There are three files required for this homework. Two are posted, `HW08_Rt.RData`, `FM_Functions.R`. Please download Fama-French 3 factor data from Dr. Kenneth French's website, see the link below.

The Asset Return Data

There are 2 object in the file `HW08_Rt.RData`, `Rt` and `syb`. `Rt` is weekly returns as percentages (%) of 26 stocks from the following 4 industries from January 1, 2006 to February 28, 2021, total of $n = 791$ observations. As usual, `syb` is the ticker symbols of the 26 stocks. The object `Rt` is a numerical matrix, its xts class is removed.

Food (1–7): Danone SA (DANOY), General Mills, Inc. (GIS), Kellogg Company (K), Hormel Foods (HRL), Conagra Brands (CAG), Campbell Soup (CPB), McCormick & Co. (MKC). These 7 companies are in the "Packaged Foods" Industry of the "Consumer Defensive" sector.

Media (8-13): The Walt Disney Company (DIS),Liberty Global (LBTYA), ViacomCBS (VIAC), Netflix (NFLX), Live Nation Entertainment (LYV), Lions Gate Entertainment (LGF-A). These 6 companies are in the "Entertainment" Industry of the"Communication Services" Sector.

Health Care (14–19): UnitedHealth Group In. (UNH), Anthem, Inc. (ANTM), Cigna Co. (CI), Centene Co. (CNC), Humana Inc. (HUM), Molina Healthcare, Inc. (MOH). These 6 companies are in the " Healthcare Plans" Industry of the "Healthcare" Sector.

Tech (20–26): Microsoft (MSFT), Oracle Co. (ORCL), Adobe Inc. (ADBE), Salesforce inc. (CRM), Autodesk, Inc. (ADSK), Amdocs Ltd (DOX), NortonLifeLock Inc. (NLOK). These 7 tech companies are in the "Software" Industry of the " Technology" Sector.

The Fama French 3-Factor Data

Please download the file from

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Select `txt` next to "Fama/French 3 Factors [Weekly]" to download the text file. You will need to remove the data description on the top and copy right statement at the bottom of the file before the data can be read into R. Trim the data to be the same time period as `Rt`, see page 185 of Handout 9. These data are also in percentage format. You can find the details of these data in this website.

The R Functions

The file `FM_Functions.R` contain several R functions for making plots that required for this homework. You will need to source the file:

```
> source("/path_you_save_the_file/FM_functions.R")
```

The file gives brief instructions about how to use these functions. Please Read them. It will be convenient to "attach" the Fama French 3 factor data with the function `attach()`.

Homework Questions

All the models you are going to fit do NOT include industry/sector as a factor. When you give your answers, please always take this factor into account if applicable and explain any pattern or difference across industries.

1. Fit the Fama-French 3 factor model to the 26 asset daily return data.

   (a) Summarize the R Squared values and betas of the 26 assets by plotting them with the R function `coef.plot()` from `FM_Functions.R`.

   ```
   > coef.plot(R.Sq, betas)
   ```

   The first argument required the vector of R Squared values, the second argument is simply `coef(fit)[-1,]`, where `fit` is the returned object of `lm()`.

   Describe what you see from the R-Squared and the betas of the excess market returns. How many of them are considered "aggressive" asset? Give the numbers for each industry and comment on your findings.

   (b) Identify those individual assets that do not follow the FF-3-factor model and their sectors.

   (c) Test if the FF-3-factor model holds for ALL the 26 assets. Use both the Wald and the LRT.

   (d) Test if FF-3-factor model holds for each industry by testing the assets from the same industry. You should have 4 tests. Please give both the Wald and the LRT tests.

   (e) In estimating the covariance matrix of the asset returns, the reduction in the number of parameters is substantial with the FF-3-factor model when $N = 26$. First, give the numbers of parameters of the sample covariance and the model based.

   The model based estimation relies on the assumption that the covariance ε_t is diagonal. Check the assumption by the `resid.summary()` function,

   ```
   > resid.summary(res)
   ```

   The argument `res` is simply the residuals from the regression `resid(fit)`. The function plots the heatmap of the residual covariance and the number of significant pairs. Carefully describe what you see from the graph and its implications.

2. Apply the factor analysis to the return data. For the sake of comparison, use 3 factors. Compute the factors, R-squared and residuals. The residual is calculated by the formula (9.13) for each t. For the data of n observations, the residual matrix can be obtained by

$$\underset{n\times N}{\hat{\mathbf{E}}} = \underset{n\times N}{\tilde{\mathbf{Z}}} - \underset{n\times 3}{\mathbf{F}}\,\underset{3\times N}{\hat{\mathbf{B}}}$$

   where $\tilde{\mathbf{Z}}$ is standardized then de-meaned excess returns. Please do not display the residuals and R Squared vector.

   (a) Examine the loading estimate $\hat{\mathbf{B}}$, can you find interpretation about these coefficients?

(b) Check the covariance structure of the errors with the `resid.summary()` function. Describe your findings.

3. Apply the principal component analysis to the return data. Please use $\tilde{\mathbf{Z}}$.

   (a) Plot the scree plot. If `pca` is your returned object from `prcomp()`, the plot can be gotten by

   ```
   > plot(pca)
   ```

   According to the plot how many principal components would you choose?

   (b) Approximate the factor model with $p = 3$ (only for the purpose of comparison, this is not necessary the best choice). Compute the estimate factors and R Squared. Please do not display them.

   (c) Examine the wights of principal components, these are the estimates of loadings, can you find interpretations form them?

   (d) Check the covariance structure of the errors with the `resid.summary()` function. Describe your findings.

4. (a) Plot the R Squared values of the 3 models using the `RSq.plot()` function from `FM_functions.R`,

   ```
   > RSq.plot(RSq.all)
   ```

   There is only one argument which is a $N \times 3$ matrix of all R squared values from 3 models. You can use `cbind()` to combine the three vectors. Please use the default ordering, (1) Fama-French model; (2) factor analysis; then (3) principal components.

   (b) Compare and comment the three approaches in pricing modeling from all the analysis you have done.