**Data Examination**

**Data Description:** Employment in various industry segments reported as a percent for thirty European nations. Note that EU stands for European Union, EFTA stands for European Free Trade Association, and Eastern stand for Eastern European nations or the former Eastern Block.

For convenience here are the definitions of the abbreviated industries.

AGR: agriculture

MIN: mining

MAN: manufacturing

PS: power and water supply

CON: construction

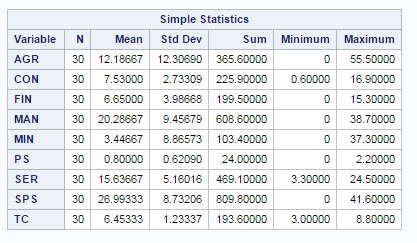
SER: services

FIN: finance

SPS: social and personal services

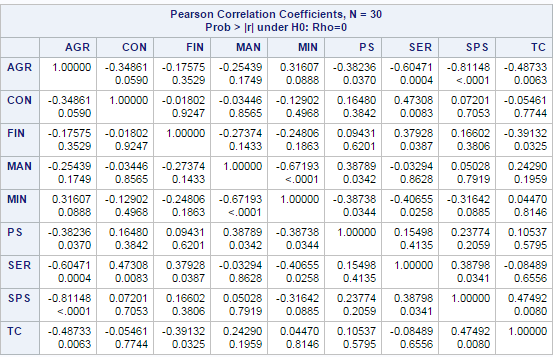
TC: transport and communications

Let’s have a look at some basic statistcs.

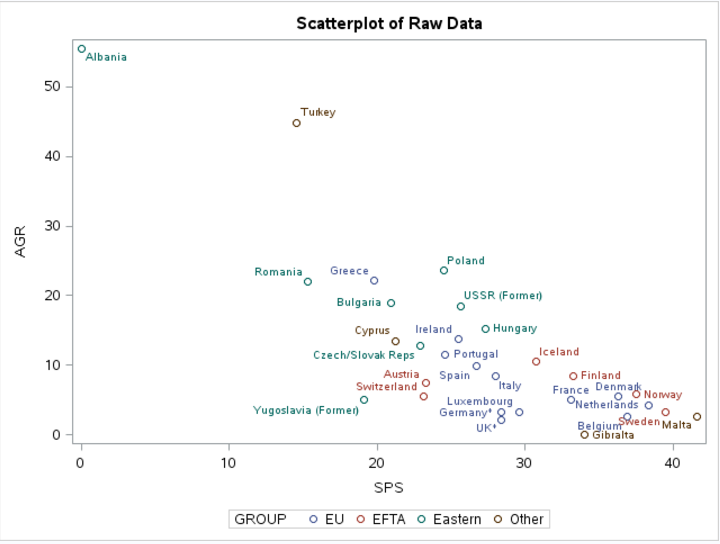


The strongest correlation, with a statistically significant test, was between AGR and SPS. The correlation being 0*.*81148 with a probability *>* |*r*| under

*H*0 : *\_* = 0 test statistic of *<* 0*.*0001.

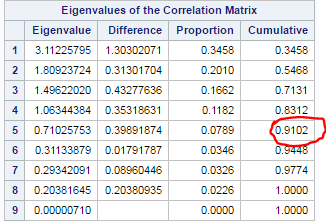


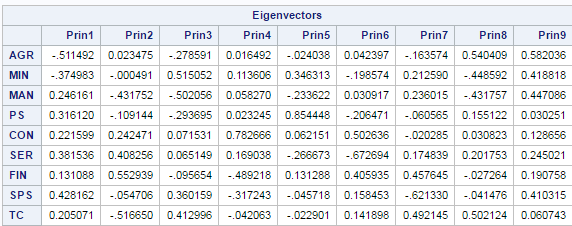
We’ll continue to examine this correlation by producing a scatter-plot:

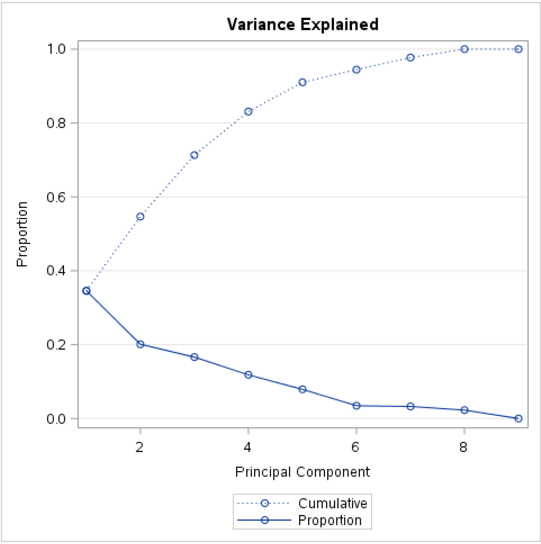


**Principal Components, Dimensionality Reduction**

As we have nine variables within the data set, we’ll use PCA as a dimensionality reduction method. We’lluse the variability table, or scree plot to examine how many components we require to account for 90% of the data variability.







**Cluster Analysis**