Economic Forecasting

Time series data

Sebastian Fossati

University of Alberta | E493 | 2023

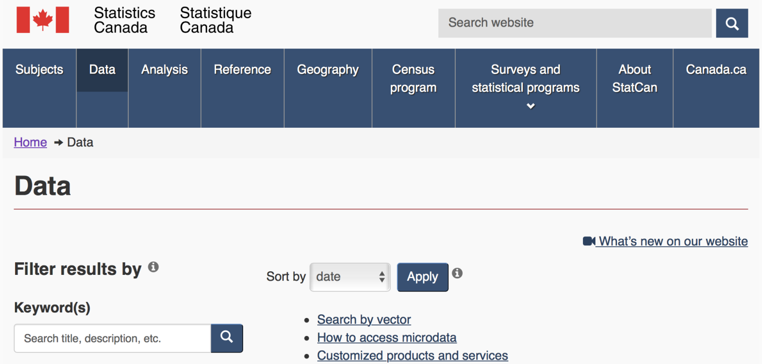
Outline

Time series data

Examples

- daily IBM stock prices
- monthly rainfall
- quarterly Australian beer production
- annual Google profits

Where can we find time series data?

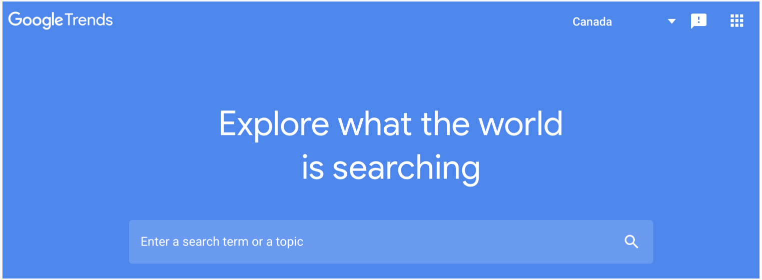


Statistcs Canada

Where can we find time series data?

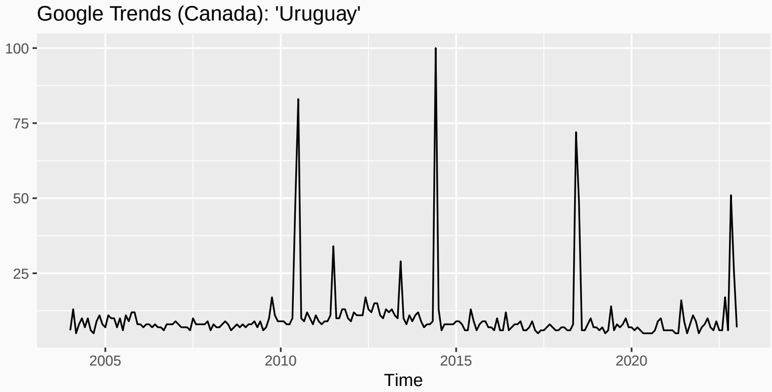
FRED

Where can we find time series data?



Google Trends

Your turn


Google Trends (Canada): 'Uruguay'
100
75
50
25
2005
2010
2015
2020
Time

`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

Year	Observation
2012	123
2013	39
2014	78
2015	52

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

`ts` objects and `ts` function

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

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

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

`ts` objects and `ts` function

Time series object: `ts(data, frequency, start)`

Type of data	frequency	start example
Annual		
Quarterly		
Monthly		
Weekly		
...		

`ts` objects and `ts` function

Time series object: `ts(data, frequency, start)`

Type of data	frequency	start example
Annual	1	1995
Quarterly		
Monthly		
Weekly		
...		

`ts` objects and `ts` function

Time series object: `ts(data, frequency, start)`

Type of data	frequency	start example
Annual	1	1995
Quarterly	4	c(1995,2)
Monthly		
Weekly		
...		

`ts` objects and `ts` function

Time series object: `ts(data, frequency, start)`

Type of data	frequency	start example
Annual	1	1995
Quarterly	4	c(1995,2)
Monthly	12	c(1995,9)
Weekly		
...		

`ts` objects and `ts` function

Time series object: `ts(data, frequency, start)`

Type of data	frequency	start example
Annual	1	1995
Quarterly	4	c(1995,2)
Monthly	12	c(1995,9)
Weekly	52.18	c(1995,23)
...		

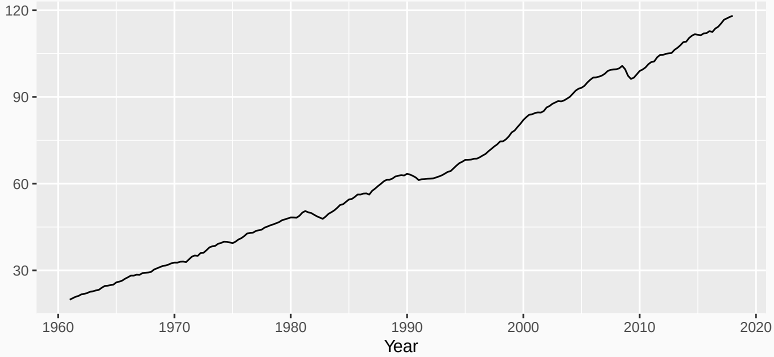
Canadian GDP

```r
# read canada quarterly real gdp data
data <- read.csv("data/NAEXKP01CAQ661S.csv", header = TRUE)
cangdp <- ts(data$NAEXKP01CAQ661S, start = 1961, freq = 4)
# display some observations
window(cangdp, start = c(2001,1), end = c(2005,4))
```

```
##       Qtr1  Qtr2  Qtr3  Qtr4
## 2001 84.43 84.65 84.59 85.11
## 2002 86.37 86.87 87.62 88.11
## 2003 88.60 88.46 88.80 89.41
## 2004 90.05 91.12 92.20 92.86
## 2005 93.18 93.84 95.00 95.94
```

Canadian Real GDP

```r
# time series plot
autoplot(cangdp) + xlab("Year") + ylab(" ")
```

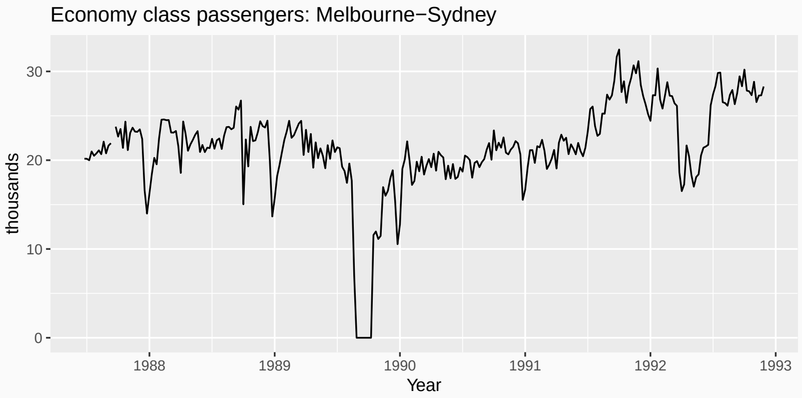


Your turn

Update the Canadian real GDP time series to include the latest observations available (you should be able to get data up to 2022Q3).

Outline

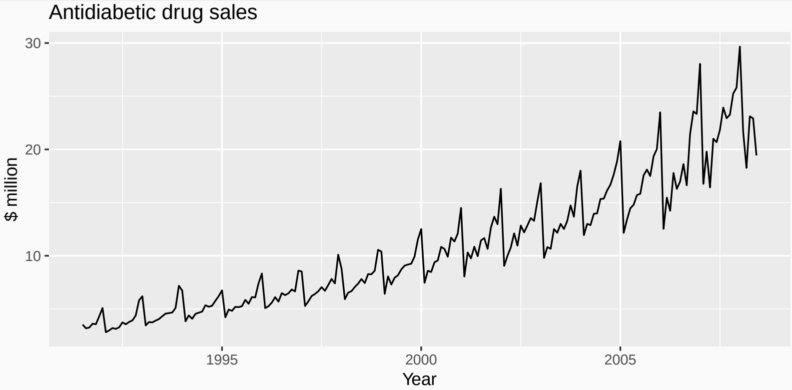
Time plots



Economy class passengers: Melbourne−Sydney
thousands
30
20
10
0
1988
1989
1990
1991
1992
1993
Year

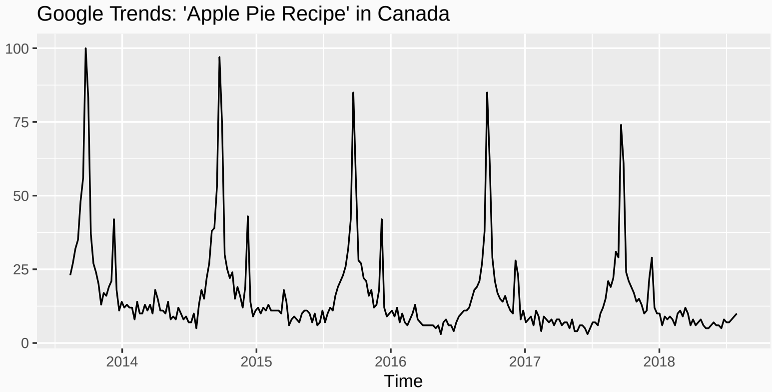
Time plots

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

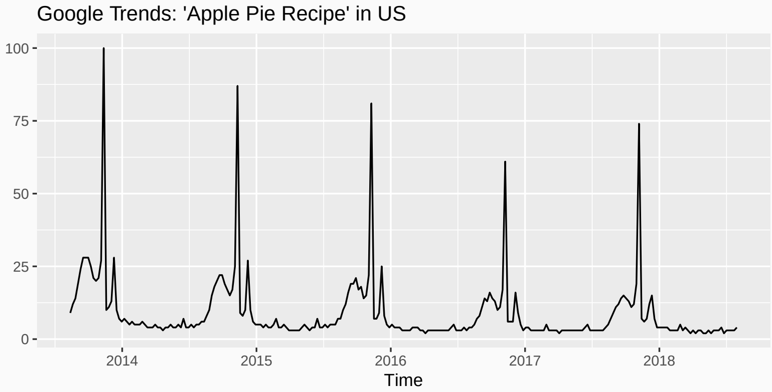


Outline

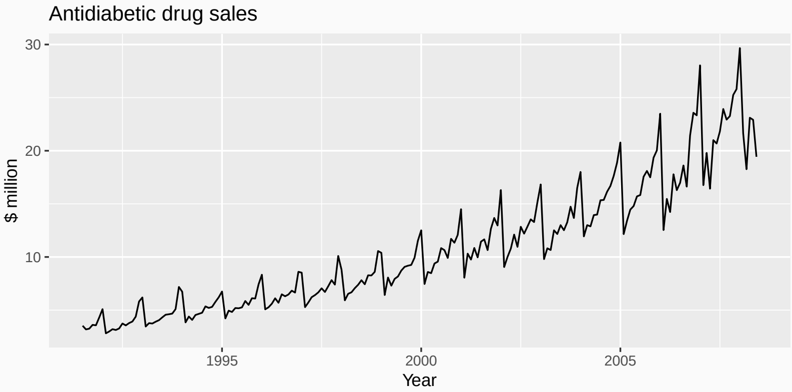
Seasonal data



Seasonal data

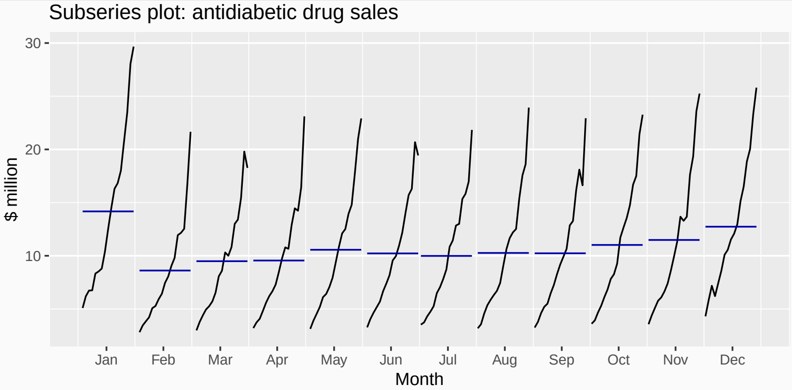


Seasonal data



Seasonal subseries plots

```
# seasonal plot
ggsubseriesplot(a10) + ylab("$ million") +
  ggtitle("Subseries plot: antidiabetic drug sales")
```



Seasonal subseries plots

Remarks:

- 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()`
- see also `ggseasonplot()`

Outline

1. Time series data
2. Time plots
3. Seasonal plots
4. **Seasonal or cyclic?**
5. Autocorrelation
6. White noise
7. Your turn

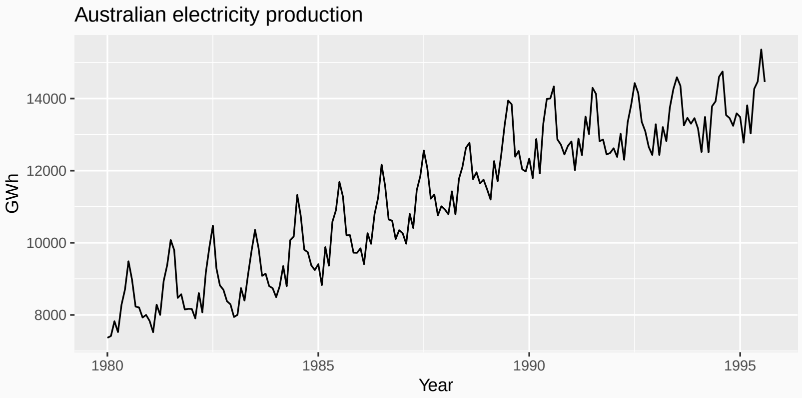
Time series patterns

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

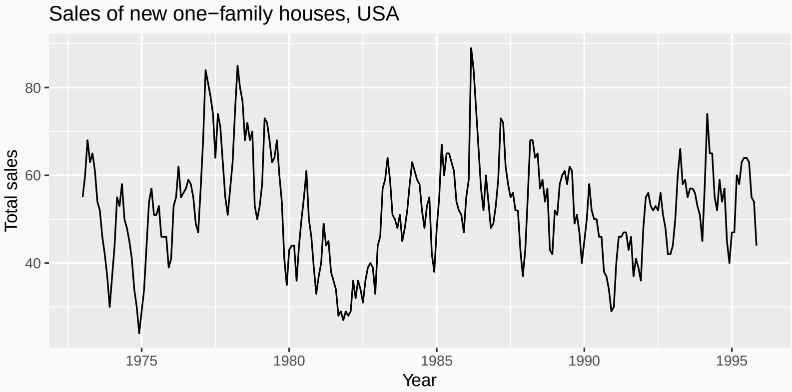
Seasonal: pattern exists when a series is influenced by seasonal factors (eg, the quarter of the year, the month, or day of the week)

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

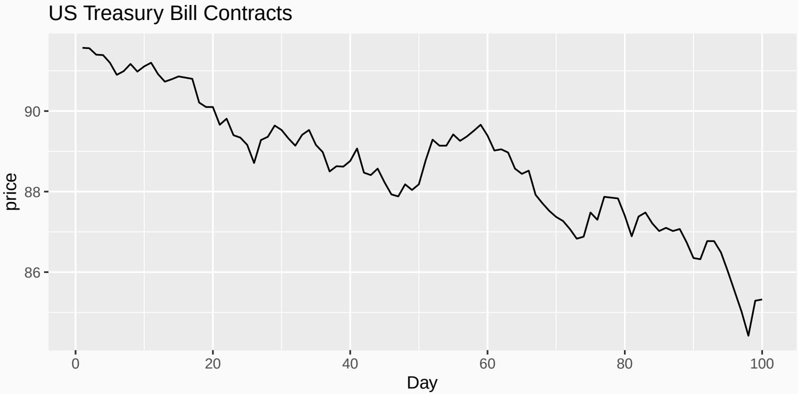
Time series patterns



Time series patterns



Time series patterns



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

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.

Outline

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 c_k and the sample autocorrelation at lag k by r_k. Define

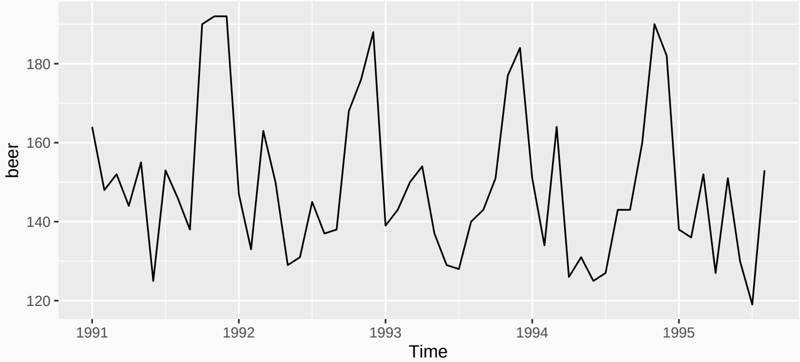
$$c_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

$$\text{and} \qquad r_k = c_k/c_0$$

- r_1 indicates how successive values of y relate to each other
- r_2 indicates how values two periods apart relate to each other
- etc.

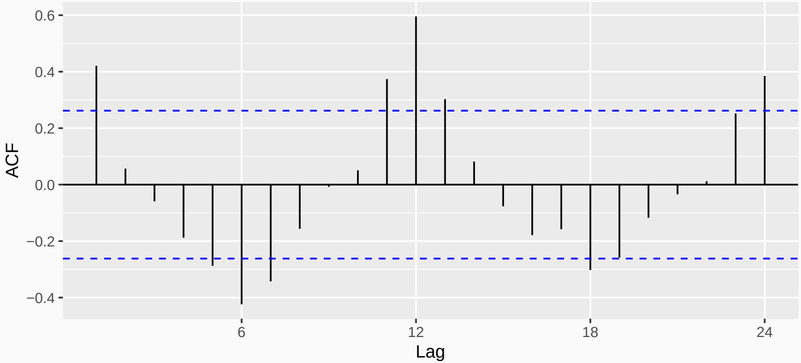
Autocorrelation

```
# plor beer data
autoplot(beer)
```



Autocorrelation

```r
# plot autocorrelations
ggAcf(beer) + ggtitle("")
```



Autocorrelation

Remarks:

- r_4 is higher than for the other lags due to **the seasonal pattern in the data**: the peaks tend to be **4 quarters** apart
- r_2 is more negative than for the other lags because troughs tend to be 2 quarters behind peaks
- together, the autocorrelations at lags $1, 2, \ldots,$ make up the *autocorrelation* or ACF
- and the plot is known as a **correlogram**

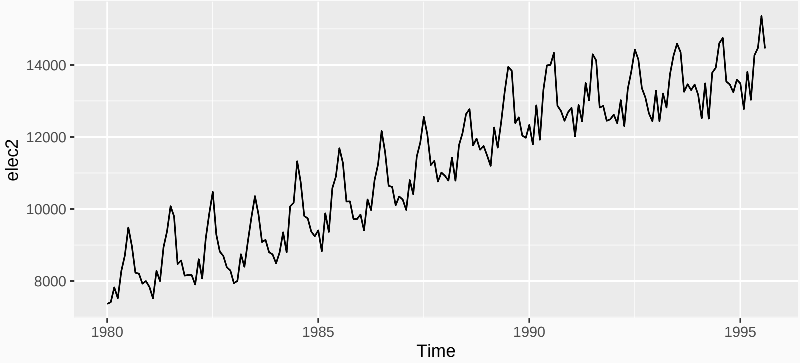
Trend and seasonality in ACF plots

Remarks:

- 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 (ie, at multiples of the seasonal frequency)
- when data are trended and seasonal, you see a combination of these effects

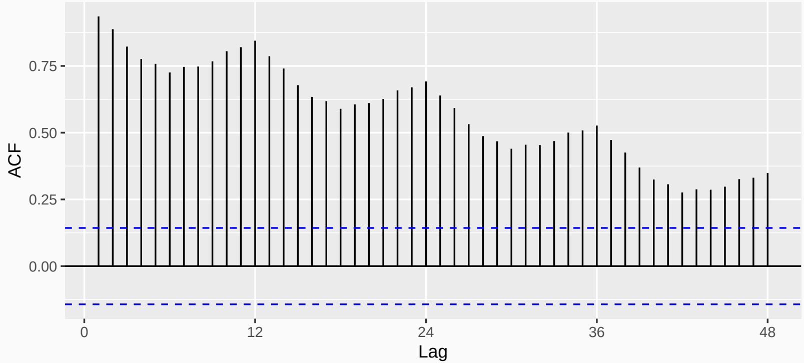
Aus monthly electricity production

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



Aus monthly electricity production

```r
# plot autocorrelations
ggAcf(elec2, lag.max = 48) + ggtitle("")
```



Aus monthly electricity production

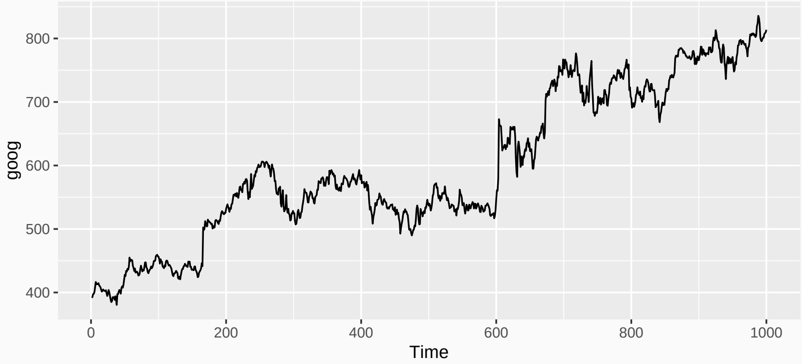
Time plot shows clear trend and seasonality.

The same features are reflected in the ACF.

- the slowly decaying ACF indicates trend
- the ACF peaks at lags 12, 24, 36, $\ldots$, indicate seasonality of length 12

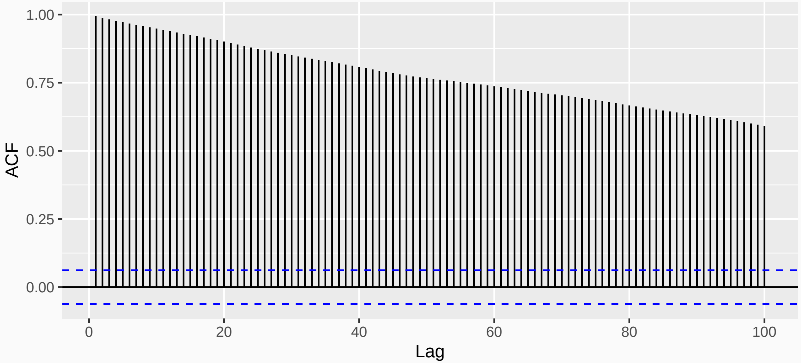
Google stock price

```r
# plot data
autoplot(goog)
```



Google stock price

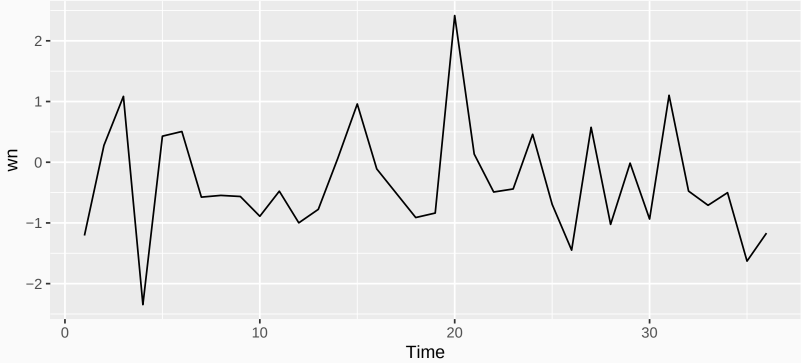
```
# plot autocorrelations
ggAcf(goog, lag.max = 100) + ggtitle("")
```



Outline

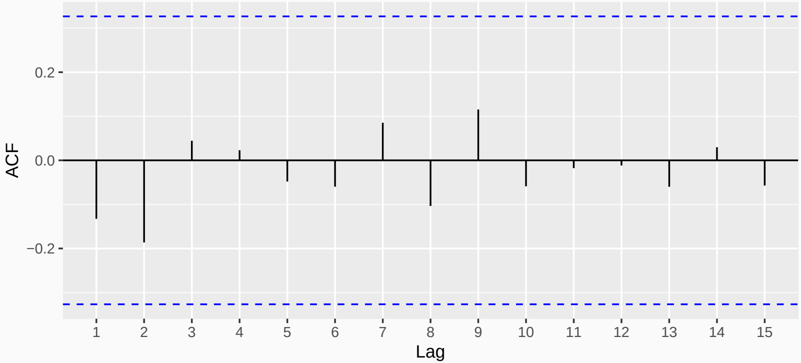
Example: White noise

```r
# simulate white noise process
wn <- ts(rnorm(36))
autoplot(wn)
```



Example: White noise

```r
# plot autocorrelations
ggAcf(wn) + ggtitle("")
```

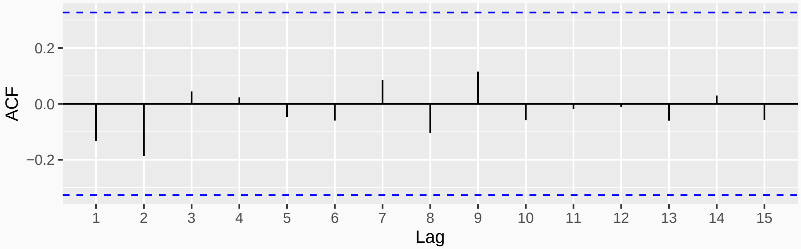


Sampling distribution of autocorrelations

Sampling distribution of r_k for white noise data is asymptotically N(0,$1/T$).

- 95% of all r_k for white noise should lie within $\pm 1.96/\sqrt{T}$
- if this is not the case, the series may not WN
- common to plot lines at $\pm 1.96/\sqrt{T}$ when plotting ACF (these are the ***critical values***)

Autocorrelation



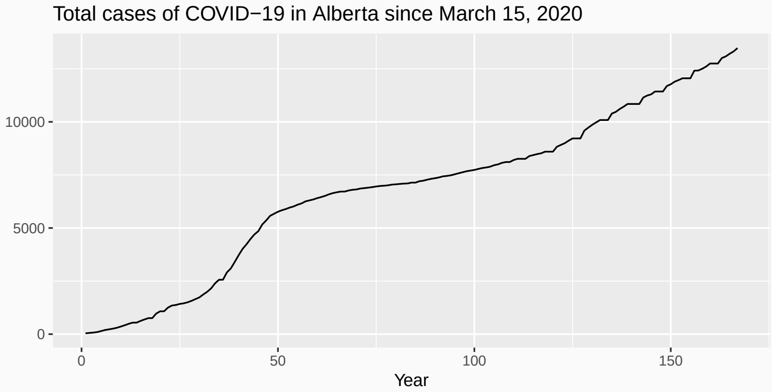
Example:

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

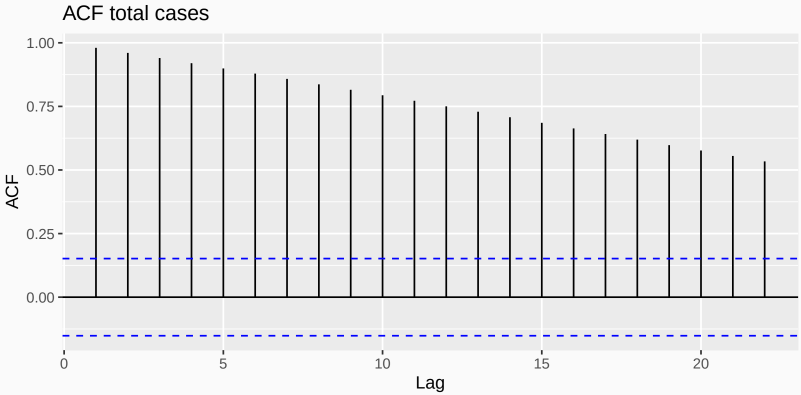
All autocorrelation coefficients lie within these limits, the data appear to be white noise.

Outline

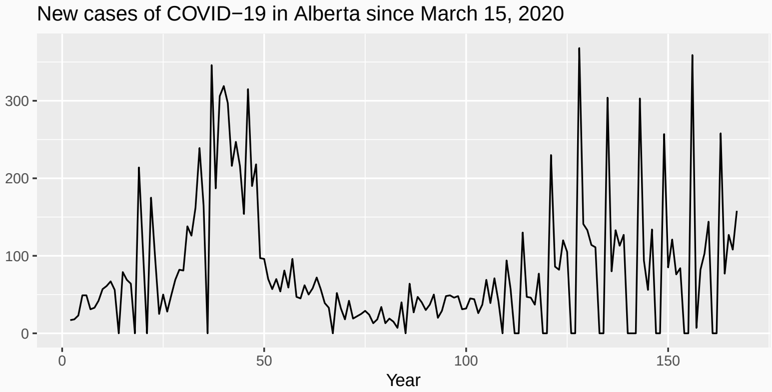
COVID-19 in Alberta



COVID-19 in Alberta



COVID-19 in Alberta



COVID-19 in Alberta