**R Diagnostic**

This is a basic R diagnostic designed to assess your familiarity with R. You will also likely come across similar problems during FRP. **Feel free to use online resources or any other materials, but it’s highly advised you work on this by yourself.**

**Part 1: Working with Data**

A large amount of time will be spent cleansing and manipulating data. This part will assess your familiarity in manipulating data types and creating new ones.

1. Please download daily data of the S&P500, Seasonally Adjusted Quarterly GDP Growth and the 10-Year Constant Maturity Rate from 2009-01-01 to 2017-01-01. Load both datasets into R. Links to both are below:

<https://fred.stlouisfed.org/series/DGS10>

<https://fred.stlouisfed.org/series/SP500>

<https://fred.stlouisfed.org/series/A191RP1Q027SBEA>

1. This data needs to be cleansed. Some days have a price level “.” which causes R to read this column as a factor type instead of numeric. Please:
   1. Remove all rows with price “.” in the CMT & SP500 datasets
   2. Reformat all price columns to type numeric
   3. Rescale the GDP dataset’s returns to be in real levels (divide everything by 100)
   4. Rename the second column of the GDP dataset to “GDPReturn”
2. It’s not a good idea to work with level data, so let’s transform the data. Please compute the daily returns of both the S&P and 10 Yr CMT and create a new column called “SP\_Return” and “CMT\_Return” respectively. The first row’s return should be NA.
3. Merge the two dataframes together into a master data frame. Please only keep rows where both dataframes have price data for.
   1. Also remove the first row since there is no return data here
4. You’ll notice we have a period mismatch; quarterly returns for GDP but daily for S&P and CMT. Please create a final table containing quarterly GDP and quarterly S&P & CMT returns. Use dates according to the quarterly GDP dataset.
   1. Also remove the first row since there is no return data here

**Part 2: Understand your Data**

Before you build models, getting a good sense of feel for your data is important. It will help you spot out outliers, skewness, etc very quickly so you can save time when building models later.

1. Provide the following information:
   1. Min, Max, 1st & 3rd quartile, Mean, Median of all 3 return columns
   2. Which days did the Max/Min returns occur for S&P and CMT columns?
      1. Bonus points: What happened on these days to justify the returns?
   3. Correlations between GDP v CMT, S&P v CMT, and GDP v S&P
   4. Standard deviation of both columns

**Part 3: Modelling & Analytics**

The final part is to build models. In R these are quite simple to do, so most of your time will be spent assessing and understanding the context of your results.

1. Let’s see if there is any predictability between the returns. Please run a regression, the explanatory variable (X) is the CMT return and GDP growth. The response variable (Y) is the SP’s return. Include an intercept term as well!
   1. Provide the coefficients
   2. Run a t-test and provide t-statistics on both coefficients (and p-values)
   3. What are the standard errors of the coefficients?
   4. What about R squared, Adjusted R Squared?
   5. Run an F-test and provide the F-statistic along with P-values
2. Is the regression model suitable for modelling this phenomena compared to just an intercept term? (Hint: look at your F-Test)