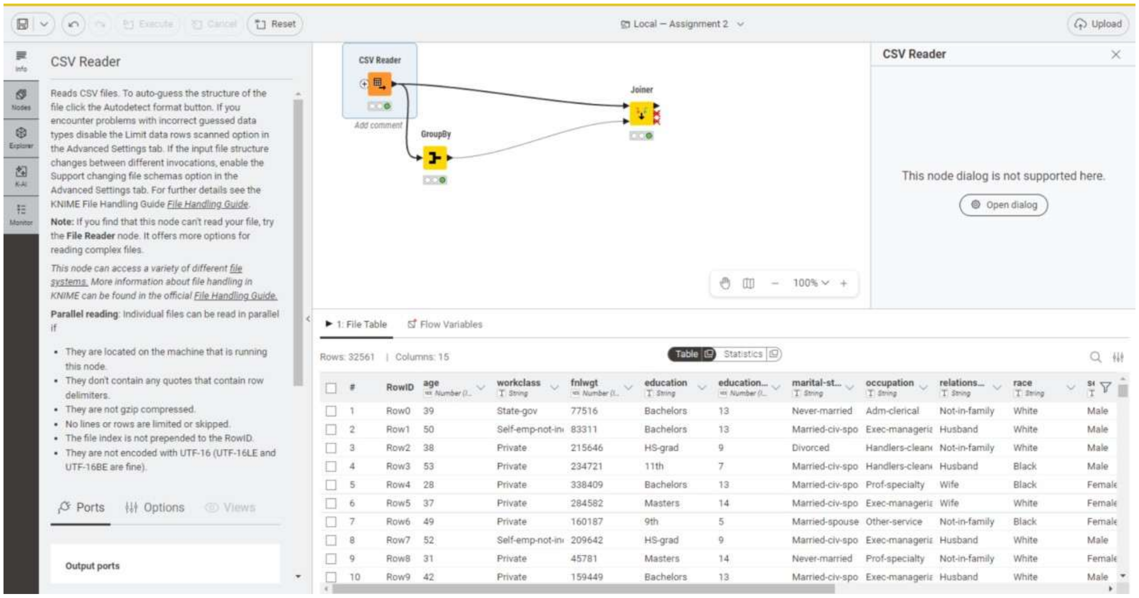
1) Read the adult.csv file available in the data folder on the KNIME Hub. The data are provided by the UCI Machine Learning Repository.

2) Calculate the average age and count for each one of the 4 groups defined by sex and income values

3) Join the two aggregated values to the original table

Step 1: Read the adult.csv file



CSV Reader

Reads CSV files. To auto-guess the structure of the file click the Autodetect format button. If you encounter problems with incorrect guessed data types disable the Limit data rows scanned option in the Advanced Settings tab. If the input file structure changes between different invocations, enable the Support changing file schemas option in the Advanced Settings tab. For further details see the KNIME File Handling Guide File Handling Guide.

Note: If you find that this node can't read your file, try the **File Reader** node. It offers more options for reading complex files.

This node can access a variety of different file systems. More information about file handling in KNIME can be found in the official File Handling Guide.

Parallel reading: Individual files can be read in parallel if

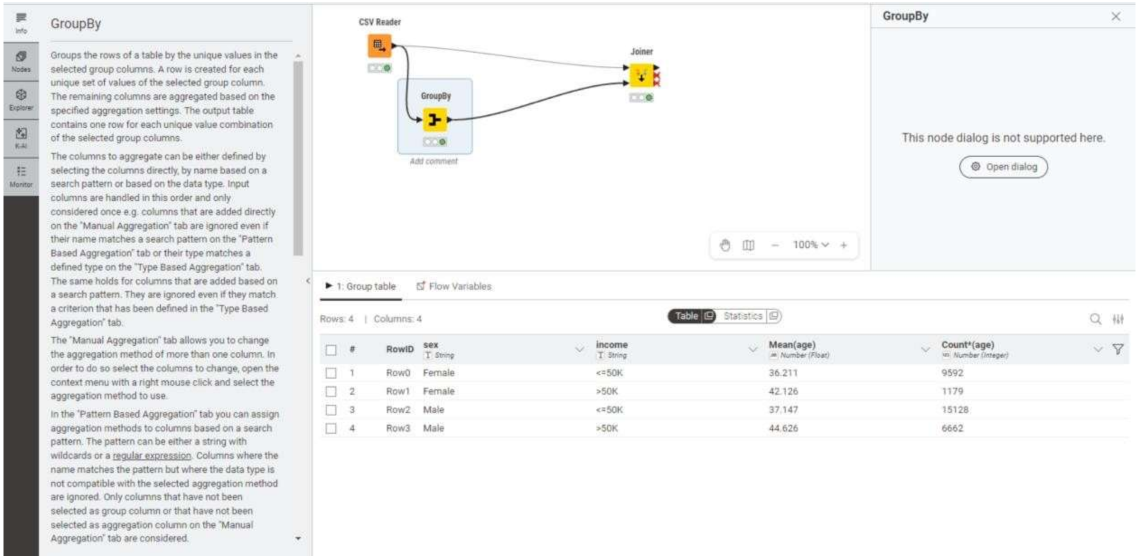
- They are located on the machine that is running this node.
- They don't contain any quotes that contain row delimiters.
- They are not gzip compressed.
- No lines or rows are limited or skipped.
- The file index is not prepended to the RowID.
- They are not encoded with UTF-16 (UTF-16LE and UTF-16BE are fine).

This node dialog is not supported here.

Rows: 32561 | Columns: 15

□	#	RowID	age Number (I...	workclass String	fnlwgt Number (I...	education String	education... Number (I...	marital-st... String	occupation String	relations... String	race String	s€
□	1	Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male
□	2	Row1	50	Self-emp-not-in	83311	Bachelors	13	Married-civ-spo	Exec-manageria	Husband	White	Male
□	3	Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-clean	Not-in-family	White	Male
□	4	Row3	53	Private	234721	11th	7	Married-civ-spo	Handlers-clean	Husband	Black	Male
□	5	Row4	28	Private	338409	Bachelors	13	Married-civ-spo	Prof-specialty	Wife	Black	Female
□	6	Row5	37	Private	284582	Masters	14	Married-civ-spo	Exec-manageria	Wife	White	Female
□	7	Row6	49	Private	160187	9th	5	Married-spouse	Other-service	Not-in-family	Black	Female
□	8	Row7	52	Self-emp-not-in	209642	HS-grad	9	Married-civ-spo	Exec-manageria	Husband	White	Male
□	9	Row8	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female
□	10	Row9	42	Private	159449	Bachelors	13	Married-civ-spo	Exec-manageria	Husband	White	Male

Step 2: Calculate the average age and count for each one of the 4 groups defined by sex and income values



GroupBy

Groups the rows of a table by the unique values in the selected group columns. A row is created for each unique set of values of the selected group column. The remaining columns are aggregated based on the specified aggregation settings. The output table contains one row for each unique value combination of the selected group columns.

The columns to aggregate can be either defined by selecting the columns directly, by name based on a search pattern or based on the data type. Input columns are handled in this order and only considered once e.g. columns that are added directly on the "Manual Aggregation" tab are ignored even if their name matches a search pattern on the "Pattern Based Aggregation" tab or their type matches a defined type on the "Type Based Aggregation" tab. The same holds for columns that are added based on a search pattern. They are ignored even if they match a criterion that has been defined in the "Type Based Aggregation" tab.

The "Manual Aggregation" tab allows you to change the aggregation method of more than one column. In order to do so select the columns to change, open the context menu with a right mouse click and select the aggregation method to use.

In the "Pattern Based Aggregation" tab you can assign aggregation methods to columns based on a search pattern. The pattern can be either a string with wildcards or a regular expression. Columns where the name matches the pattern but where the data type is not compatible with the selected aggregation method are ignored. Only columns that have not been selected as group column or that have not been selected as aggregation column on the "Manual Aggregation" tab are considered.

GroupBy

This node dialog is not supported here.

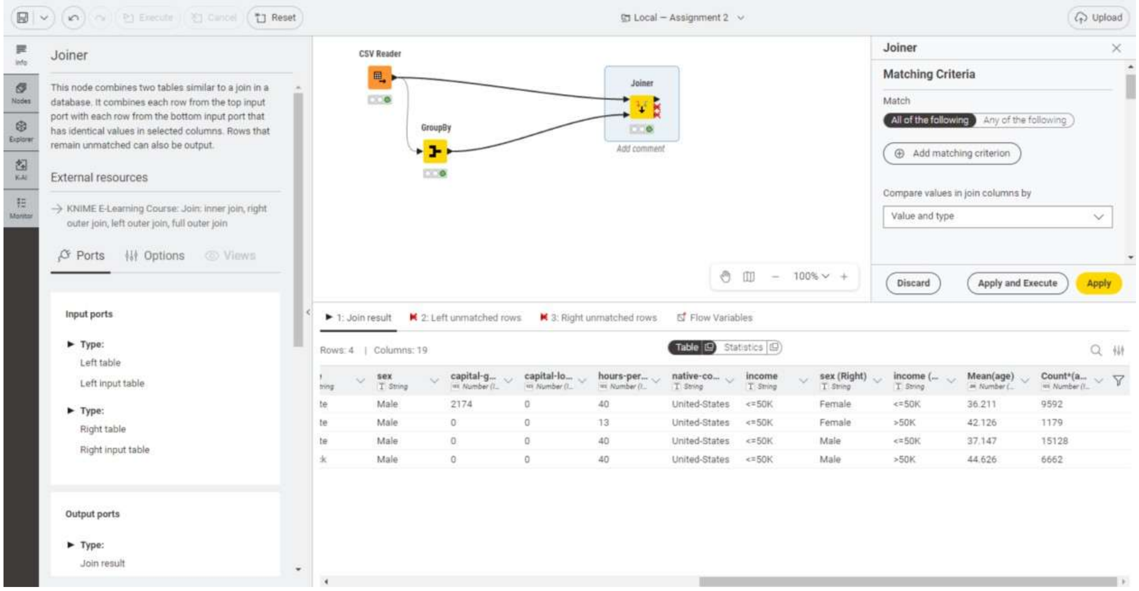
1: Group table | Flow Variables

Rows: 4 | Columns: 4

Table | Statistics

□	#	RowID	sex String	income String	Mean(age) Number (Float)	Count*(age) Number (Integer)
□	1	Row0	Female	<=50K	36.211	9592
□	2	Row1	Female	>50K	42.126	1179
□	3	Row2	Male	<=50K	37.147	15128
□	4	Row3	Male	>50K	44.626	6662

Step 3: Join the two aggregated values to the original value



Joiner

This node combines two tables similar to a join in a database. It combines each row from the top input port with each row from the bottom input port that has identical values in selected columns. Rows that remain unmatched can also be output.

External resources

→ KNIME E-Learning Course: Join: inner join, right outer join, left outer join, full outer join

Ports | Options | Views

Input ports

- Type:
  Left table
  Left input table
- Type:
  Right table
  Right input table

Output ports

- Type:
  Join result

Joiner

Matching Criteria

Match: All of the following | Any of the following

Add matching criterion

Compare values in join columns by: Value and type

1: Join result | 2: Left unmatched rows | 3: Right unmatched rows | Flow Variables

Rows: 4 | Columns: 19

Table | Statistics

… …tring	sex String	capital-g... Number (I...	capital-lo... Number (I...	hours-per... Number (I...	native-co... String	income String	sex (Right) String	income (... String	Mean(age) Number (...	Count*(a... Number (I...
te	Male	2174	0	40	United-States	<=50K	Female	<=50K	36.211	9592
te	Male	0	0	13	United-States	<=50K	Female	>50K	42.126	1179
te	Male	0	0	40	United-States	<=50K	Male	<=50K	37.147	15128
:k	Male	0	0	40	United-States	<=50K	Male	>50K	44.626	6662