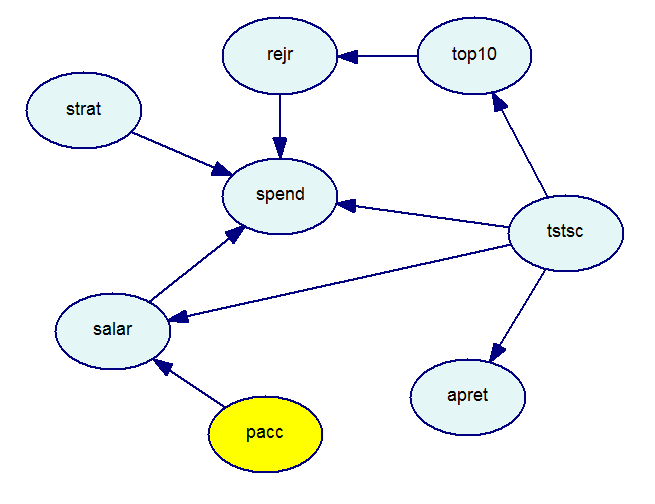
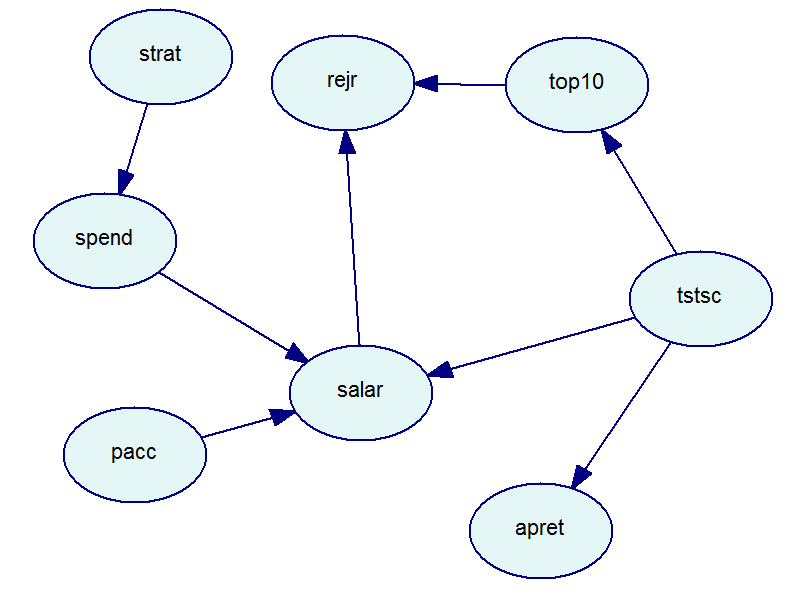
Team: Venkatesh Duvvuri, Haifa Alnasser, Gopi Tata

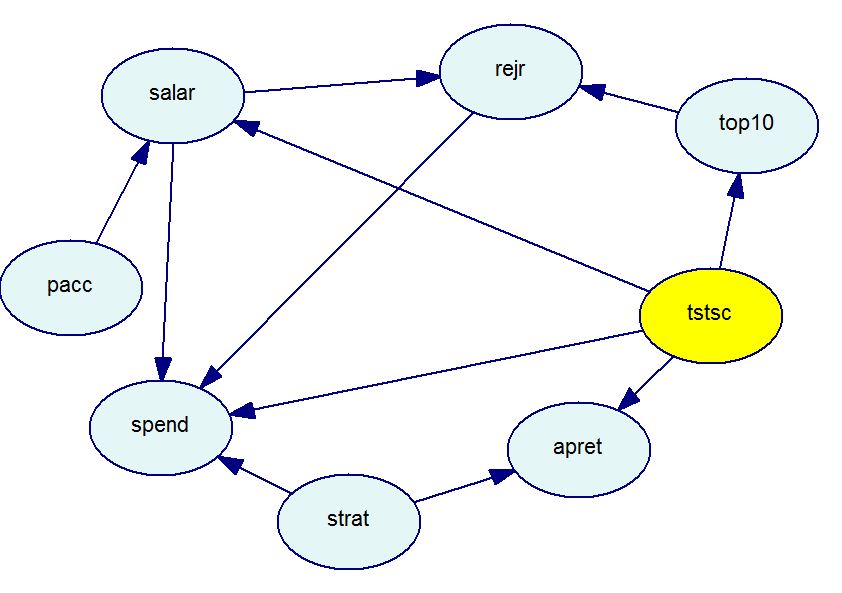
Significance Level = 0.01



Significance Level = 0.001



Significance Level = 0.05



**Findings:**

1. Do the 1993 data support Drudzel and Glymour’s conclusions?

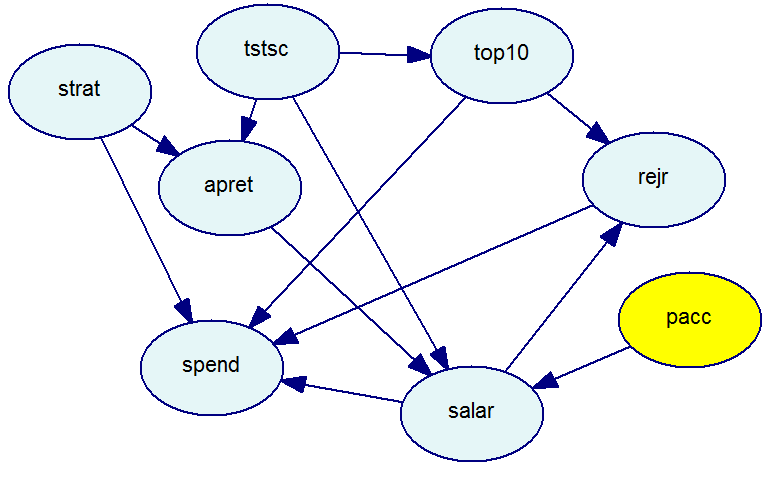
The findings from 1993 data using GeNIe matches that of 1992 data using TETRAD II.

From all the above three Directed Acyclic Graphs (DAG) for three different Significance levels 0.01,0.001 and 0.05 we strongly believe that the Average Percent of Retention (*apret*) is directly related to the Test Scores (*tstsc*)and students standing in the class (top10). The reason is because as per Markov’s condition of implications the *apret* is probabilistically dependent as there exists direct active path from *apret* to *tstsc* and *tstsc* to top10.

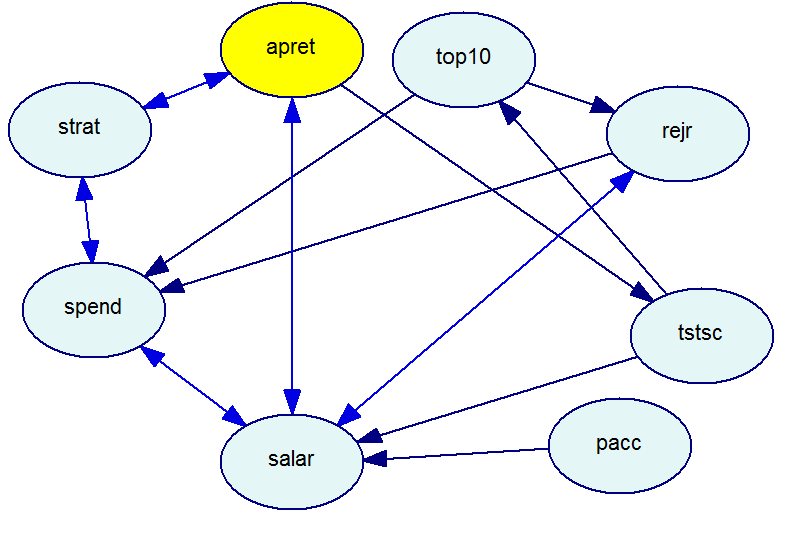
1. Can you find anything else going on in the data?

Observer the below plots for higher significance values:

Significance Level = 0.1



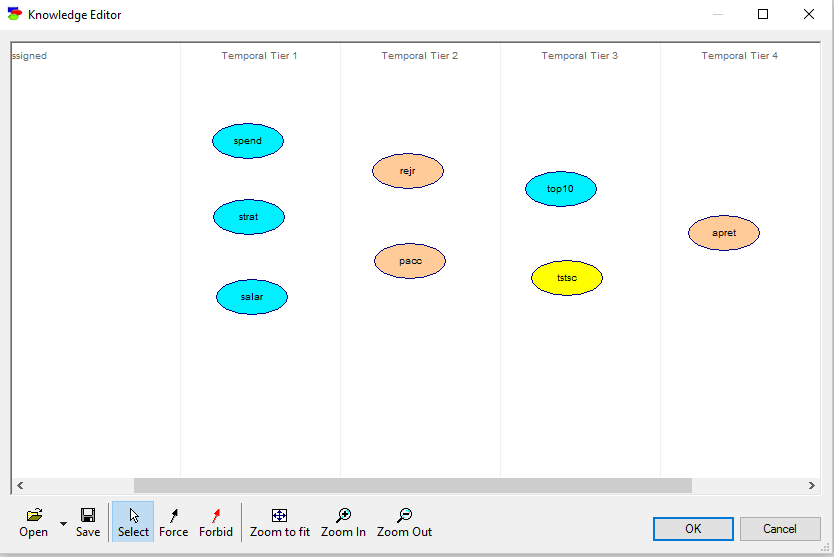
Significance Level = 0.2



* + 1. From all the DAGs it is evident that there is no Arc between *apret* and *pacc*. That is these two are not directly related.
    2. From all the three DAG it is also evident that the *rejr* and *apret* are probabilistically dependent.
    3. The *salar* and *strat* are directly related; the lower the *strat* the lower the *salar*

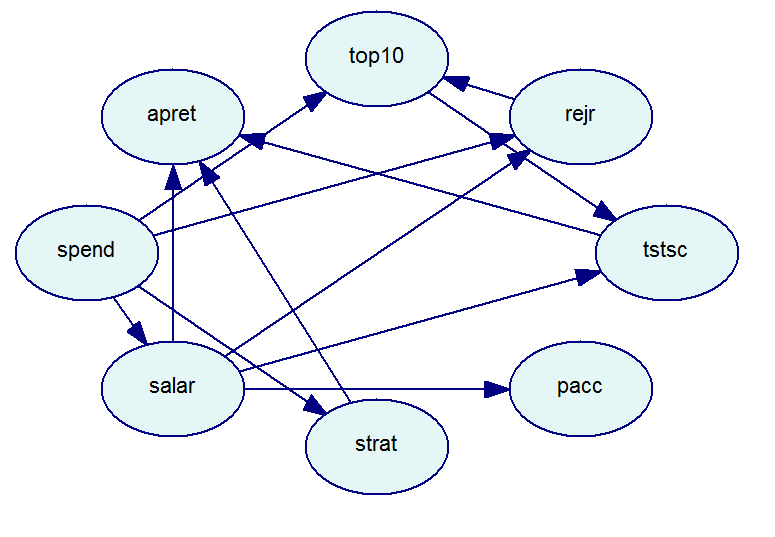
1. What are the causal graphs suggested by GeNIe? What causes student retention? Please, feel free to help GeNIe using your common sense knowledge of interactions between the variables included in the data set (not necessarily the knowledge inputed by Druzdzel & Glymour).

If we edit provide prior knowledge using the knowledge editor as in the figure below :

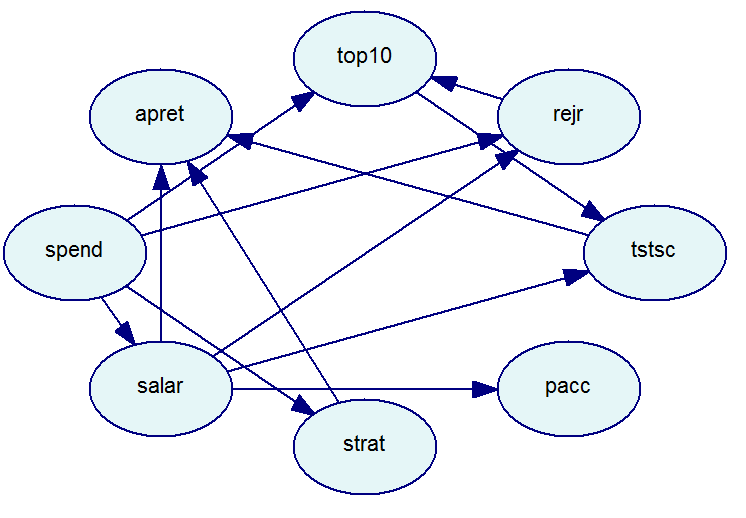


We can DAGs for different Significance levels :

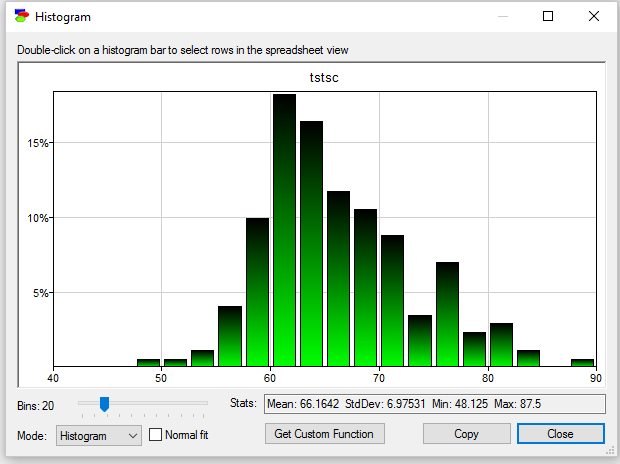
Significance Level = 0.2

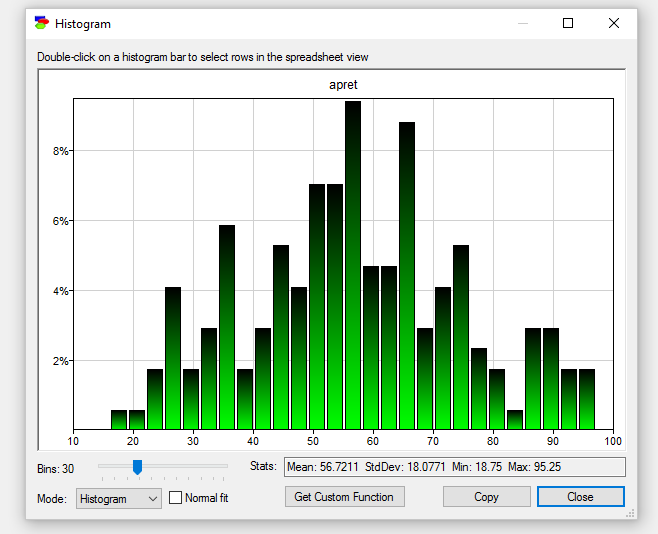


Significance Level = 0.1



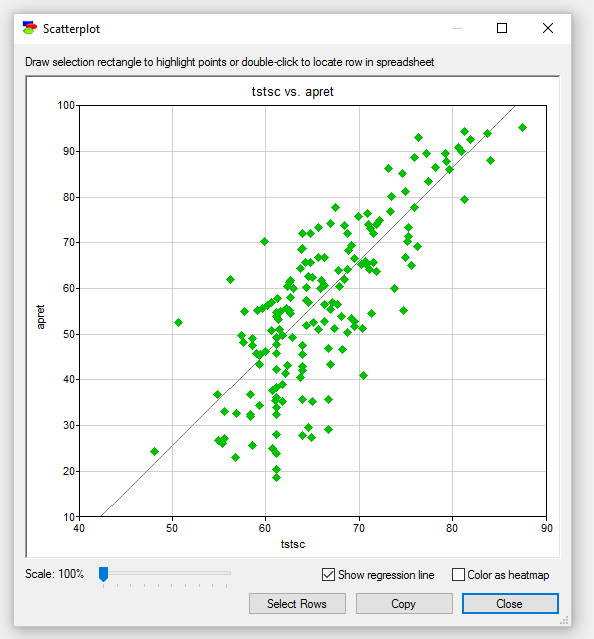
* + 1. For every Significance level we can observe that top10 and tstsc cause apret



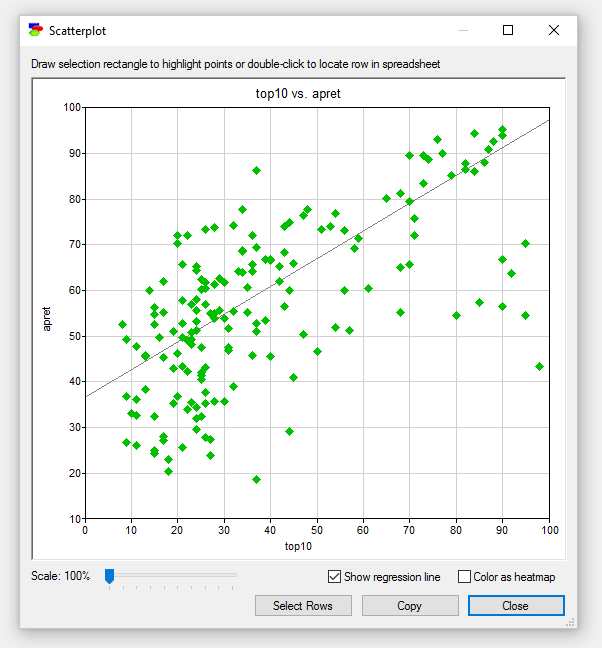


Druzdzel & Glymour made the assumption that the data are normally distributed and linearly dependent. Let us plot some and check ourselves

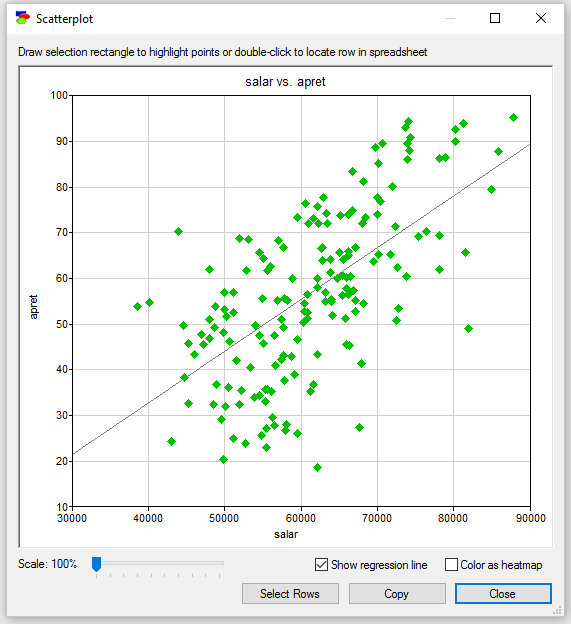
From the below Scatter plot between *tstsc* and *apret* these two variable are positively correlated.



From the Scatter plot blow between *top10* and *apret* these two variables are positively correlated.



From the Scatter Plot below the *salar* and *apret* are linearly related.



From above plots we can clearly infer that variables are linearly related.