

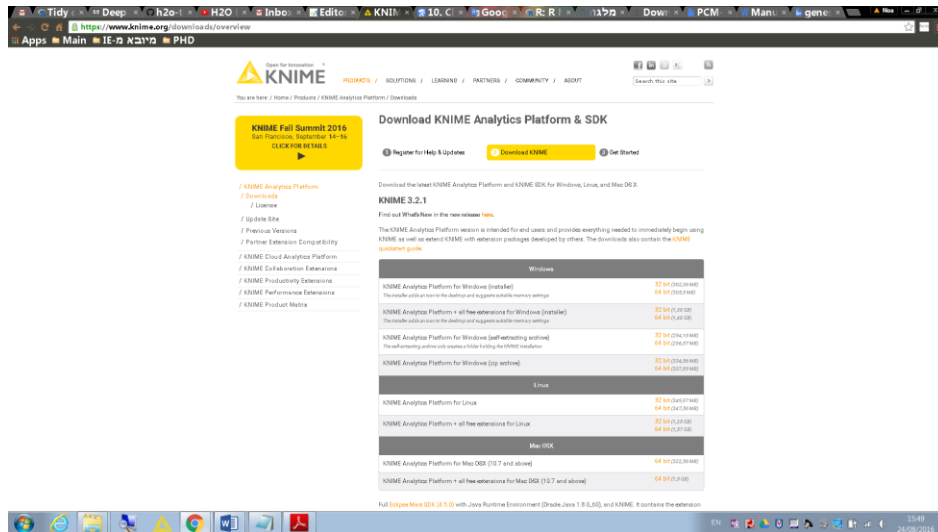
PCM-SABRE Manual

Basic running with example data

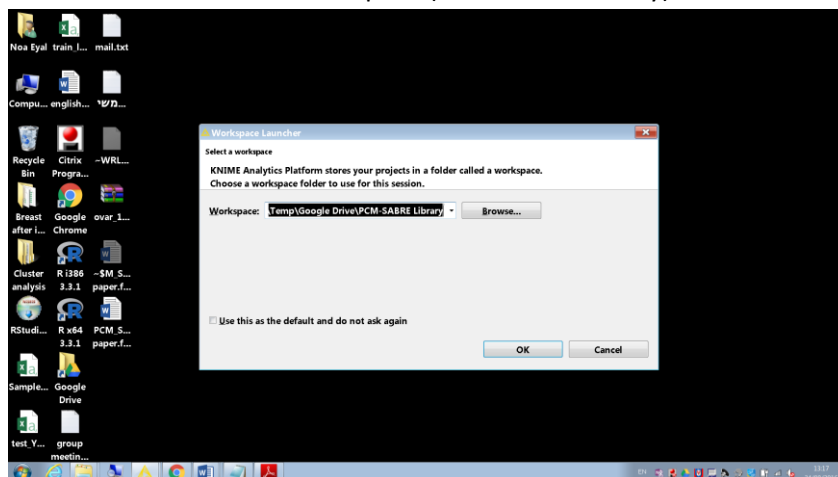
1. If R is not installed on your computer, please download it from <https://www.r-project.org/> and install.
2. Download KNIME Analytics Platform & SDK from the following link: <https://www.knime.org/downloads/overview>

please download the "KNIME Analytics Platform + all free extensions"

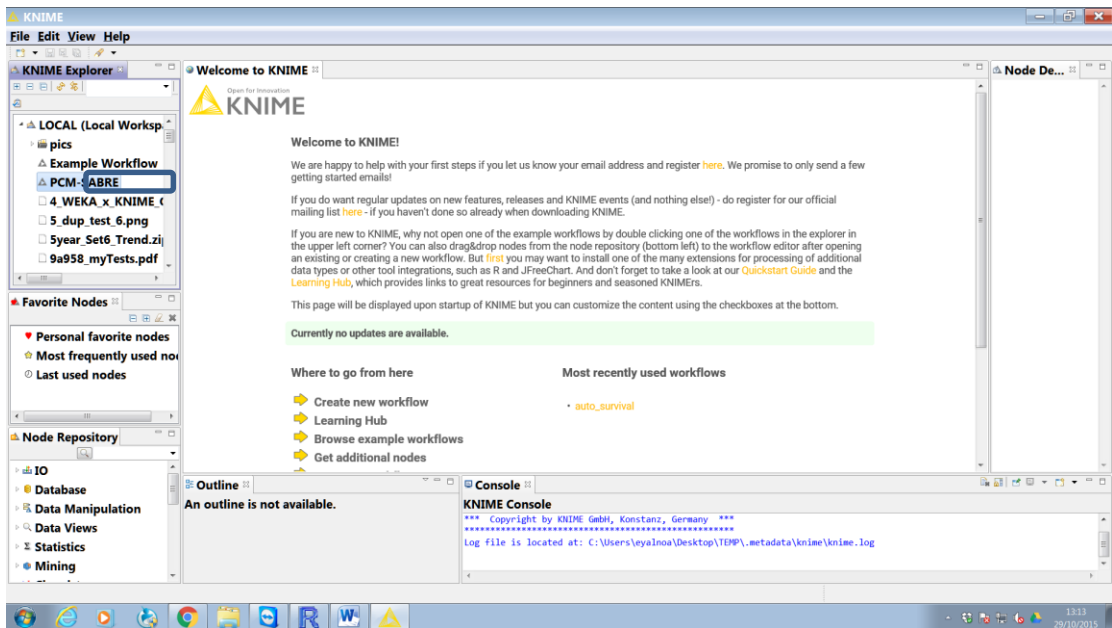
*** PCM-SABRE was built under KNIME version 3.1.0



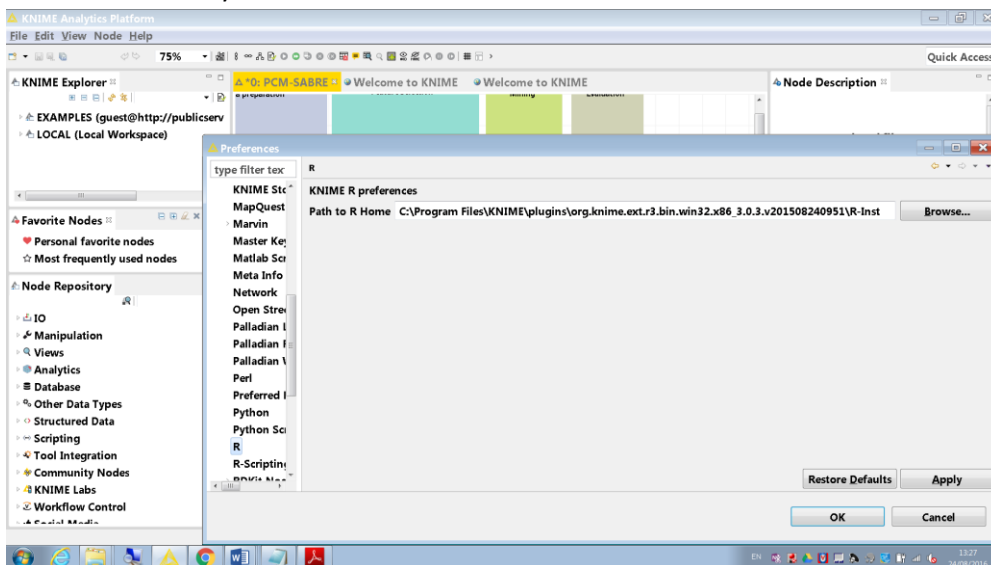
3. Install KNIME on your computer
4. If not yet downloaded, download PCM-SABRE library zip file from the following link: <https://drive.google.com/open?id=0B9pANNI-7eDdRmdoTGdvNUJFVIE>
5. Unzip the file
6. Open KNIME and select the workspace (PCM-SABRE library).



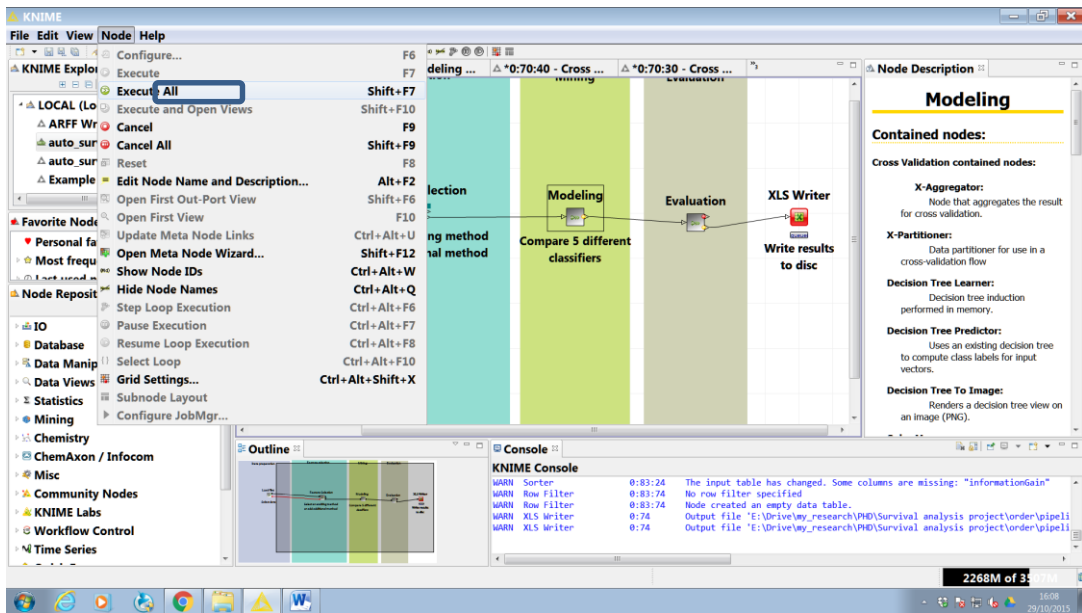
7. If the PCM-SABRE workflow do not open automatically, you can find it under the KNIME explorer box -> LOCAL -> PCM-SABRE (double click to open)



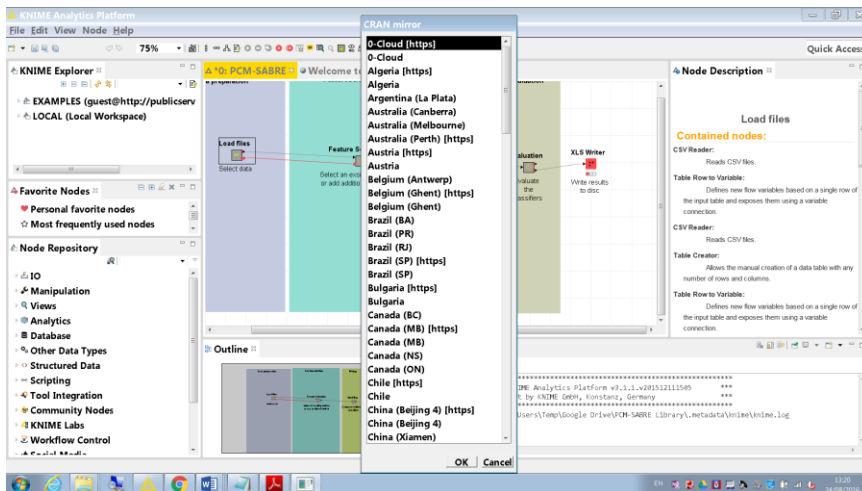
8. In order to change the default definitions, open the Flow_variables.csv file located at PCM-SABRE//csv files
 - Feature selection method:
 - 0 – information gain based feature selection (default)
 - 1 – ANOVA based feature selection
 - Number of chunks for columns split: continues variable, this variable control the number of loops for the information gain calculation in order to shorten the running time.
 - ER status and Lymph node status:
 - 0 – Negative
 - 1 – Positive
 - 2 – All patients
9. Before you run PCM-SABRE workflow, make sure that R home directory is correct. You can check it by File > Preferences > KNIME > R



10. To run the workflow, select on the menu: node -> execute all



11. Select CRAN mirror



12. You have two options to see the results:

- right-click on the "Evaluation" node -> connected to: joined table
- All the results files can be found under the PCM-SABRE\results folder

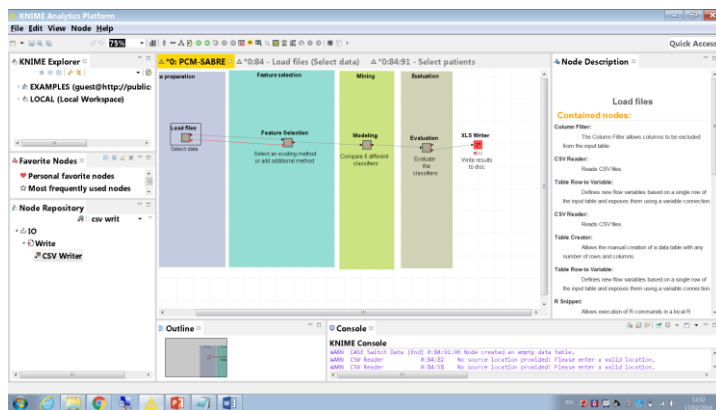
Running PCM-SABRE with different dataset

The default dataset supplied with the PCM-SABRE workflow was downloaded from the website <http://kmplot.com/analysis/>

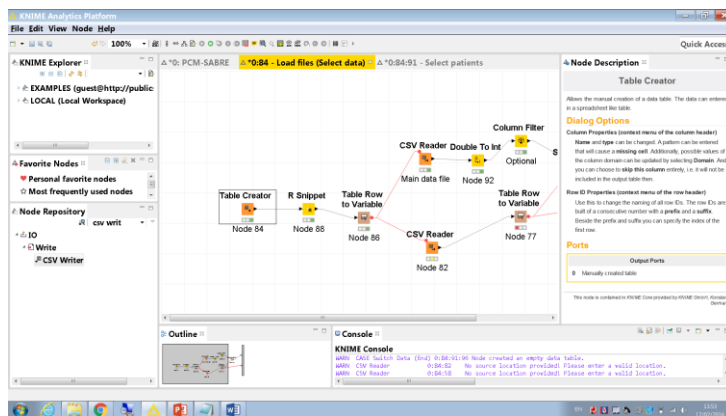
Four different versions of the dataset are supplied in the PCM-SABRE/csv files folder:

- abs_train_5.csv – The original dataset (1367 patients, 22,218 attributes)
- 5_train_without_missing.csv – The same as "1" but without missing values for the "ER status" and "Lymph node status" attributes.
- Sample.csv - The same as "2" contains only 1370 attributes for fast running of "PCM-SABRE" – default
- genes_5_train_without_missing3.csv – The same as "2" but the attributes were transformed from probe level to gene level. This file was used for the analysis.

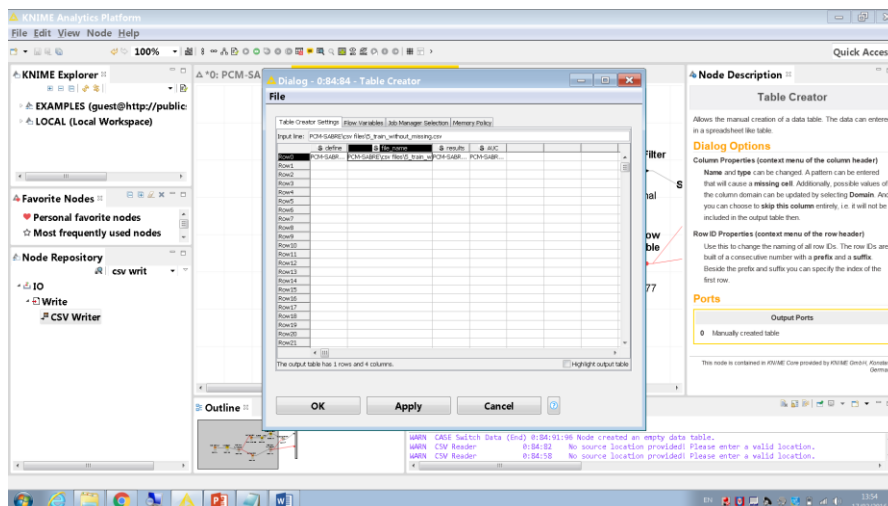
You can change the data file by double-click on the Loud files node



Double-click on the Table creator node.

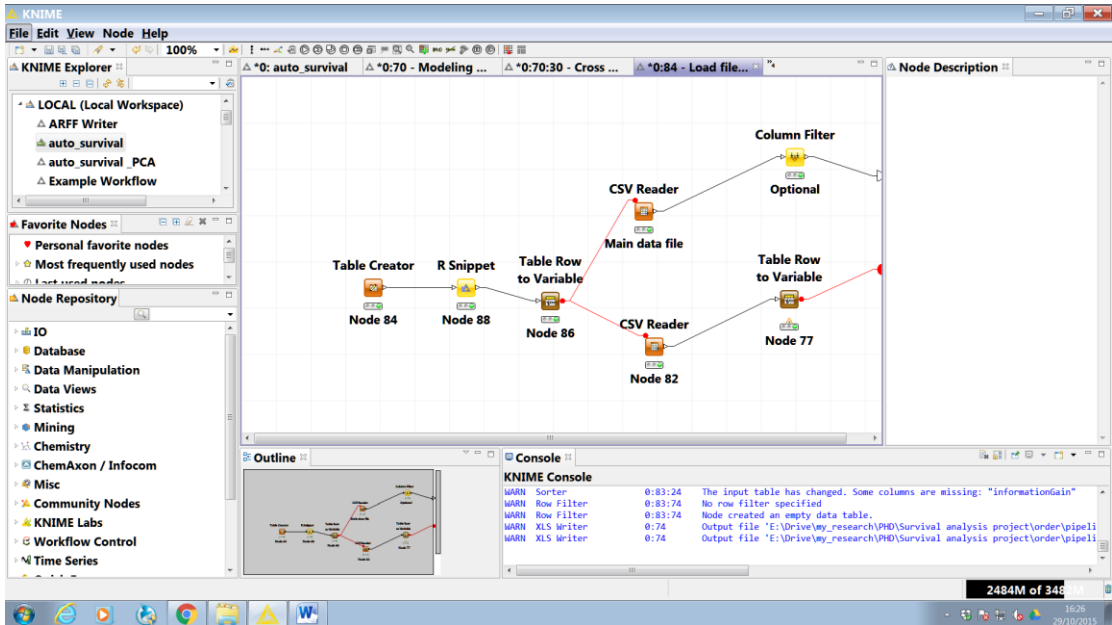


Under the "file name" column write: PCM-SABRE\csv files\ and after the "\" add the file name.

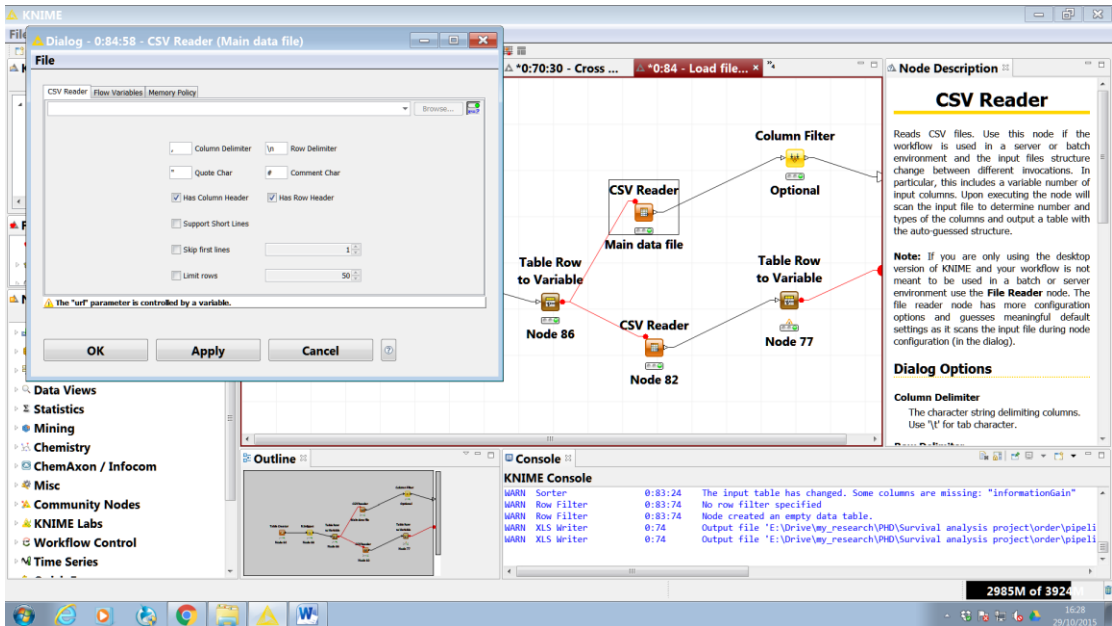


You can test your own data by:

1. double click on the CSV Reader (Main data file) node



2. Click on the flow variable button and uncheck the "use variable" option, click ok



3. Click on the browse button and select your data file, click ok

Make sure that your data files include the following instructions:

1. The prediction attribute name is "Prediction" and it is the last attribute
2. The ER status and Lymph node status attribute names are "ER status" and "Lymph node status" respectively.