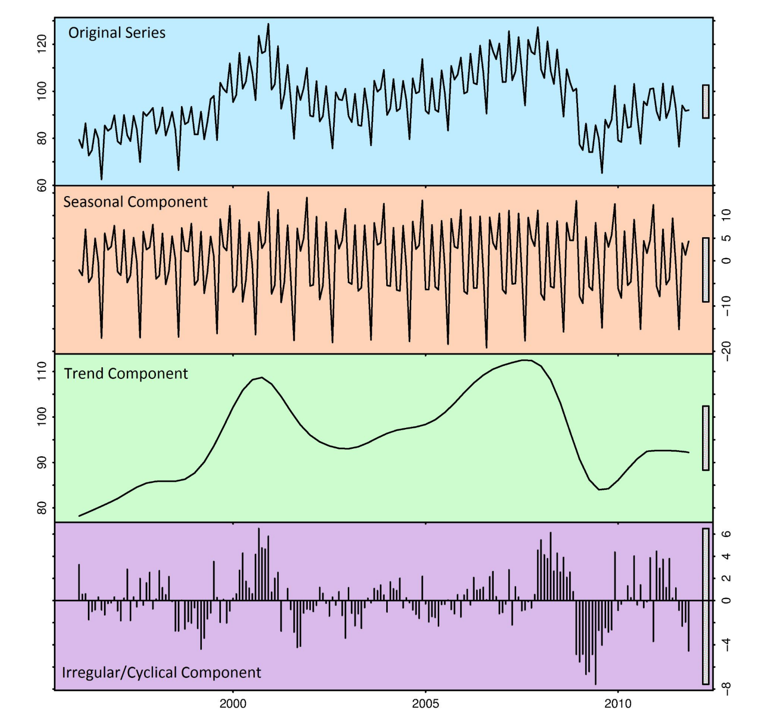
Time Series Decomposition

By: Anna Moy, Richie Rivera, Zig Dukurav, Enid Roman, Natalie Kalukeerthie

What is time series decomposition?

- Time series decomposition is breaking down the data into individual components such as trend, seasonality and residual.

- This process can help simplify complex data and discover patterns and variations within time series datasets.



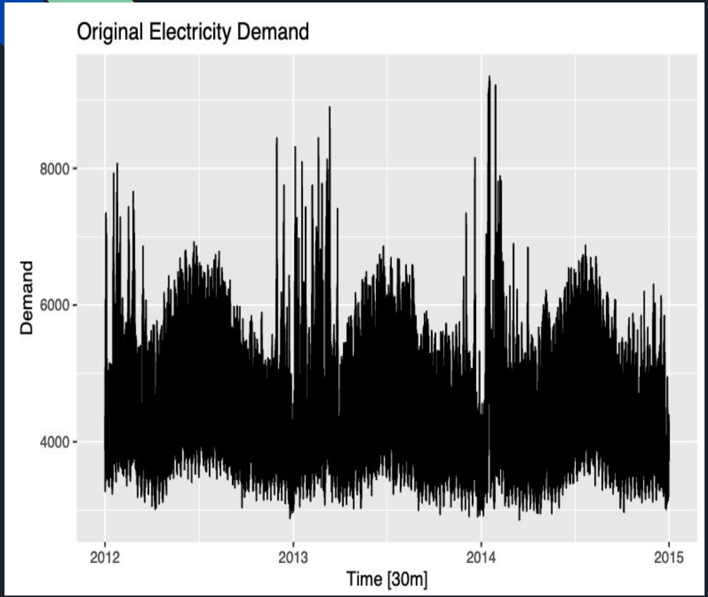
How is it useful?

- Businesses increasingly use time series data to understand changing trends and patterns.

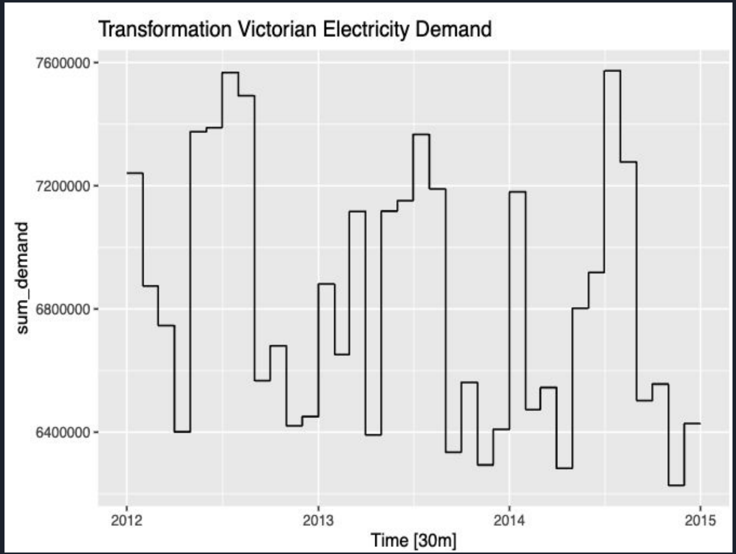
- By analyzing temporal data patterns, companies can better forecast demand, allocate resources, predict customer behavior, and make informed financial decisions.

Types of Transformations

Calendar Adjustments

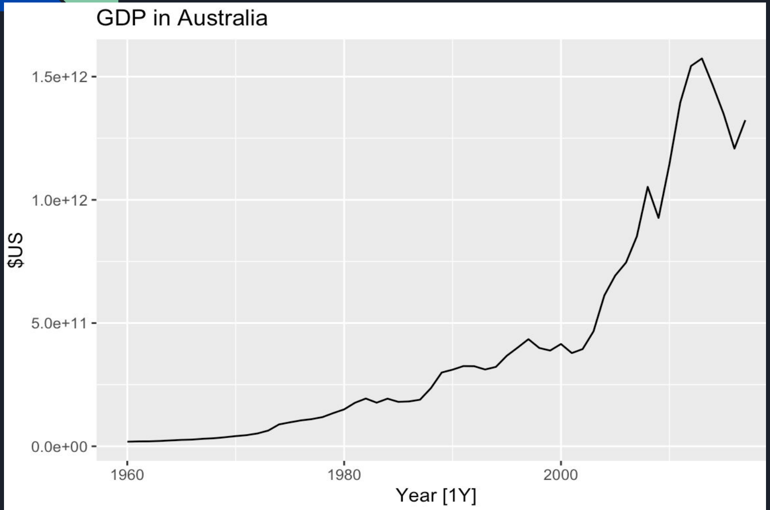


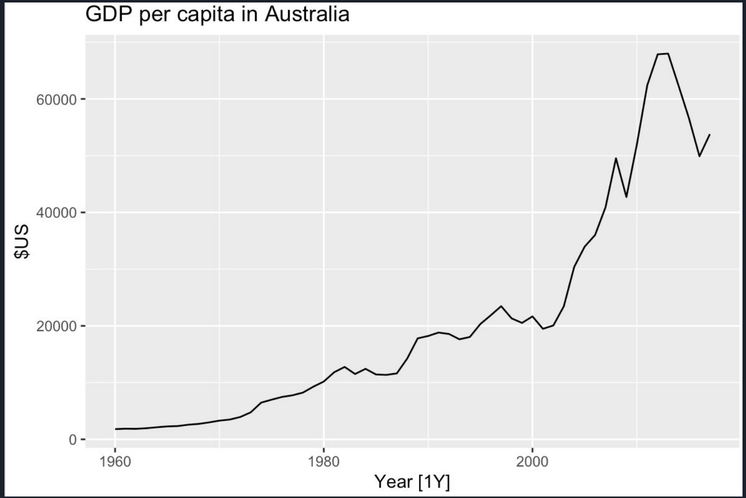

```r
vic_elec %>%
  autoplot(Demand) +
  labs(title = "Original Electricity Demand")
```




```r
vic_elec %>%
  mutate(month = yearmonth(Date)) %>%
  group_by(month) %>%
  mutate(sum_demand = sum(Demand)) %>%
  autoplot(sum_demand) +
  labs(title = "Transformation Victorian Electricity Demand")
```

Population Adjustments



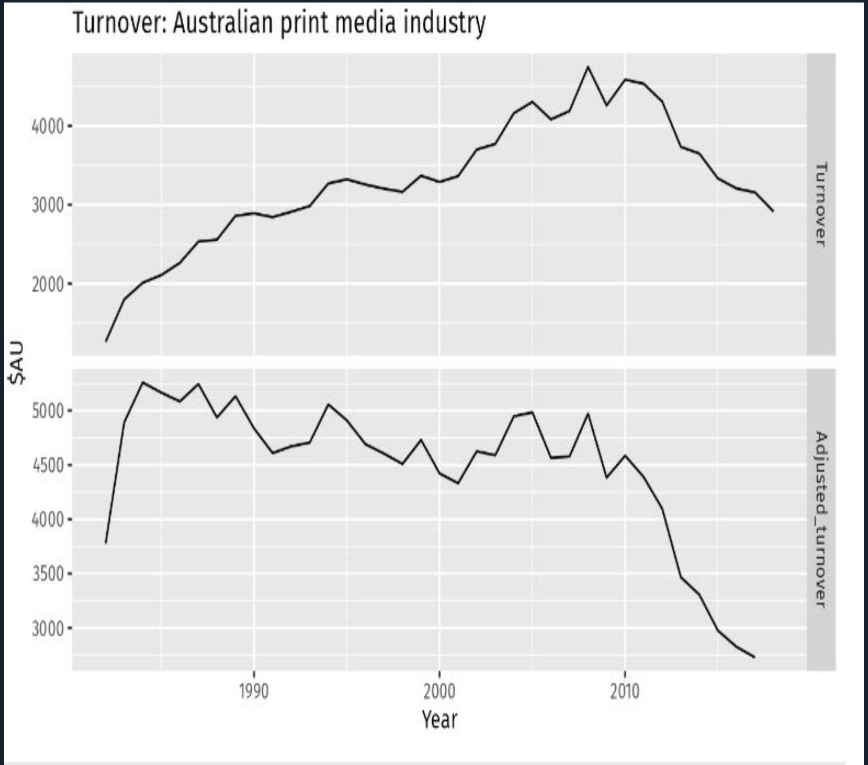



```
global_economy %>%
  filter(Country == "Australia") %>%
  autoplot(GDP) +
  labs(title = "GDP in Australia" , y = "$US")
```

```
global_economy %>%
  filter(Country == "Australia") %>%
  autoplot(GDP/Population) +
  labs(title = "GDP per capita in Australia" , y = "$US")
```

Inflation Adjustments

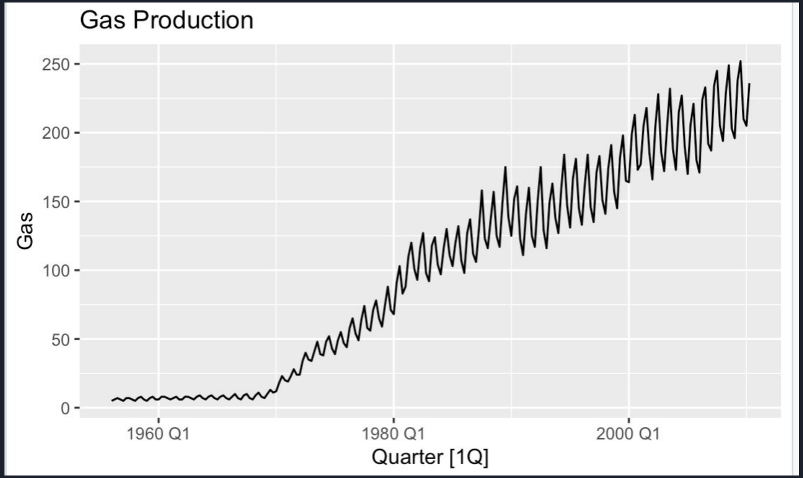
```r
print_retail <- aus_retail |>
  filter(Industry == "Newspaper and book retailing") |>
  group_by(Industry) |>
  index_by(Year = year(Month)) |>
  summarise(Turnover = sum(Turnover))
aus_economy <- global_economy |>
  filter(Code == "AUS")

print_retail |>
  left_join(aus_economy, by = "Year") |>
  mutate(Adjusted_turnover = Turnover / CPI * 100) |>
  pivot_longer(c(Turnover, Adjusted_turnover),
               values_to = "Turnover") |>
  mutate(name = factor(name,
         levels=c("Turnover","Adjusted_turnover"))) |>
  ggplot(aes(x = Year, y = Turnover)) +
  geom_line() +
  facet_grid(name ~ ., scales = "free_y") +
  labs(title = "Turnover: Australian print media industry",
       y = "$AU")
```



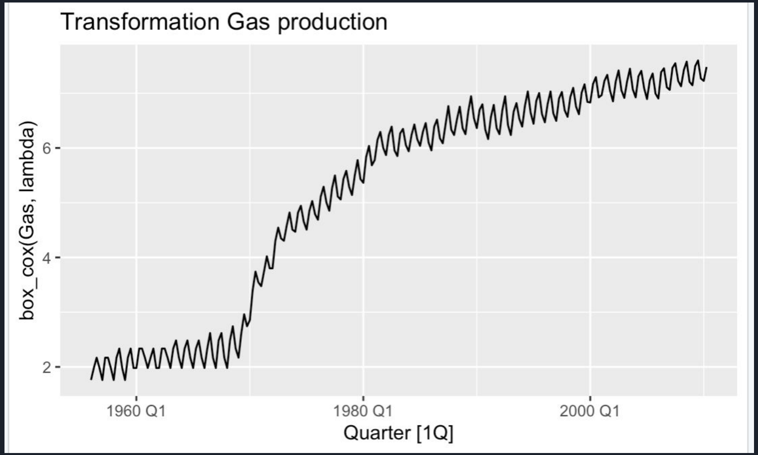
Mathematical Transformation

$$w_t = \begin{cases} \log(y_t) & \text{if } \lambda = 0; \\ (\text{sign}(y_t)|y_t|^\lambda - 1)/\lambda & \text{otherwise.} \end{cases}$$

Box-Cox Transformation - logarithms and power transformation



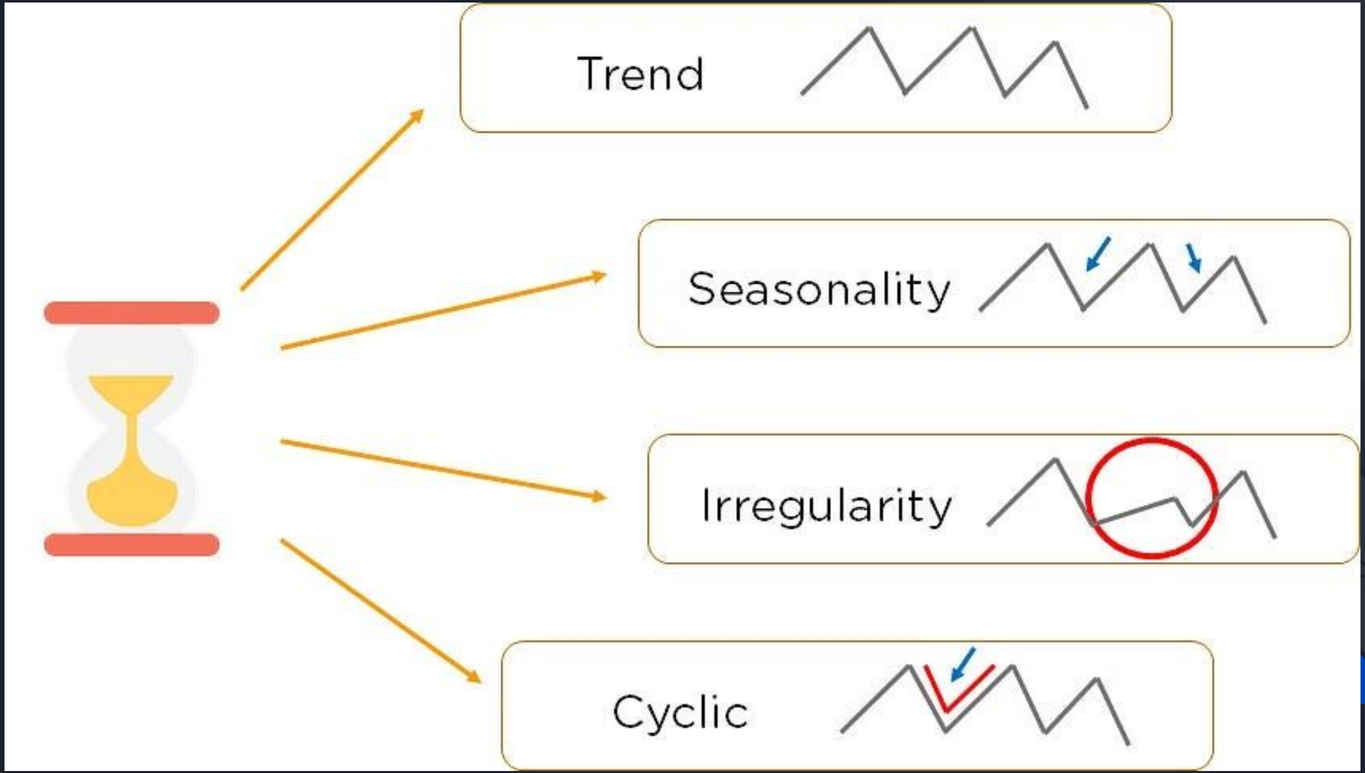

```r
aus_production %>%
  autoplot(Gas) +
  labs(title = "Gas Production")
```




```r
lambda <- aus_production %>%
  features(Gas, features = guerrero) %>%
  pull(lambda_guerrero)

aus_production %>%
  autoplot(box_cox(Gas, lambda)) +
  labs(title = "Transformation Gas production")
```

```
> lambda
[1] 0.1095171
```

Time Series Components

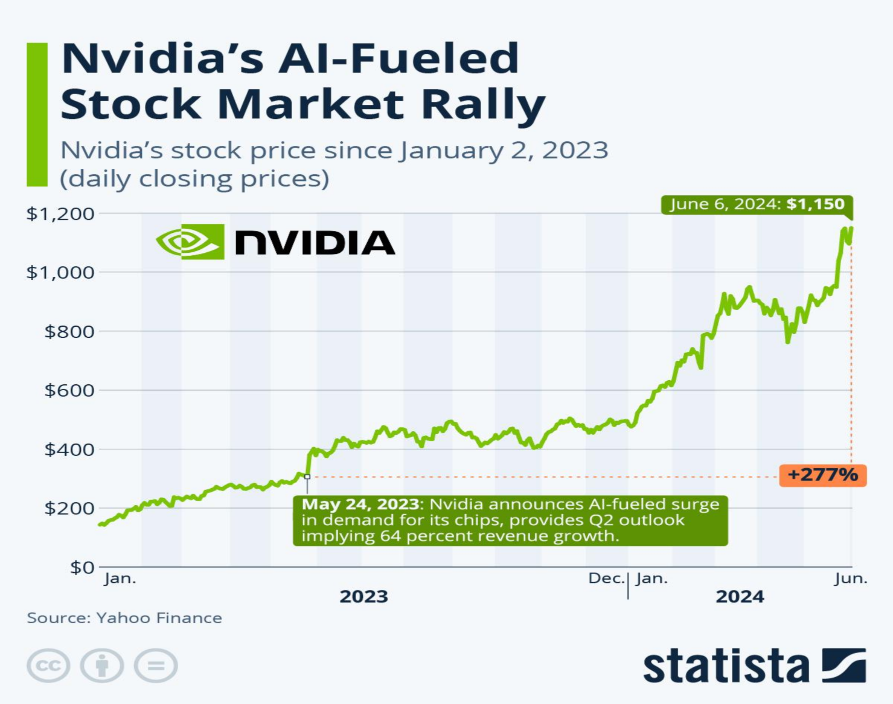


Trend

- The trend component of time series reflects the general direction of the data during a specified period of time.
- Trend displays continuous movement in one direction whether that's positive which is known as an Uptrend or negative, which is a downtrend.
- Identifying trends helps businesses, economists and policy makers to create long term strategies and forecast future outcomes.
- An example of an uptrend is S&P 500 index which on average increases 10% per year

Lets see what an Uptrend looks like!

NVDA stock has experienced astronomical gains this year with the explosion of Artificial Intelligence, which utilizes the technology of NVDA's chips. This has caused NVIDIA stock to go on a massive uptrend resulting in billions of dollars in gains for NVDA shareholders. NVDA stock experienced a 1,000% gain from October 2022 to June 2024!



Seasonality

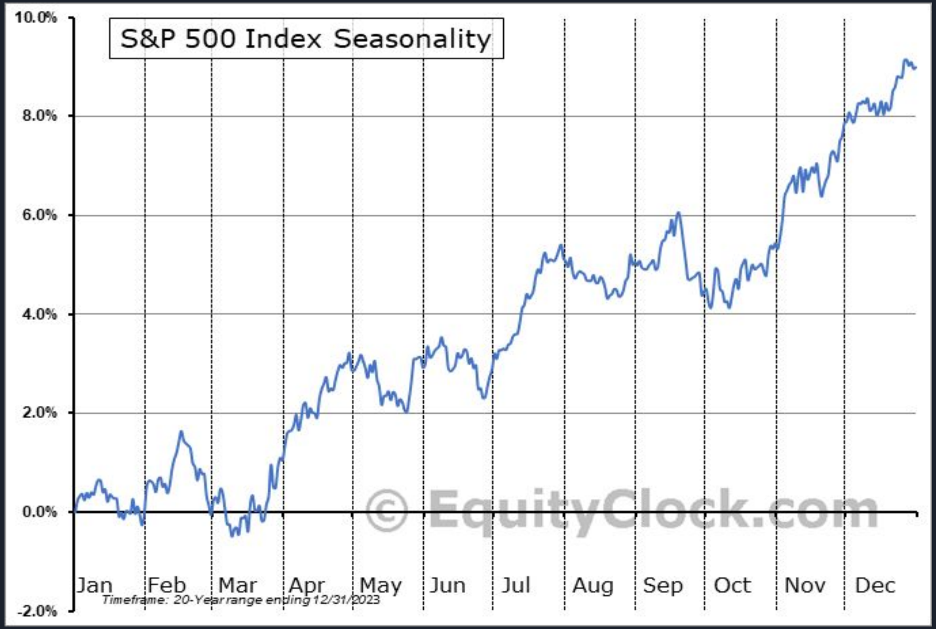
Seasonality refers to predictable patterns that repeat at regular intervals. These are regular intervals such as daily, weekly, monthly and yearly but seasonality patterns usually weekly or monthly intervals. These patterns are often driven by factors such as weather, holidays, or cultural events. Some great examples of seasonality are

Key Characteristics of Seasonality:

- Fixed Frequency: Seasonal patterns occur consistently within fixed periods, like quarters of a year or specific months **(ex: I think Jeff would agree the 4th quarter is usually the best for retail brands such as Best Buy).**
- Predictable
- Influences

S&P 500 Seasonality Chart

A great resource that can be utilized to view time series components such as seasonality of charts can be explored on equityclock.com. This chart below shows the different patterns that the S&P 500 index has displayed during time periods of the year.



Cycles

Cyclic components refer to long-term, irregular fluctuations around the trend that are not as fixed or regular as seasonal patterns. It is crucial to understand cycles and understanding cyclic patterns is crucial for long-term business and economic planning. These examples include:

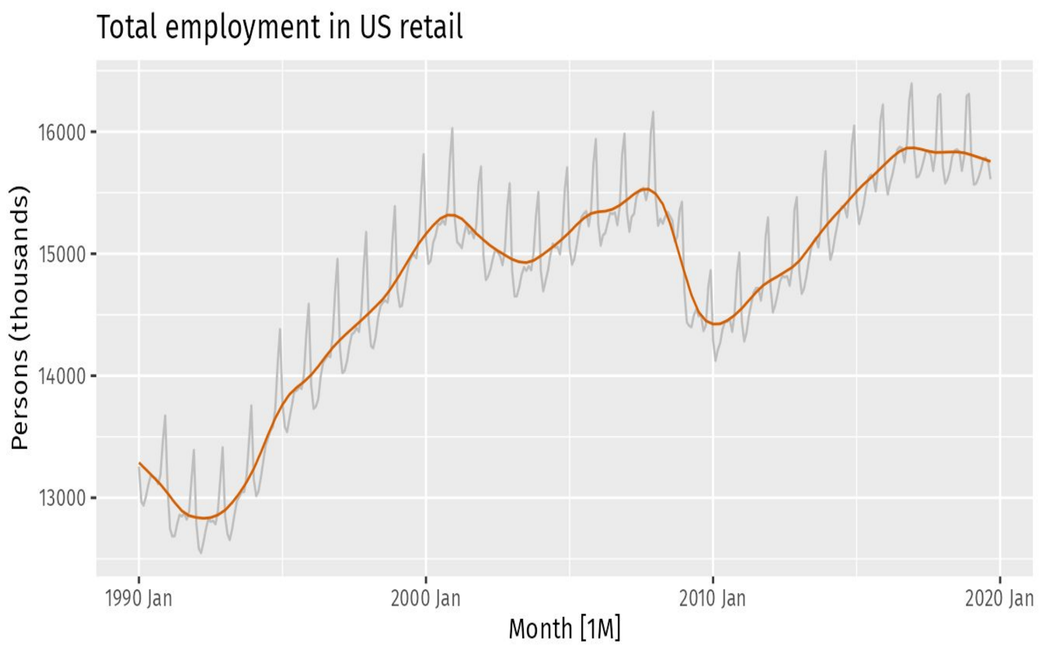
- Business Cycles: Expansion, peak, contraction, and trough phases of an economy.
- Real Estate Market: Property prices may follow a cycle of growth and decline over several years.
- A line graph showing long-term rises and falls, not evenly spaced and distinguished like seasonality.

Key Characteristics:

- No Fixed Period: Cycles occur over long periods, but the duration of each cycle can vary, unlike seasonality.
- Economic Drivers: Usually driven by economic conditions like booms and recessions which **QE can play a role in**
- Difficult to Predict: (Ex: Covid is a condition that changed many cycles but isn’t something we could have forecasted.)

Cycles

This is a chart from the Forecasting: Principles and Practice book. This chart displays the overall cycles of total employment in US retail. We can also see the seasonality curves throughout the overall cycles. There are also uptrends and downtrends that occur within the cycle. There was a sharp uptrend from the mid 1990's to about 2001. There was then a sharp downtrend from about 2008 to 2010.



Moving Averages

History

- Developed in the 1920s, moving averages smooth time series data.
- They remove noise to reveal trends in statistical analysis.

Moving Average Time Series Model in Time Series Forecasting

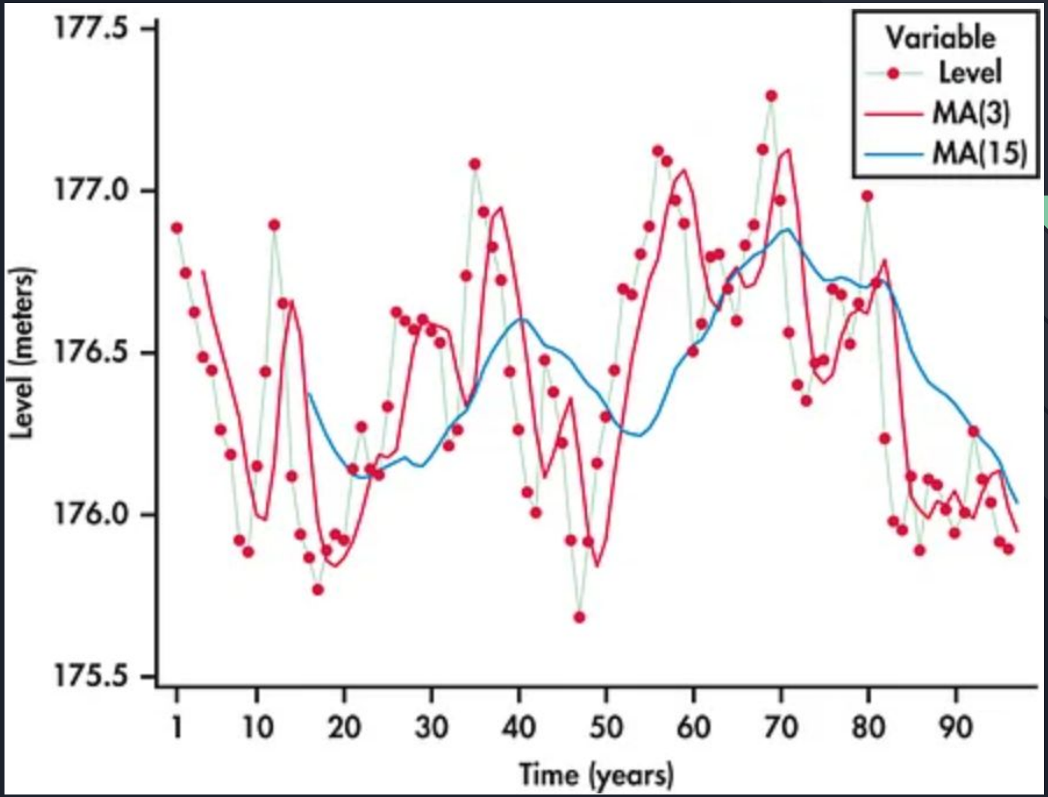
- Smooths short-term fluctuations and predicts long-term trends.
- Prevents misleading insights by managing recent spikes (e.g., stock prices, COVID-19).

What is the Moving Average Method in Time Series Analysis?

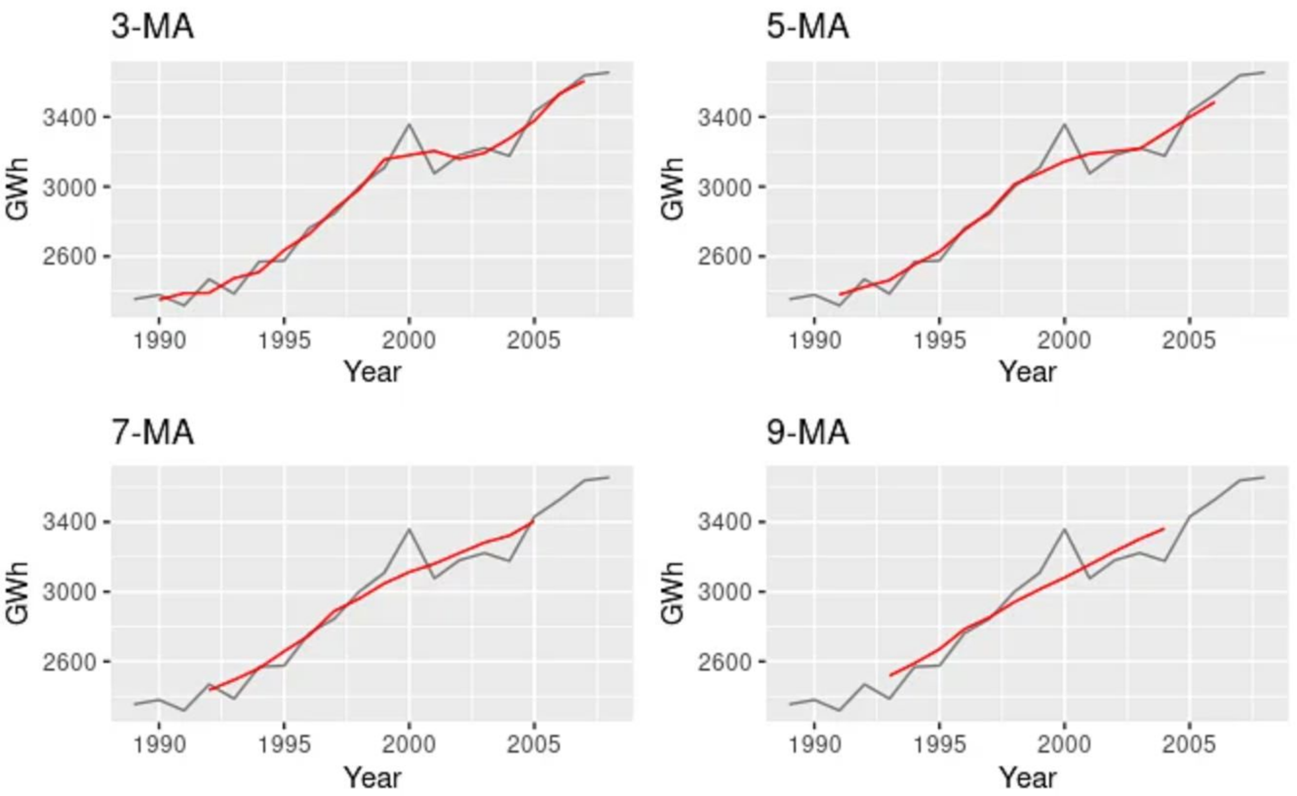
- Helps traders generate signals by analyzing price movements.
- A model used in time series analysis to forecast trends and identify patterns.
- Calculated by averaging data points over a specific period.

Example: A sliding window size of 3 and data points 10, 20, 30, 40, 50, 60, the first average would be (10 + 20 + 30) / 3 = 20. The window then shifts forward, averaging the next three points: (20 + 30 + 40) / 3 = 30, and so on.

Moving averages (MA) of two window sizes are applied and plotted.



Similarly, an upward trend is more apparent when using a larger moving average sliding window size.



Types of Moving Averages

- Centered moving averages
- Moving Averages of Moving Averages
- Weighted Moving Averages

Advantages

- Noise Reduction
- Stationarity
- Ease of Understanding
- Combination of Model

Disadvantages

- Assumption of Stationarity
- Limited Forecasting Range
- Sensitivity to Outliers
- Identification of Model Order

Examples of Moving Average

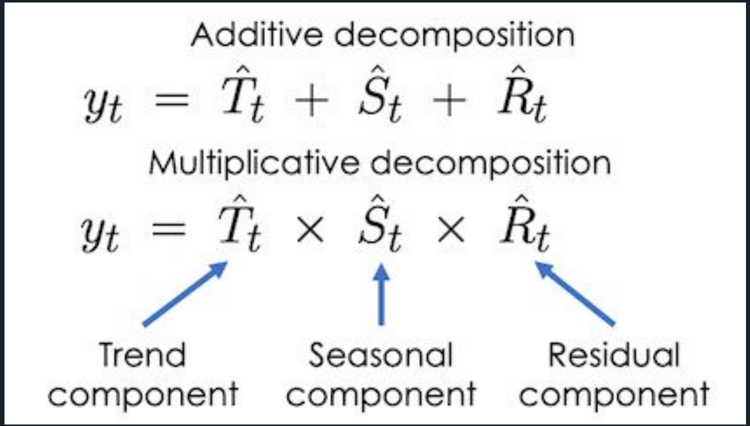
- Hospitality Industry
- Weather Study Forecast
- Fiscal Data

Methods of Decomposition

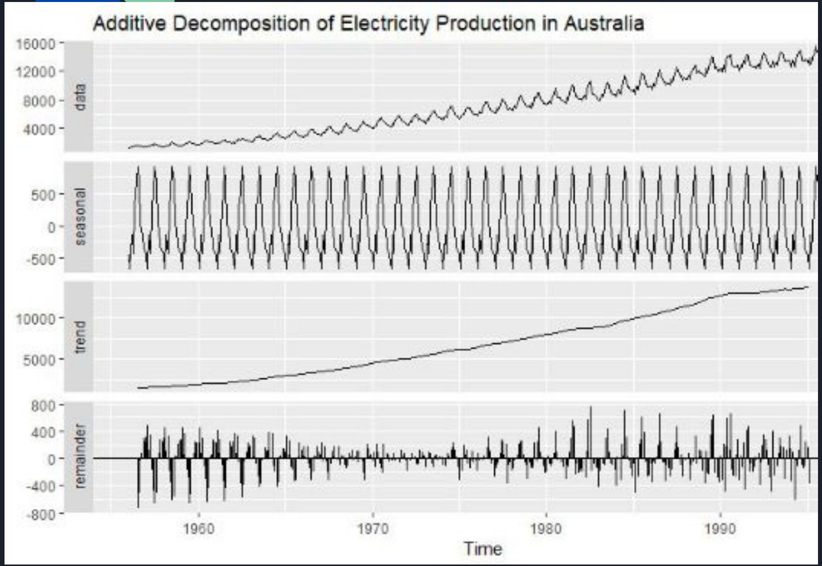
Classical Decomposition

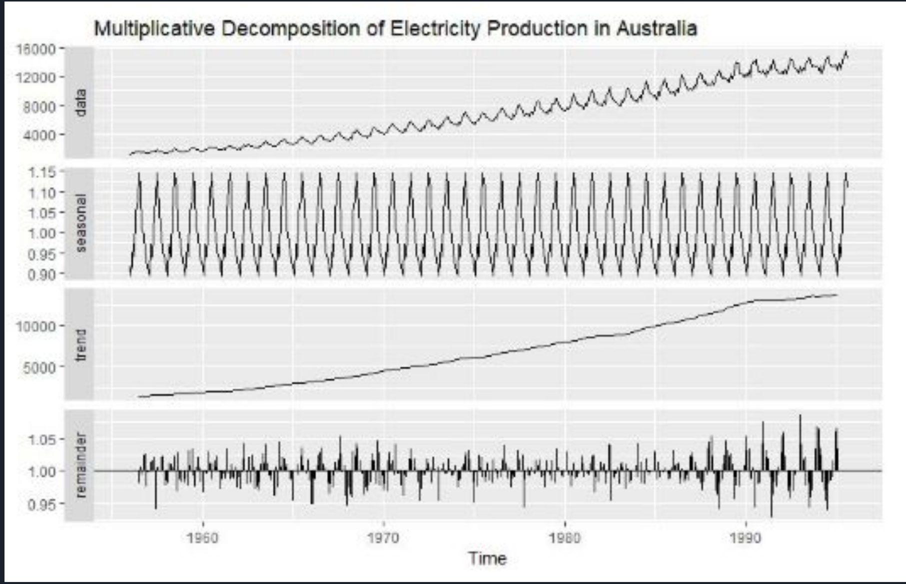
- Two forms: Additive and Multiplicative

  The additive model is useful when the seasonal variation is relatively constant over time.
  The multiplicative model is useful when the seasonal variation increases over time.
- We assume that the seasonal component is constant from year to year.



Classical Decomposition






```
elec %>%
decompose(type="additive") %>%
autoplot() +
ggtitle("Remainder of Electricity
Production in Australia")
```

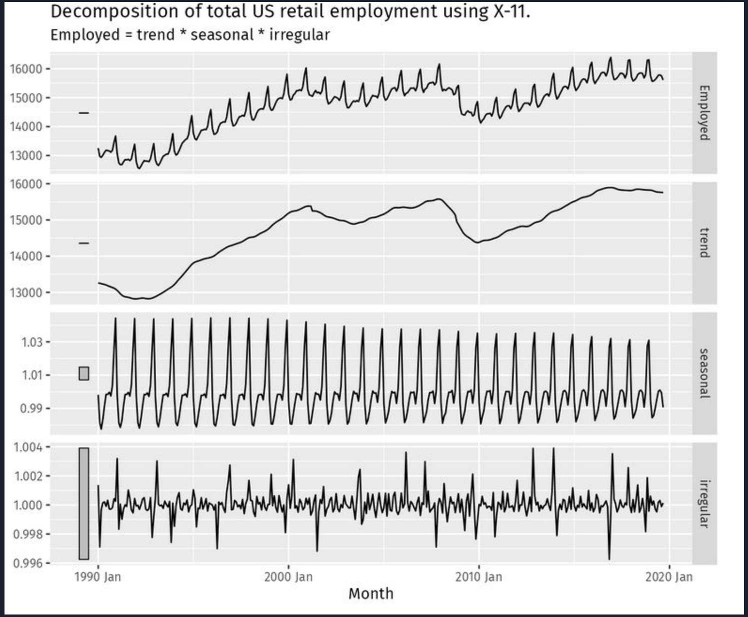
```
elec %>%
decompose(type="multiplicative") %>%
autoplot() +
ggtitle("Remainder of Electricity
Production in Australia")
```

Classical Decomposition

- While classical decomposition is still widely used, it is not recommended for some of these reasons:
    - Considers the seasonal component to be constant
    - Trend at beginning and end of the time period are unavailable.
    - Over smoothing trends
    - Not robust to unusual values from small number of periods

X11 Decomposition

- Trend-cycle estimates are available for all observations
- Variations in trends are more visible
- X-11 can handle trading day variation, holiday effects and the effects of known predictors.
- Methods available for both Additive and Multiplicative decomposition

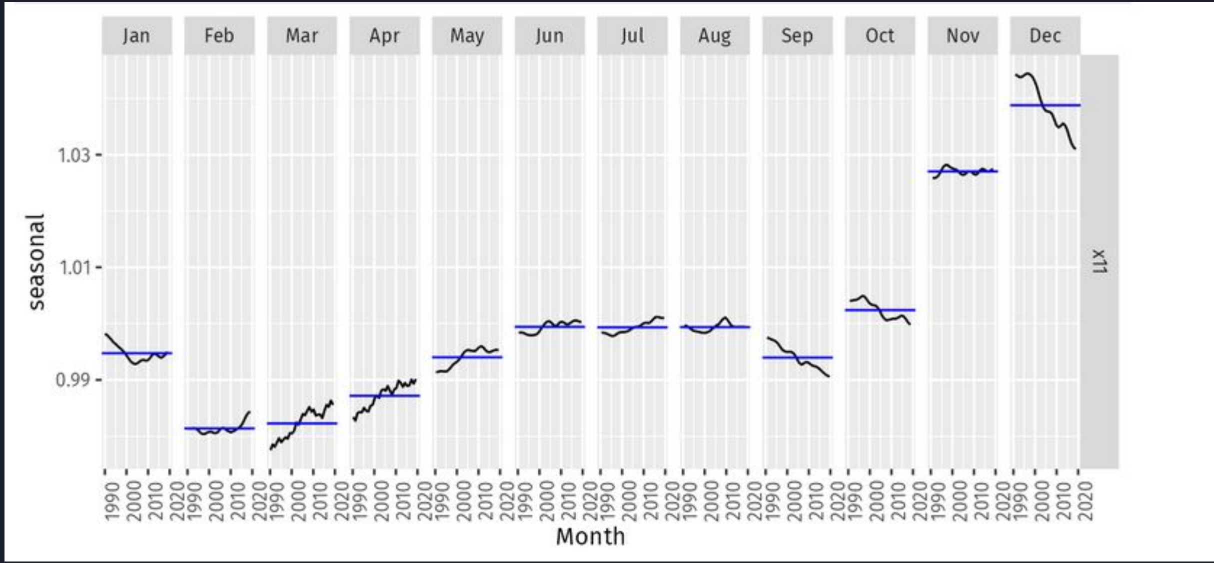



```r
x11_dcmp <- us_retail_employment |>
  model(x11 = X_13ARIMA_SEATS(Employed ~ x11())) |>
  components()
autoplot(x11_dcmp) +
  labs(title =
    "Decomposition of total US retail employment using X-11.")
```

X11 Decomposition - Seasonal Plot

```r
x11_dcmp |>
  ggplot(aes(x = Month)) +
  geom_line(aes(y = Employed, colour = "Data")) +
  geom_line(aes(y = season_adjust,
                colour = "Seasonally Adjusted")) +
  geom_line(aes(y = trend, colour = "Trend")) +
  labs(y = "Persons (thousands)",
       title = "Total employment in US retail") +
  scale_colour_manual(
    values = c("gray", "#0072B2", "#D55E00"),
    breaks = c("Data", "Seasonally Adjusted", "Trend")
  )
```

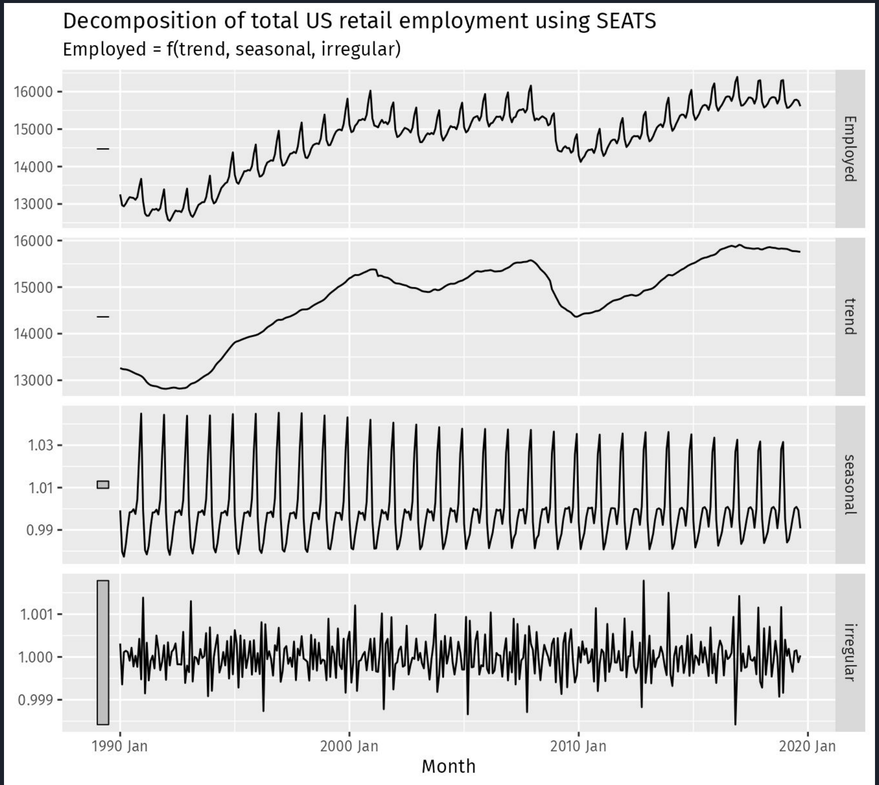
X11 Decomposition - Subseries Plot




```
x11_dcmp |>
  gg_subseries(seasonal)
```

SEATS Decomposition

- Acronym for “Seasonal Extraction in ARIMA Time Series”
- ARIMA is used to build each of the components
    - We will go over ARIMA in Weeks 7 & 8
- Better than Classical Decomposition and X11 at removing noise and handling randomness

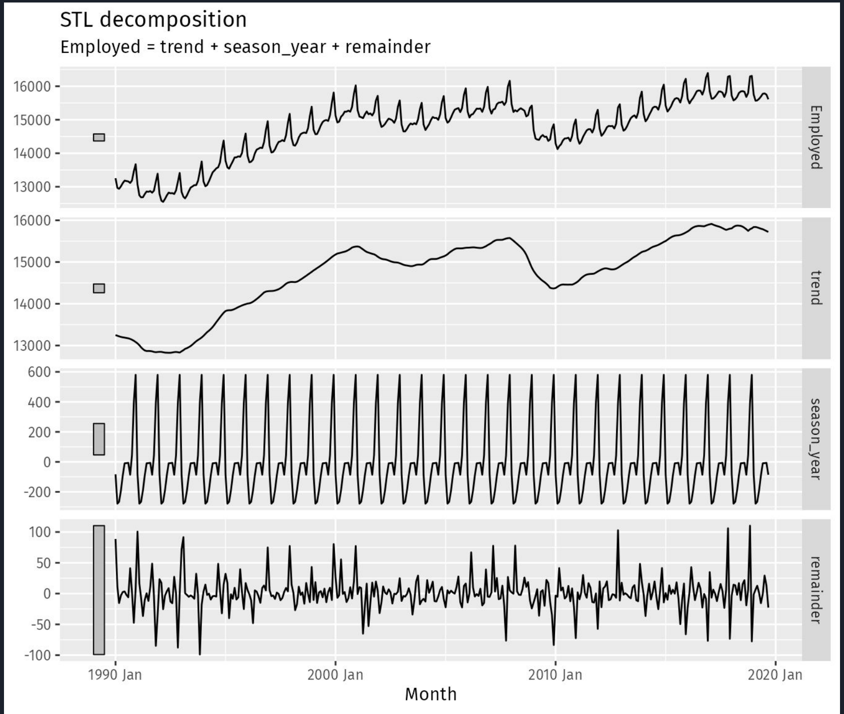



```
seats_dcmp <- us_retail_employment |>
  model(seats = X_13ARIMA_SEATS(Employed ~ seats())) |>
  components()
autoplot(seats_dcmp) +
  labs(title =
    "Decomposition of total US retail employment using SEATS")
```

STL Decomposition

- Acronym for “Seasonal and Trend decomposition using Loess”
  - Loess is a method to estimate non-linear relationships

- The trend-cycle window and seasonal window must be chosen




```
us_retail_employment |>
  model(
    STL(Employed ~ trend(window = 7) +
                   season(window = "periodic"),
    robust = TRUE)) |>
  components() |>
  autoplot()
```

Pros & Cons of STL Decomposition

Pros

- Handles all types of seasonality
- Seasonal component can vary over time
- Trend-Cycle can vary over time
- Outliers won’t affect trend or seasonal components
- Useful for data with occasional anomalies

Cons

- Cannot natively handle multiplicative decompositions
- Requires manual calendar adjustments

Appendix:

- https://www.projectpro.io/article/moving-average-time-series-model/716#:~:text=Pipeline%20%2D%20Part%202-,What%20is%20the%20Moving%20Average%20Method%20in%20Time%20Series%20Analysis,points%20with%20the%20computed%20value.
- https://afit-r.github.io/ts_moving_averages#centered-moving-averages
- https://statisticsbyjim.com/time-series/moving-averages-smoothing/
- https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/weighted-moving-average-wma