Time Series Patterns

Handout

Descriptors for Time Series Plots

- **Trend**
    - Exists when there is a long-term increases or decrease in data
    - It does not have to be linear
- **Level**
    - The level of a series refers to its height on the ordinate axis
    - Series with a trend will have a changing level **but** a series whose level is changing may not have a trend
- **Seasonal**
    - Exists when a series is influenced by factors such as quarters of the year, month, day of the week, or time of day.
    - Periods of data
        - Months → 12
        - Quarterly → 4
- **Cyclic**
    - Exists when there are rises and falls that are not of a fixed period.
    - With one cycle alone, it is difficult to distinguish from a trend
- **Questions for identifying notable features of a time series plot**
    1. Is there a trend?
        a. Is it linear or non-linear?
        b. Does it change over time?
    2. Is there a seasonal effect?
        a. Is it obvious what the period is?
    3. Are there sudden dips or spikes? When did they occur?
        a. This could signal certain effects happening

4. Does the process display a non-constant variance?
    a. This could be observed if the variability across time increases over time.
    b. Note that to observe this, we may need to decompose it first into the trend, seasonality and residuals components to observe which component has non constant variance

Seasonal Effects

- Refers to the patterns that repeat at regular intervals within a time series data.
- These patterns correspond to certain time periods (e.g. days, weeks, month, or years)

> 💡 **Period** - Time duration before a season repeats itself. (For instance, Monthly Data will repeat every 12 months and Quarterly Data will repeat every 4 months)
> **Season** - Individual parts of a period (For instance, Monthly Data → Each month is a season)

gg_season()

- Time plot is chopped up into individual periods (contains one cycle of the seasons) aligned and plotted on an axes for a single period
- For the below example, each of the period (years) is a line and they are plotted against the seasons

▼ **Code Example**

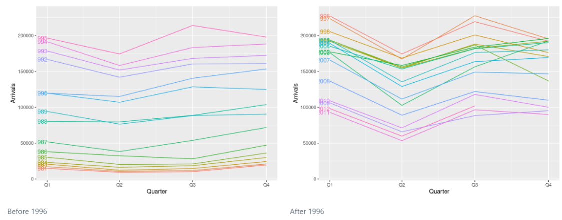
```
filter(aus_arrivals, Origin=="Japan",
       Quarter <= yearquarter("1995 Q4")) %>%
  gg_season(y=Arrivals, labels="left") +
  lims(y=c(9000, 230000))
filter(aus_arrivals, Origin=="Japan",
       Quarter > yearquarter("1995 Q4")) %>%
  gg_season(y=Arrivals, labels="left") +
  lims(y=c(9000, 230000))
```

gg_subseries()

- We facet is based on each of the seasons and each period of a particular season is extracted out and plotted in one time plot.
- For the below example, each quarter is a facet and it is plotted across the time periods which are years

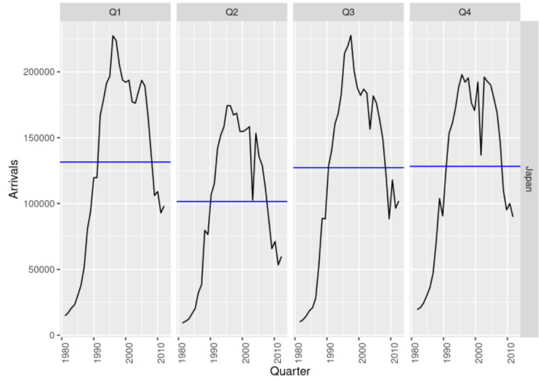
▼ **Code Example**

```
filter(aus_arrivals, Origin=="Japan") %:
  gg_subseries(y=Arrivals)
```



▼ **When should I use this?**

- Examine the behaviour and variations within individual seasons
- Trying to understand how the data behaves within each season independently



▼ **When should I use this?**

- Compare the patterns across different seasons side by side
- Identifying seasonality or other recurring patterns across the entire dataset

Autocorrelation Function (ACF)

▼ Sample ACF

$$r_h = \frac{\sum_{t=h+1}^{T}(y_t - \bar{y})(y_{t-h} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

where

- T is the length of the series
- $\bar{y}$ is the sample mean of the T observations

▼ Meaning of Formula

- Each r_h is meant to estimate the correlation between y_t and y_{t-h}, which we denote as ρ_h which is the correlation at lag h

▼ Assumptions

Note that for this sample ACF, we have the following assumptions, if these assumptions are not meant then it is not a good estimate of the correlation and could give misleading results.

1. **Constant mean** across time periods is assumed.
    a. If mean is changing over time, the ACF plot does not seem to decay as expected
2. Formula assumes that the Covariance between points with the same time lags are the same.
    a. For example, (y_{10}, y_{20}) and (y_{20}, y_{30}) are assumed to have the same covariance
3. **Constant variance** is assumed. (This can be seen from the denominator)
    a. If variance is changing over time, there could be varying amplitudes of the ACF values as you move through the lags. The variability in ACF values might not be consistent

▼ **Uses**

1. Checks if the residuals are white noise
2. Indicate if there is a strong trend remaining in the data
3. Indicates seasonality, if any
4. If it dies down relatively quickly, ARIMA Models can be appropriate

▼ **Things to look out for in ACF Plots**

1. Does not decay
    a. This means that there is strong trend in the original time series plot. This is because it will take a very big time lag for the time data points to no longer be correlated.
2. Seasonality. We will look at for the different periods, maybe there are high ACF values or negative values.
    a. This indicates seasonality effects that we should adjust for. This is only if there are observable peaks at each of the seasons
3. White Noise
    a. If there does not seem to be any significant deviation from 0 for the ACF lag points, then it shows that it could most likely be a white noise.

b. Note that there are 95% Confidence Intervals (CI) for each lag and it helps to determine if the values are significantly different from 0

▼ White Noise Model

- When we have a WN model, we will see that it just waivers around the mean and fluctuates randomly
- Time Series that consists of uncorrelated random variables $\{e_t\}$ such that for all $t \geq 1$

$$E(e_t) = 0, Var(e_t) = \sigma_e^2$$

▼ Representation

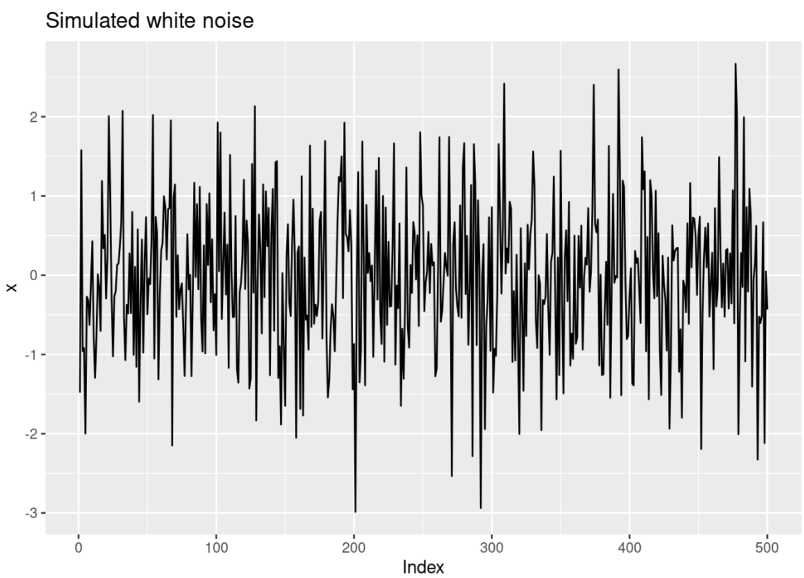
$$e_t \sim WN(0, \sigma_e^2)$$

▼ Purpose

- Many useful processes (time series) can be constructed from White Noise despite it being uninteresting.

▼ Gaussian White Noise

- This is when $e_t \sim N(0, \sigma_e^2)$ for all t.
- This is **i.i.d** Gaussian WN.



Sample plot of Gaussian White Noise

▼ Random Walk Model