Data Analysis, Clustering and Visualisation

*TIME SERIES DATA:THE SP500 INDEX:*

1. The candles in the upper part of the following chart shows the market’s open, high, close and low price for a each single day. The real body of the candles is the wider area in the middle. Which represents the price range between the open and close of the day’s trading.

The green candlesticks mean that on a particular day the opening price was lower than the closing price. Which means that price moved up on that day. The orange candlesticks mean that the opening price was higher than the closing price. Which means that the price went down on that day.

For example in the following chart on 03/Sep/2009, it has a green candle so the price went up on that day. While on 01/Sep/2009 the price went down by quite a bit.

A screenshot of a cell phone

Description automatically generated

1. R code use to calculate average price and plot showing average price Av(t):

A picture containing phone

Description automatically generated

A close up of a map

Description automatically generated

1. R code to calculate returns.

A picture containing phone

Description automatically generated

Figure showing the return for last 12 months.

A picture containing bird

Description automatically generated

Figure showing the candleChart for the last 12 months.

A screenshot of a cell phone

Description automatically generated

*DEGREE DISTRIBUTION : NETWORKS ->*

1. The maximum degree for the network is 183 and the minimum is 5.
2. Comment on what the single figure with two plots tells you about the network.

A screenshot of a cell phone

Description automatically generated

A screenshot of a video game

Description automatically generated

1. R code for calculating p(k) which is the probability of observing a node with degree k and the plot of p(k) versus k.

A picture containing monitor, screen, phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. The plot above shows that the network is likely to be scale free because nodes with the higher probability are more likely to see new edges as compared to the one with the lower probability. It means that when the network grows the underlining structure or properties stays the same. That is the characteristics of this network are independent of the size of the network.

*DATA QUALITY AND STRUCTURE:*

1. The plot represents the explanatory variables x1, x2 on the x-axis and the response variable y on the y-axis. The plot shows that most of the data is pretty close to the centre middle line and with change in the explanatory variables there is an effect in the response variable. This means that x1 and x2 are good explanatory variables. And is useful method for examining the quality of the dataset.

A picture containing rain

Description automatically generated

1. The resulting plot shows that x1 and x2 are not good explanatory variables to predict y1 (response variable). The data is everywhere and there is no linear change in the response variable by changing the explanatory variable, which shows that x1 , x2 has no effect on the response variable and that is why are not good explanatory variables and is not a useful method for examining the quality of the dataset.

In summary3

A screenshot of a cell phone

Description automatically generated

A picture containing nature, rain

Description automatically generated

*VISUALISATION AND CLUSTERING :*

1. Breifly state the meaning of each variable for the data.

PopDensity = is the measurement of population. In other words how dense an area is term of it population.

IncomeperCapital = That is the average income earned per person in a given area.

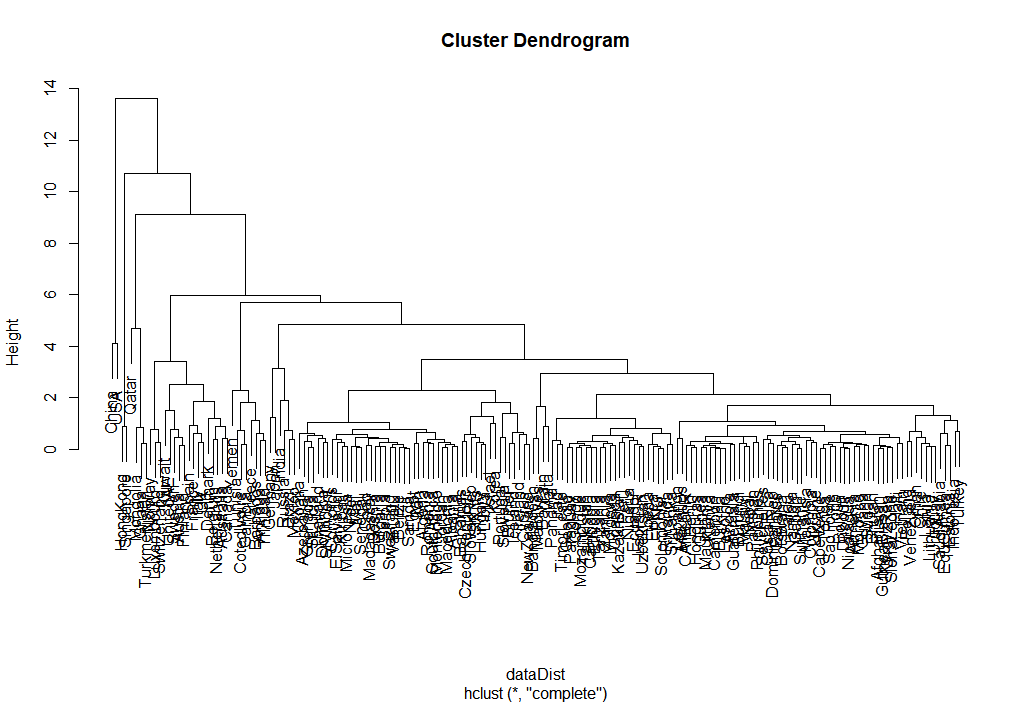
Purchasing parity = is the measurement of prices in differnt countries.

changeGDP = is to measure or make an assessment on how the economy of a country is doing.

1. Which country is similar to New Zealand and which countries are the strongest or weakest.
2. Dendrograms

A screenshot of a cell phone

Description automatically generated



1. Plot for tsne: Comment on why it’s the same or different here???

A screenshot of a cell phone

Description automatically generated