Executive Summary:

The data used for analysis is our top priority in this business problem is to identify companies in bankruptcy. And this data were collected from the Taiwan Economic Journal for the years 1999 to 2009. Company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange. We trying to predict company Bankruptcy and as well as classify the features that lead to bankruptcy for companies. Our loading that will be useful in our modelling and feature selection or variable be reduction techniques.

Our data consists of 6819 companies and 96 variables . All data were extracted from formal financial statements including balance sheets, cash flow statements and income statements taken from the TSEC financial databases, which imply that the findings of this research can be generalized to firms outside of Taiwan. Moreover, the proposed methodology and experimental results could be of use to other stock markets worldwide. Our goal is to train model that will predict with high accuracy the companies that went or did not go bankrupt.

In order to analyze the dataset, preprocessing was done to be able to analyze useful information. Preprocessing is essentially cleaning the data in order to make it usable and reduce any problems with analysis. We used multiple methods in order to analyze the dataset. The methods we used were Principal component analysis (PCA), …………………

We used PCA to group variables into components that would most greatly affect a bankrupt . Since the dataset had many variables, this helped group up different component of given variable and attributes to help condense the model. after grouping up a variables we come up with 3 component of with PCA they are Earnings Per Share, Inflation and Interest rate expectations and third is Return on Assets.

First test applied on the selected dataset is the **Cronbach's alpha** it tests the internal consistency of the data; in other words, it tests how closely the related the set of variables to each other. Secondly other test applied **is Kaiser-Meyer-Olkin (KMO)** Test for verifying that the selected data is suitable for factor analysis the minimum should be 0.6. Third and last test to assess the assumption is **Bartlett's Test of Sphericity** which verifies the variables possesses correlations or not. Selected data for factor analysis give p< 2.22e-16 which means it rejects the null hypothesis (no correlations present among the variables) and accepts the alternat hypothesis (correlation is present among the variables).

There is one limitations for dataset which has analysis is inherent strict assumption such as linearity , normality ,independence among predictor variables and per- existing functional forms relating to the criterion variable and the predictor variable. And other limitations suggest that future research could validate its applicability in other markets or select other explanatory variables in order to improve the predictive ability.

From our research we can conclude that a company bankrupt. is highest impacted by 3 component which we got by using tow types of analysis. The 3 component are named as Earnings Per Share, Inflation and Interest rate expectations and third is Return on Assets. And These 3 component are resulted from the PCA and CFA. as Earnings Per Share is the largest component that explains a company bankrupt. Future research might need to be done to uncover such an effect

Technical Summary:

Abstract:

In this paper we determine a Company bankruptcy was defined based on the business regulations of the Taiwan Stock Exchange. For the analysis we used data Company bankruptcy containing the 96 variables for analysis our dataset and 6819 observations in other words 6819 company are records for analysis bankruptcy. Firstly, **Cronbach's alpha** it tests the internal consistency of the data; then we test applied **is Kaiser-Meyer-Olkin (KMO)** and then we test to assess the assumption is **Bartlett's Test of Sphericity.**  To add to it Principal component analysis(PCA) and For model validation we used training and testing sets. We used this test because it shows how well the model predicts new data as well as making sure our model is usable

**Methods**

**Principal Component Analysis :** Principal component analysis (PCA) was conducted using the prcomp function, and principal function from the Psych package in RStudio. For the latter function, varimax rotation was used. The correct number of components were determined by considering both the eigenvalue and scree methods.

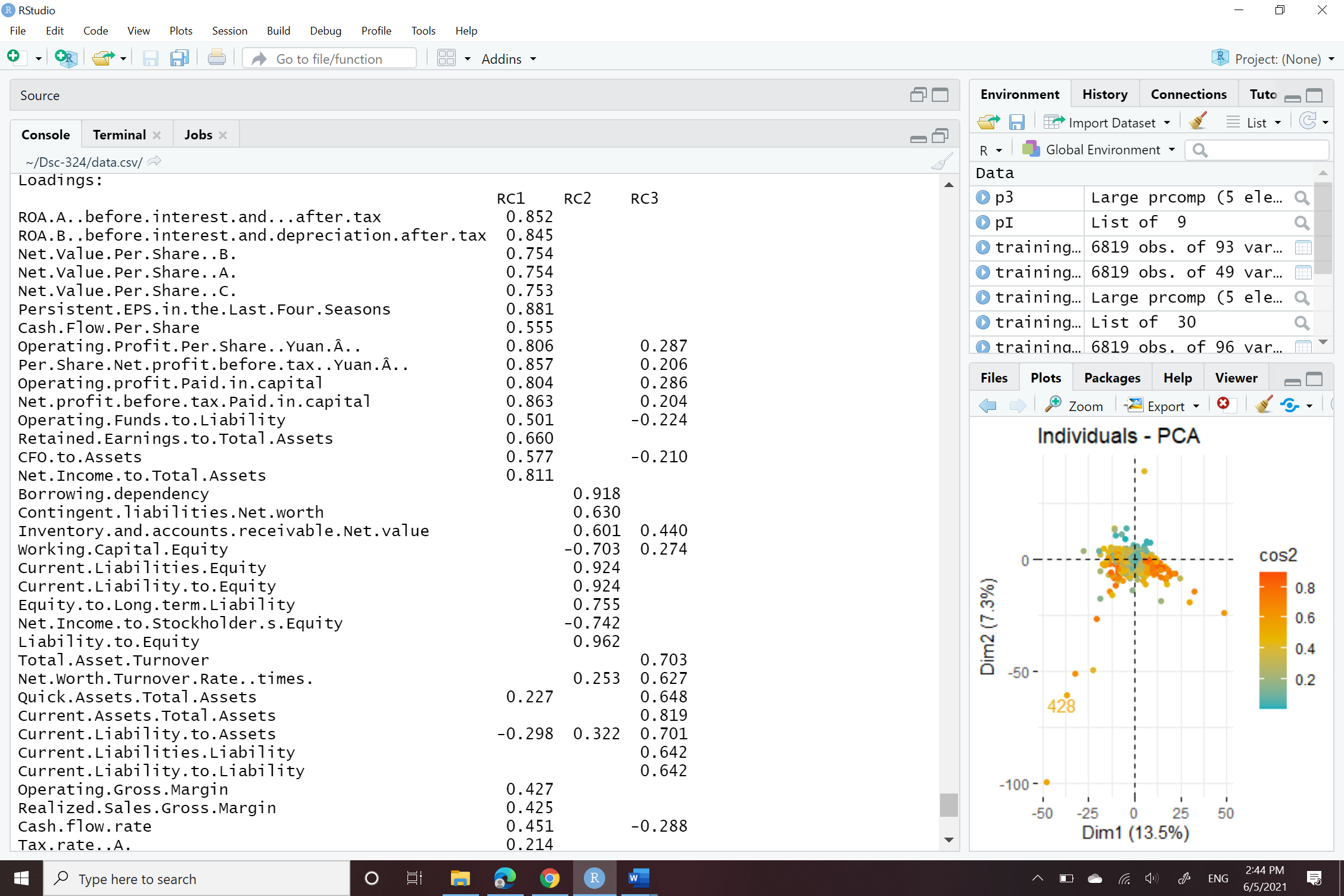
Principal Component Analysis The original dataset contained 96 variables and 6819 observations. After cleaning and preprocessing, the resulting dataset contained 93 variables and 6819 observations. The dependent variables related to bankrupt were removed from the dataset as well, leaving 2 total variables. Before starting the PCA process, the following techniques were used to verify assumptions about the dataset: Chronach’s Alpha, Kaiser-Meyer-Olkin (KMO), and Bartlett’s Test of Sphericity. Cronbach's Alpha assesses the consistency of each summated scale in the dataset. A value of 0.02 was recorded, suggesting good reliability for the dataset. Next, KMO was checked, recording a value of 0.50, indicating a relatively strong relationship and durable sample size. Lastly, Bartlett’s Sphericity Test was checked to evaluate shared variance. A value of p < 2.22e-16 was recorded. The null hypothesis that there are no correlations present in the dataset was rejected, accepting the alternative hypothesis that there are correlations, shared variance between the variables. In order to choose the number of components for analysis to best represent the variance in the dataset, both the eigenvalue and scree methods were considered. The scree method **Figure -1** in contrast, suggested all of 10 components. To test the components suggested by the eigenvalue method, 3 components were used, using varimax rotation and a final cutoff value of 0.2, in three separate analyses.. Using 4 components, the fourth component was questionable from an interpretability standpoint, as the variables did not give a clear picture of what the component meant in terms of application. Lastly, analysis was conducted and evaluated using 3 components. The main difference here, compared to the analyses with 3 and 4 components, was that all of the components were easily interpretable. Components 3 and 4, respectively, contained variables that made sense together, in terms of the application. Thus, 3 components were determined to be ideal for analysis. After configuring the PCA model with 3 components, the results are shown in **Table ( ).** Using 3 components, 46.8% of the variance was accounted for overall. Components 1 and 2, respectively, made up 21.1% and 36.5 % of the variance. Components 2 and 3 added 14.3% and 10.3%, Proportion Var respectively. In additional information about the variance of the independent variables in the dataset that was relevant to the application.

As we see in the give **Figure -1()** that component 1 the minimum score is -91.0578and the maximum score is 170.6308.

As we see in the give **Figure -1()** for component 2 the minimum score is -110.7711and the maximum score is 337.5442.

Then we see in the give **Figure -1()** for component 3 the minimum score is -21.1098 and the maximum score is 44.8143.

After consulting the loadings **Figure -()** , Component 1 was identified as being as net value.per share and also persistent Eps in the last four seasons, it was cash flow per Share and operating profit per share yuan assets therefore Component 1 define as **Earnings Per Share.** Component 2 was identified as being strongly borrowing dependency , current liabilities of equity and net income to stockholder share equity that why it labelled it **Inflation and Interest rate expectations.** Lastly, component 3 was correlated with net worth Assets , fixed assets turnover frequency and cash total Assets that why it labelled **Return on Assets.**



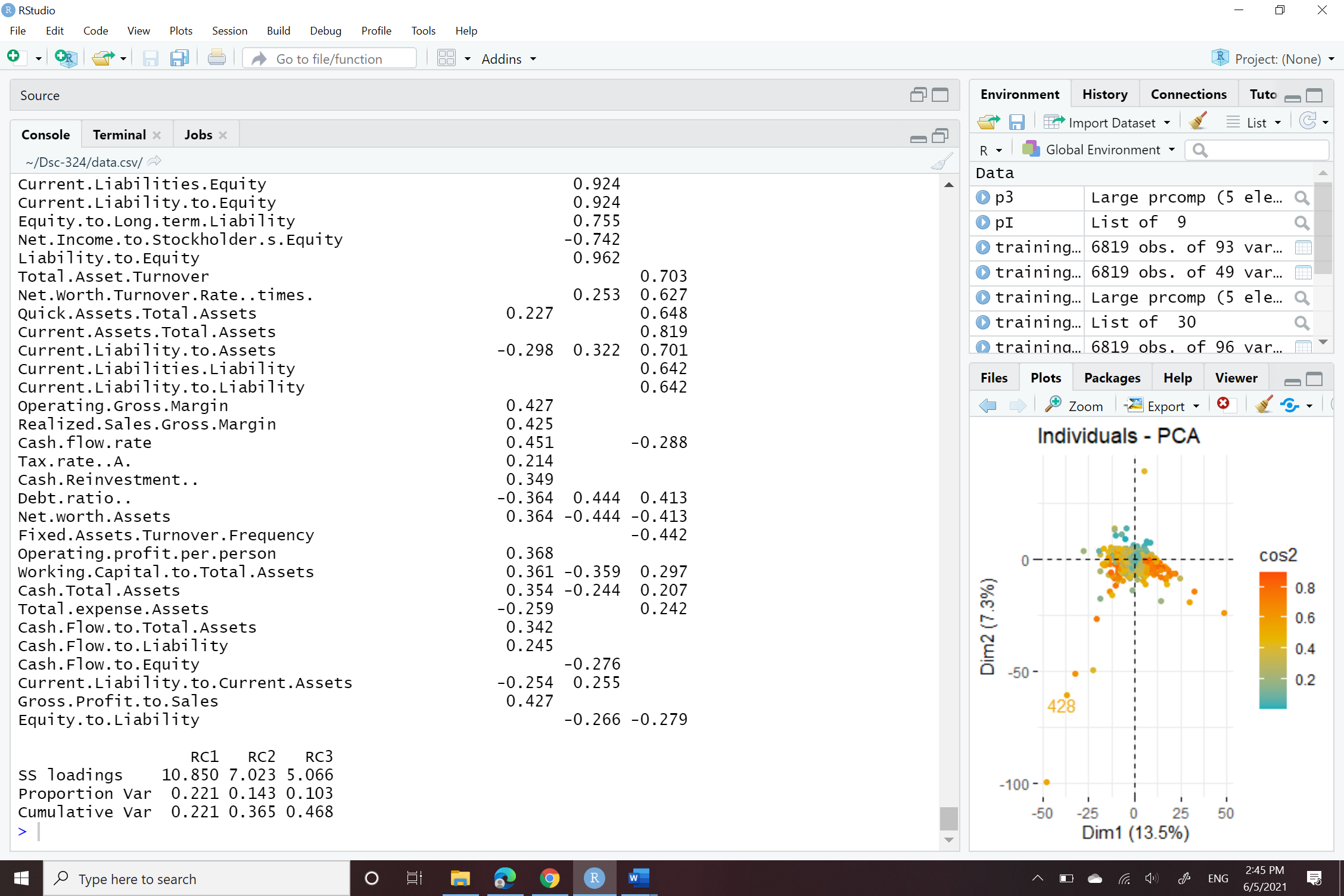


Figure -1

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Figure -2

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Earnings Per Share** | **Inflation and Interest rate expectations** | **Return on Assets** |
| **loadings** | 10.85 | 7.02 | 5.06 |
| **Proportion Var** | 22.1% | 14.3% | 10.3% |
| **Cumulative Var** | 22.1% | 36.5% | 46.8% |

Table -1

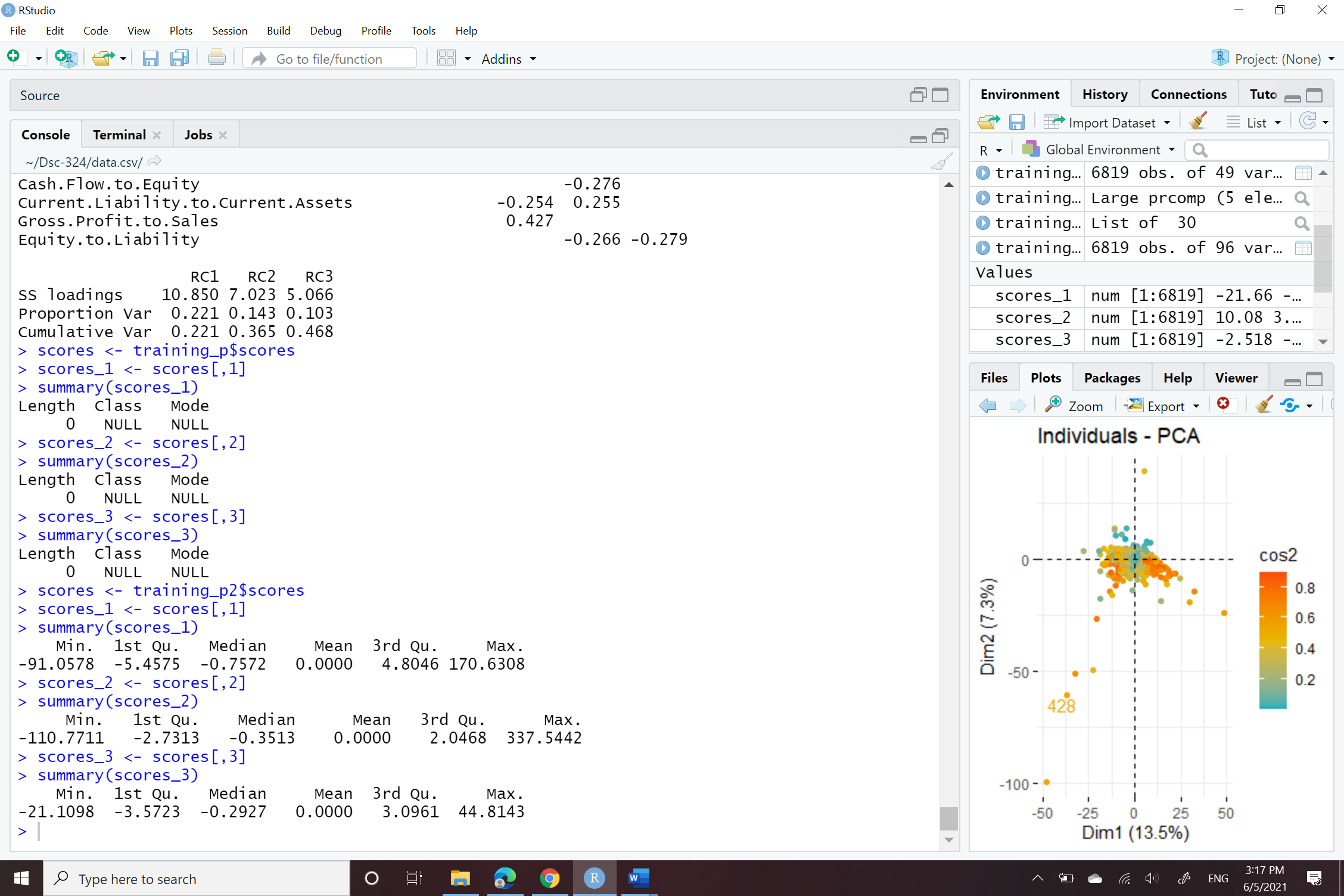


Figure -1