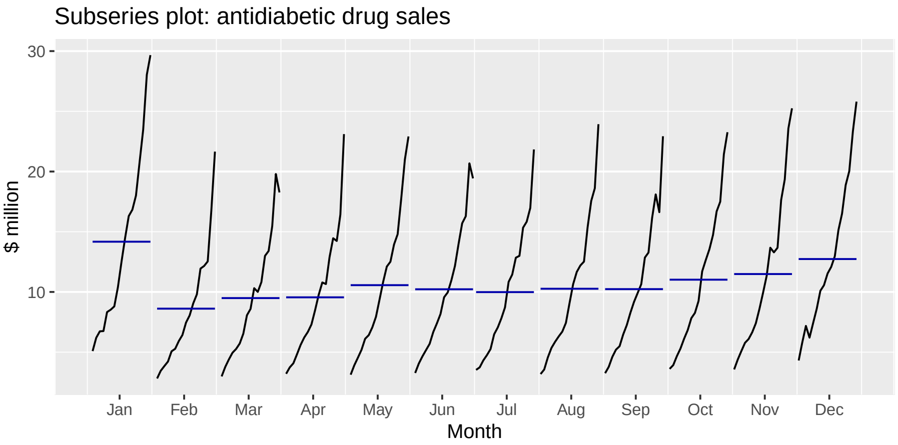
Seasonal polar plots

```r
ggseasonplot(a10, polar = TRUE) + ylab("$ million")
```



Seasonal subseries plots

```r
ggsubseriesplot(a10) + ylab("$ million") + ggtitle("Subseries plot:
```

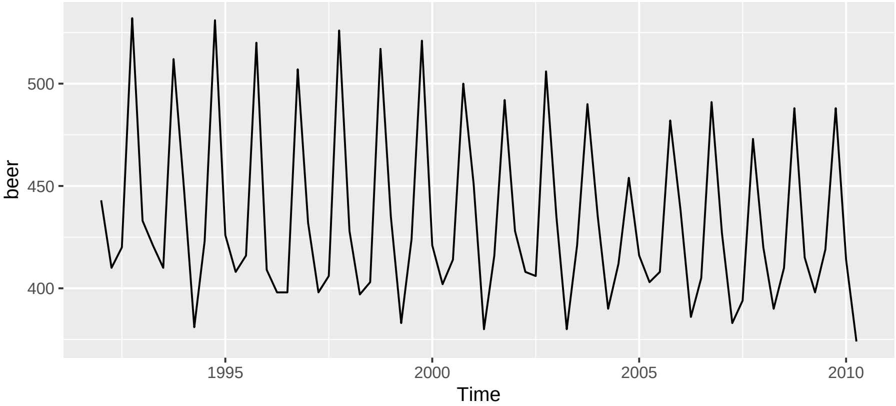


Seasonal subseries plots

- Data for each season collected together in time plot as separate time series.
- Enables the underlying seasonal pattern to be seen clearly, and changes in seasonality over time to be visualized.
- In R: `ggsubseriesplot()`

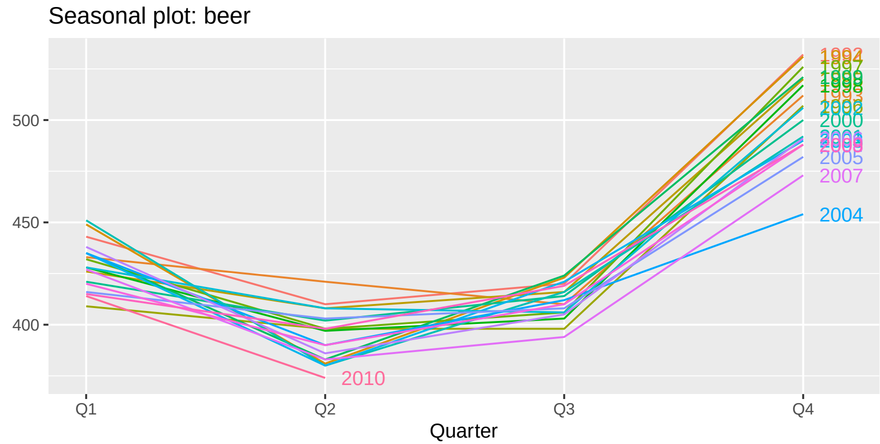
Quarterly Australian Beer Production

```r
beer <- window(ausbeer, start = 1992)
autoplot(beer)
```



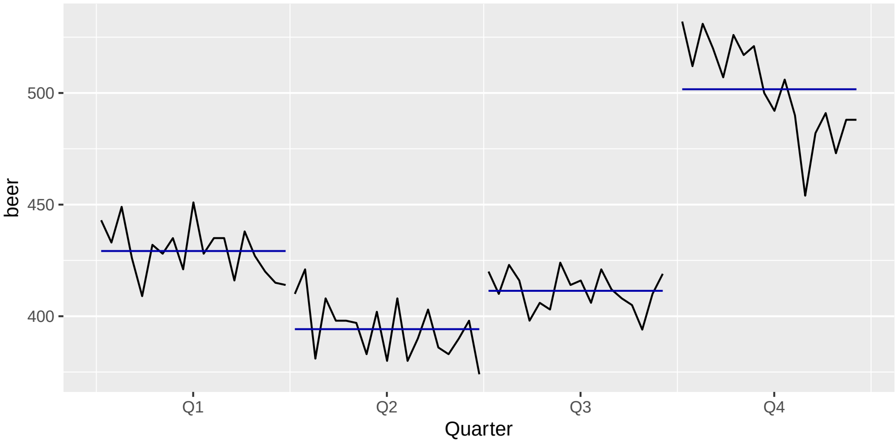
Quarterly Australian Beer Production

```r
ggseasonplot(beer, year.labels = TRUE)
```



Quarterly Australian Beer Production

```r
ggsubseriesplot(beer)
```



Let's Practice!!!

The `arrivals` data set comprises quarterly international arrivals (in thousands) to Australia from Japan, New Zealand, UK and the US.

- Use `autoplot()` and `ggseasonplot()` to compare the differences between the arrivals from these four countries.
- Can you identify any unusual observations?

Section 4

Seasonal or cyclic?

Time series patterns

Trend pattern exists when there is a long-term increase or decrease in the data.

Seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).

Cyclic pattern exists when data exhibit rises and falls that are *not of fixed period* (duration usually of at least 2 years).

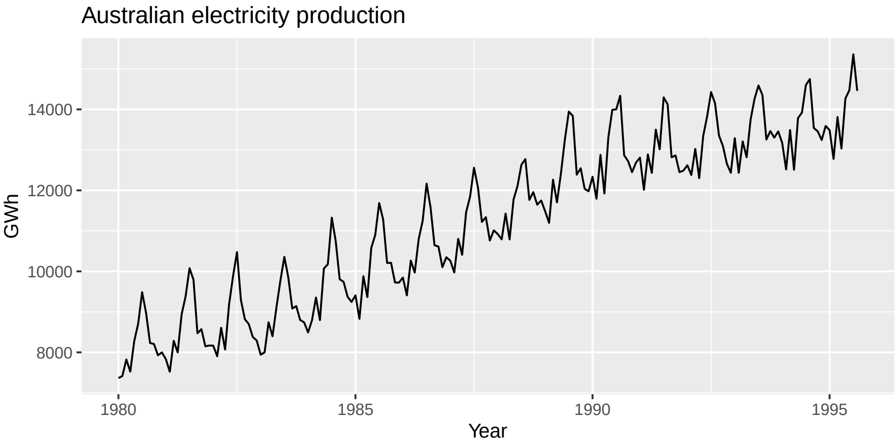
Time series components

Differences between seasonal and cyclic patterns:

- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

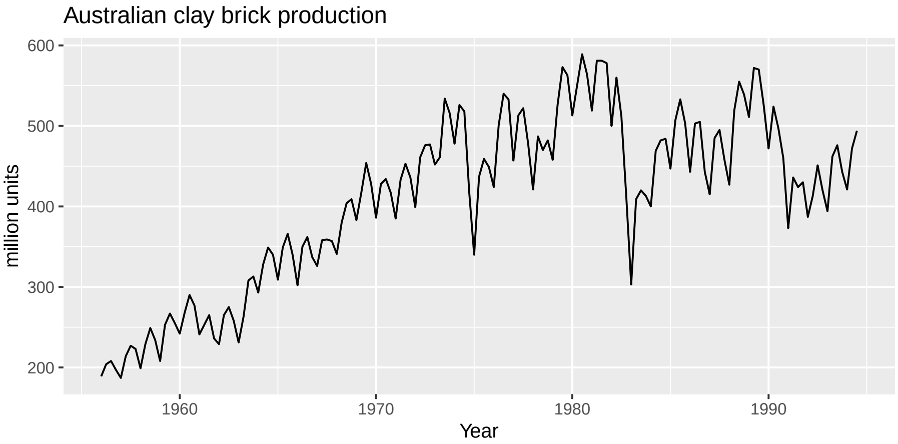
Time series patterns

```r
autoplot(window(elec, start = 1980)) + ggtitle("Australian electrici
    labs(x = "Year", y = "GWh")
```



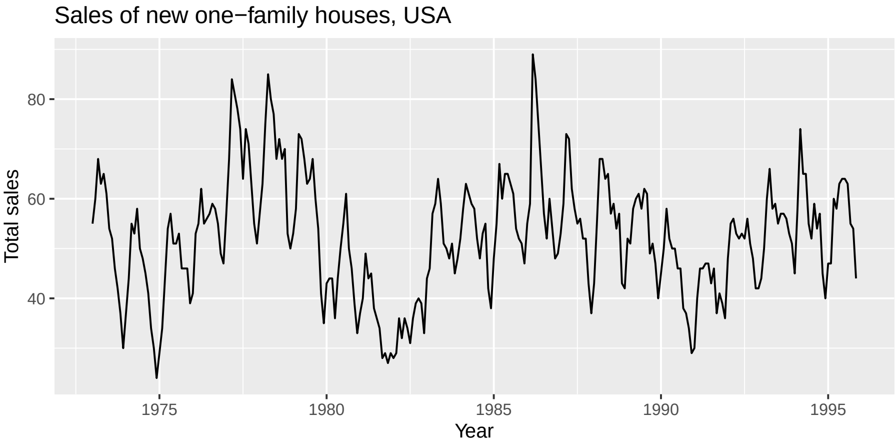
Time series patterns

```r
autoplot(bricksq) + ggtitle("Australian clay brick production") +
    xlab("Year") + ylab("million units")
```



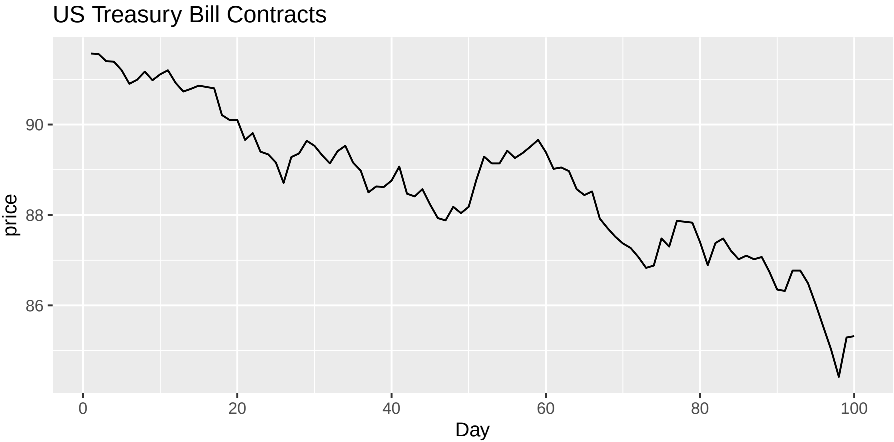
Time series patterns

```r
autoplot(hsales) + ggtitle("Sales of new one-family houses, USA") +
    labs(x = "Year", y = "Total sales")
```



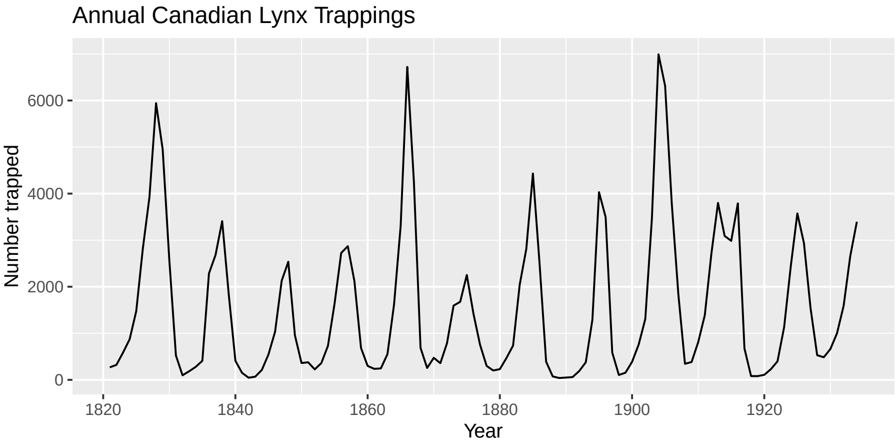
Time series patterns

```
autoplot(ustreas) + ggtitle("US Treasury Bill Contracts") +
    labs(x = "Day", y = "price")
```



Time series patterns

```r
autoplot(lynx) + ggtitle("Annual Canadian Lynx Trappings") +
    xlab("Year") + ylab("Number trapped")
```



Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

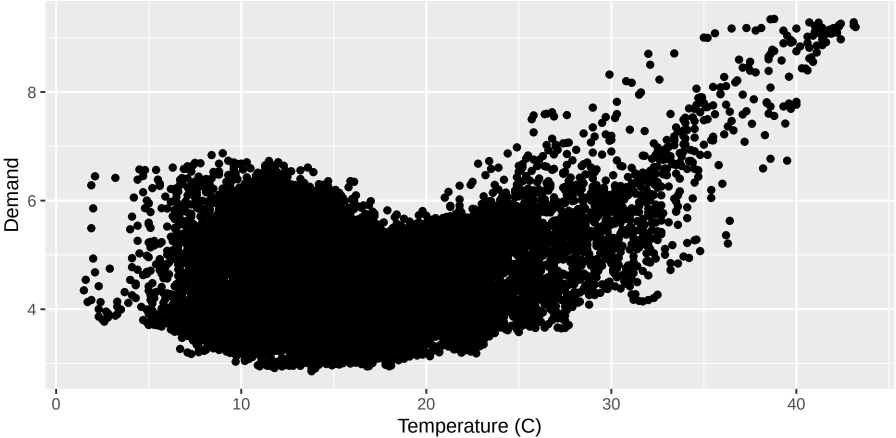
Section 5

Scatterplots

Scatterplots: Relationship *Between* Time Series

Scatterplots are commonly used tools for visualizing the relationship between variables. Take the text example of Electricty Demand and Temperature in Victoria, Autralia:

```
qplot(x = Temperature, y = Demand, data = as.data.frame(elecdemand)) +
    labs(y = "Demand", x = "Temperature (C)")
```



Scatterplots: Relationship *Between* Time Series

We notice that higher temperatures are associated with higher demand for electricity.

What if we would like to get a measure of the strength of this relationship?

- For this, we will use the **correlation coefficient**:

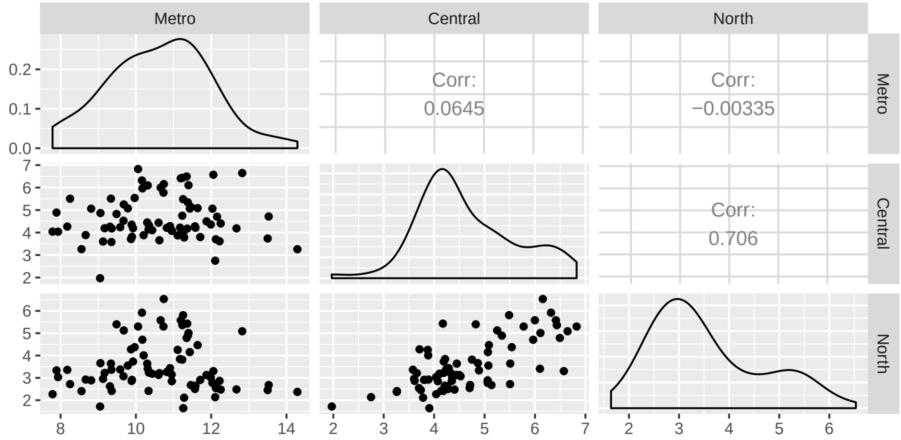
$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y - \bar{y})^2}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sigma_x^2}\sqrt{\sigma_y^2}}, \quad -1 \leq \rho_{xy} \leq 1$$

- Negative values indicate a negative *linear* relationship between `x` and `y`.
- Positive values indicate a positive *linear* relationship between `x` and `y`.

Scatterplot Matrices: Relationship *Between* Time Series

We can visualize the correlation between multiple series using a scatterplot matrix. In R, we will use the ggpairs() command from the GGally

```
# install.packages('GGally') #unmute when running
# for the first time
GGally::ggpairs(as.data.frame(visnights[, 6:8]), columnLabels = c("Metro",
    "Central", "North"))
```

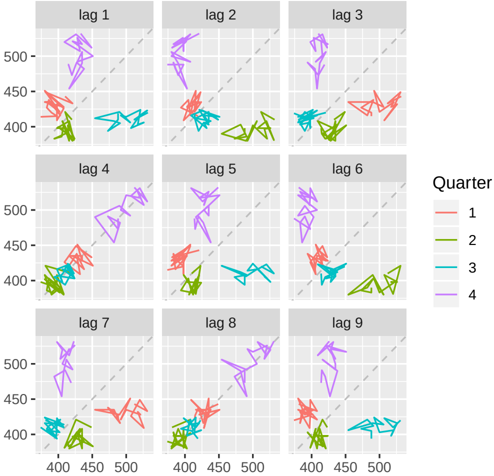


Section 6

Lag plots and autocorrelation

Example: Beer production

```r
beer <- window(ausbeer, start = 1992)
gglagplot(beer)
```



Lagged scatterplots

- Each graph shows y_t plotted against y_{t-k} for different values of k.
- The autocorrelations are the correlations associated with these scatterplots.

Autocorrelation

Covariance and **correlation**: measure extent of **linear relationship** between two variables (y and X).

Autocovariance and **autocorrelation**: measure linear relationship between **lagged values** of a time series y.

We measure the relationship between:

- y_t and y_{t-1}
- y_t and y_{t-2}
- y_t and y_{t-3}
- etc.

Autocorrelation

We denote the sample autocovariance at lag k by γ_k and the sample autocorrelation at lag k by ρ_k. Then define

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

$$\text{and} \qquad \rho_k = \frac{\gamma_k}{\gamma_0}$$

- It is easy to see that γ_0 is the variance of y. Let `k = 0` then:

$$\gamma_0 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})(y_t - \bar{y}) = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2$$

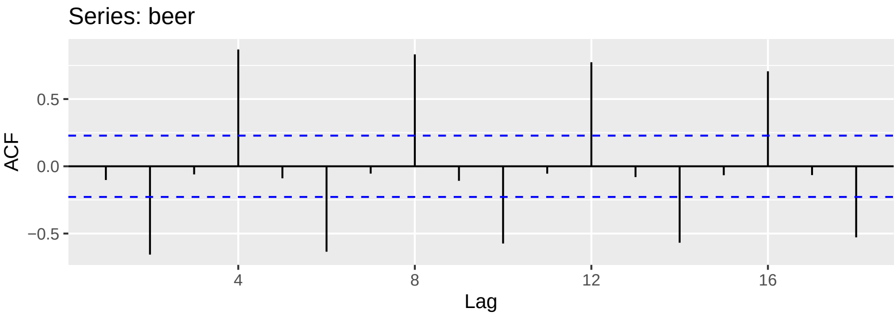
- ρ_1 indicates how successive values of y relate to each other
- ρ_2 indicates how y values two periods apart relate to each other
- ρ_k is *almost* the same as the sample correlation between y_t and y_{t-k}.

Autocorrelation

Results for first 9 lags for beer data:

ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9
-0.102	-0.657	-0.060	0.869	-0.089	-0.635	-0.054	0.832	-0.108

```
ggAcf(beer)
```

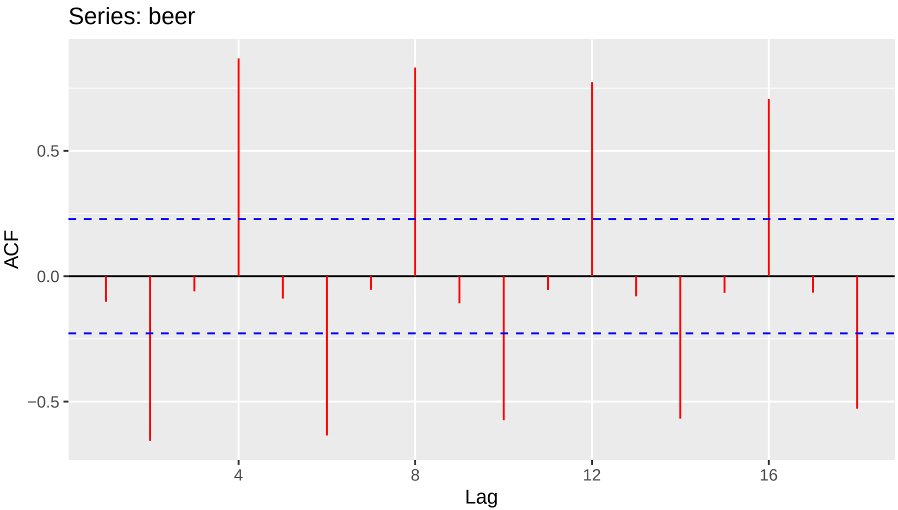


Autocorrelation

- ρ_4 higher than for the other lags. This is due to **the seasonal pattern in the data**: the peaks tend to be **4 quarters** apart and the troughs tend to be **2 quarters** apart.
- ρ_2 is more negative than for the other lags because troughs tend to be 2 quarters behind peaks.
- Together, the autocorrelations at lags 1, 2, . . . , make up the *autocorrelation function* or ACF.
- The plot is known as a **correlogram**

ACF

```r
ggAcf(beer, col = "red")
```

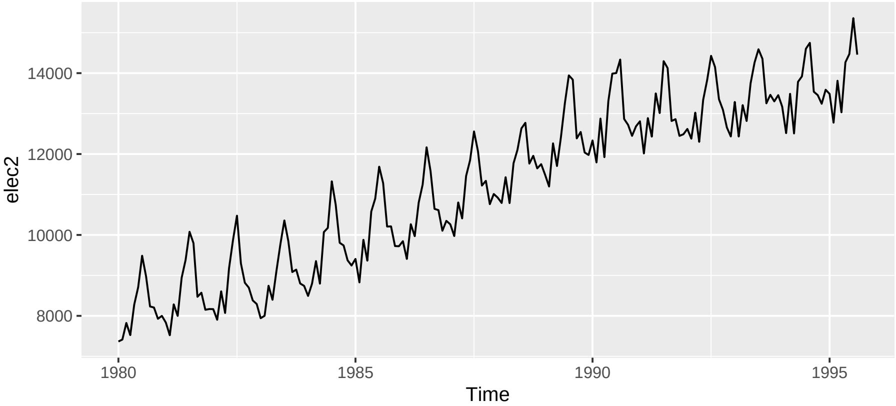


Trend and seasonality in ACF plots

- When data have a trend, the autocorrelations for small lags tend to be large and positive.
- When data are seasonal, the autocorrelations will be larger at the seasonal lags (i.e., at multiples of the seasonal frequency)
- When data are trended and seasonal, you see a combination of these effects.

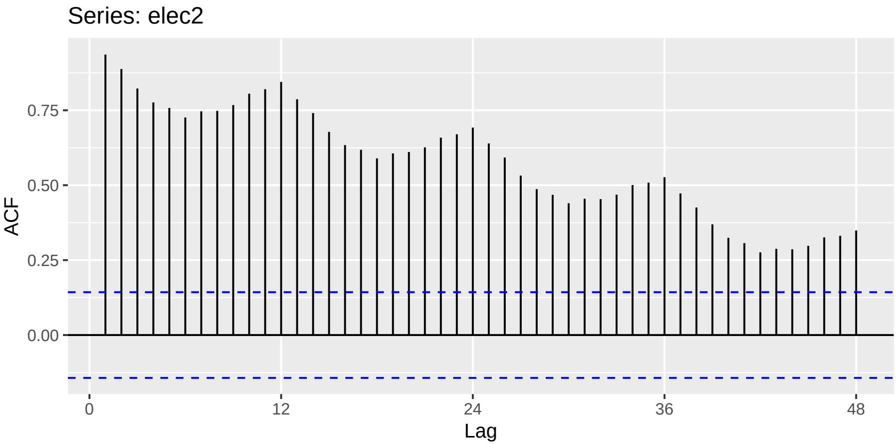
Aus monthly electricity production

```r
elec2 <- window(elec, start = 1980)
autoplot(elec2)
```



Aus monthly electricity production

```
ggAcf(elec2, lag.max = 48)
```



Aus monthly electricity production

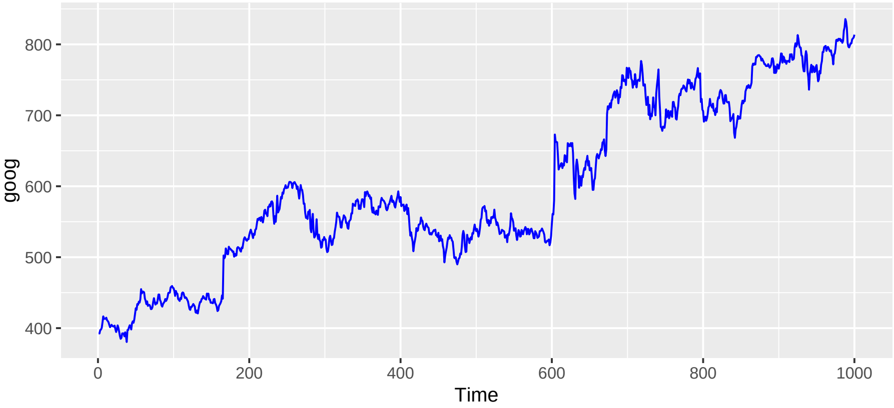
Time plot shows clear trend and seasonality.

The same features are reflected in the ACF.

- The slowly decaying ACF indicates trend.
- The ACF peaks at lags 12, 24, 36, . . . , indicate seasonality of length 12.

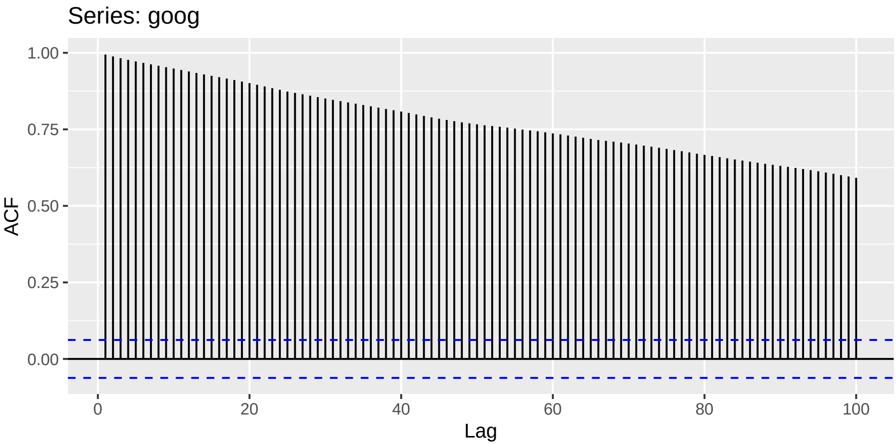
Google stock price

```r
autoplot(goog, col = "blue")
```



Google stock price

```r
ggAcf(goog, lag.max = 100)
```



Let's Practice!!!

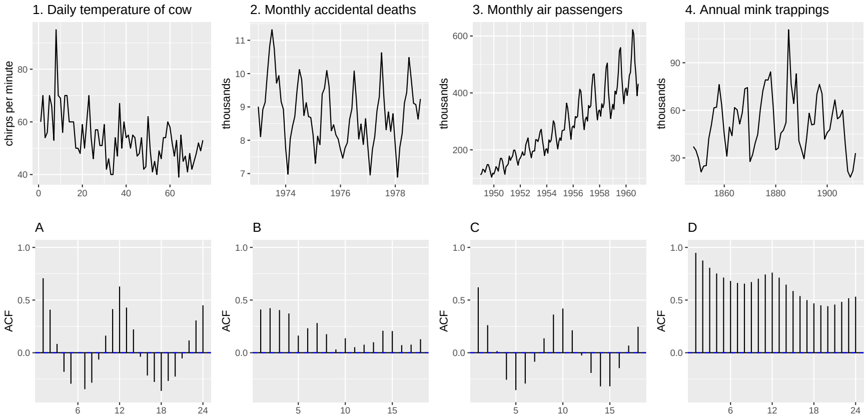
We have introduced the following graphics functions:

- `gglagplot`
- `ggAcf`

Explore the following time series using these functions. Can you spot any seasonality, cyclicity and trend? What do you learn about the series?

- `hsales`
- `usdeaths`
- `bricksq`
- `sunspotarea`
- `gasoline`

Which is which?

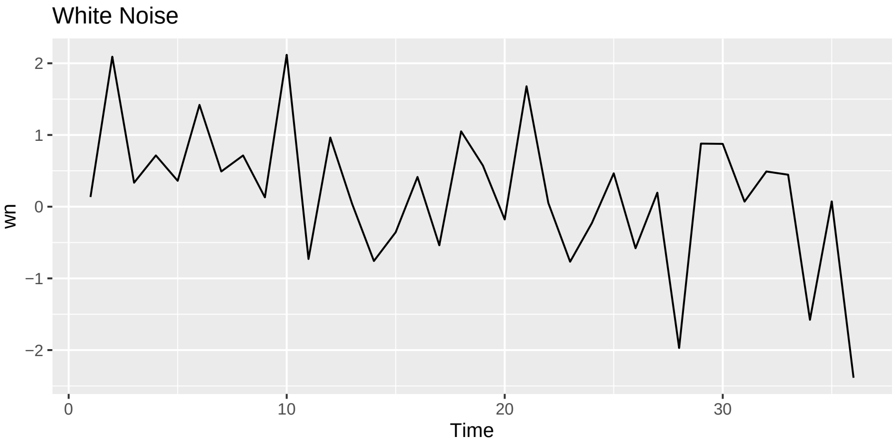


Section 7

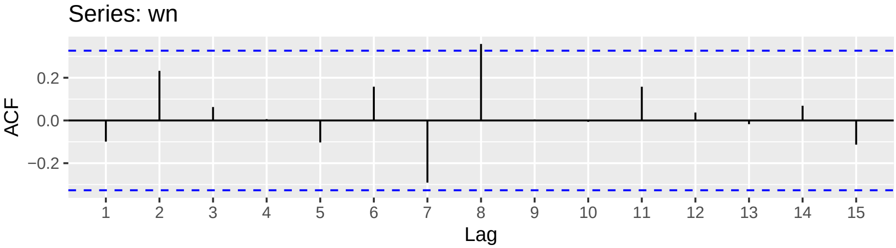
White noise

Example: White noise

```
set.seed(14567)
wn <- ts(rnorm(36))
autoplot(wn) + ggtitle("White Noise")
```



Example: White noise



ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7
-0.099	0.232	0.063	0.006	-0.103	0.158	-0.291

We expect each autocorrelation to be close to zero.

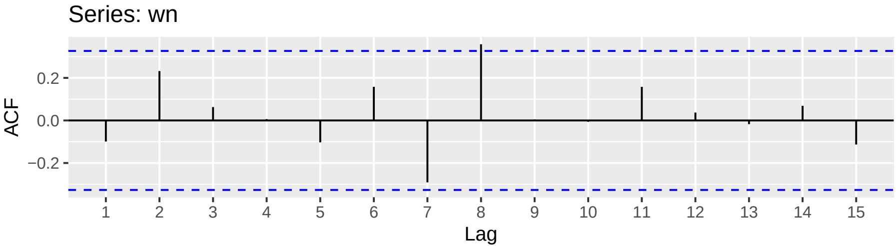
Sampling distribution of ρ_k for white noise data is asymptotically $\mathrm{N}(0, \frac{1}{T})$.

- 95% of all ρ_k for white noise must lie within $\pm\frac{1.96}{\sqrt{T}}$.
- If this is not the case, the series is probably not WN.
- Common to plot lines at $\pm 1.96/\sqrt{T}$ when plotting ACF. These are the ***critical values***.

Autocorrelation

Example:

```
ggAcf(wn)
```



$T = 36$ and so critical values at $\pm\frac{1.96}{\sqrt{36}} = \pm 0.327$.

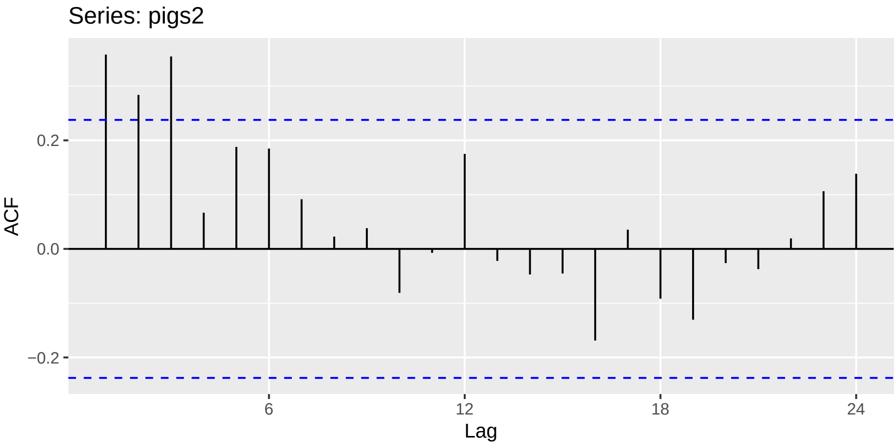
All autocorrelation coefficients lie within these limits. Data cannot be distinguished from white noise.

Example: Pigs slaughtered

```
pigs2 <- window(pigs, start = 1990)
autoplot(pigs2) + xlab("Year") + ylab("thousands") +
    ggtitle("Number of pigs slaughtered in Victoria")
```

Example: Pigs slaughtered

```r
ggAcf(pigs2)
```



Example: Pigs slaughtered

Monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. (Source: Australian Bureau of Statistics.)

- Difficult to detect pattern in time plot.
- ACF shows some significant autocorrelation at lags 1, 2, and 3.
- ρ_{12} relatively large although not significant. This may indicate some slight seasonality.

These show the series is **not a white noise series**.

Let's Practice!!!

1. Compute the daily changes in the Google stock price (`ddgoog`).
2. Does `ddgoog` look like white noise?
    - Present a time plot of `ddgoog`
    - Present the ACF plot

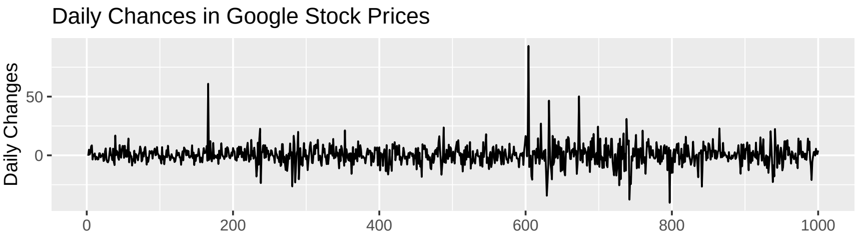
Let’s Practice (Solution)

Manually:

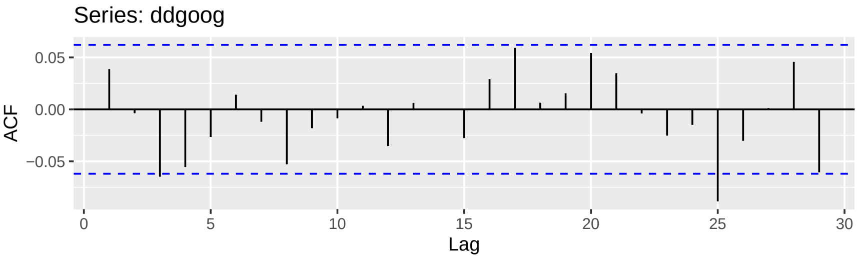
```r
# initialize the ddgoog variable to be a object
# with NAs
ddgoog <- NA

# We can use the loop to do the changes
# calculations Remember with differencing we lose
# the first obs so we start the counter at 2
for (i in 2:length(goog)) {
    ddgoog[i] <- goog[i] - goog[i - 1]  # Store results in ddgoog at position [i]
}
# Declare as a ts and drop first obs since it's NA
ddgoog <- ts(ddgoog[-1], start = c(2))
round(head(ddgoog), 3)
```

```
## Time Series:
## Start = 2
## End = 7
## Frequency = 1
## [1] -0.318  4.794  0.705  2.479  7.606  8.495
```

```r
autoplot(ddgoog) + ggtitle("Daily Chances in Google Stock Prices") +
    labs(y = "Daily Changes", x = "")
```




```r
ggAcf(ddgoog)
```



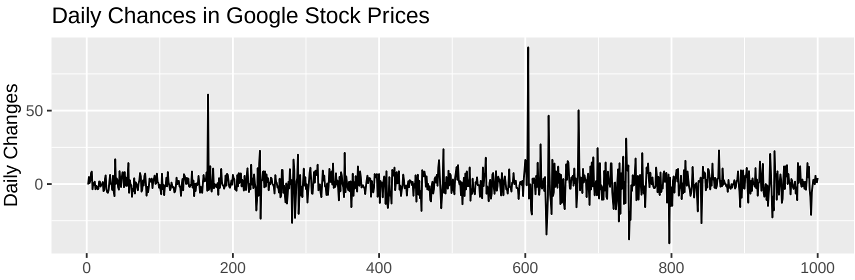
Let's Practice (Solution)

Using the fpp2 package:

```
dgoog <- diff(goog)
round(head(dgoog), 3)
```

```
## Time Series:
## Start = 2
## End = 7
## Frequency = 1
## [1] -0.318  4.794  0.705  2.479  7.606  8.495
```

```
autoplot(dgoog) + ggtitle("Daily Chances in Google Stock Prices") +
    labs(y = "Daily Changes", x = "")
```



Let's Practice (Solution)

So, what is your conclusion?