

# INSURANCE PREMIUM ANALYSIS USING KNIME

Project Guide:-  
SAYAN CHAKBRABOTY

**Presented by**

**RIAJ IKBAL**

**33200121135**

**BTECH | CSE | 3RD  
YEAR**

**FROM  
TECHNO  
INTERNATIONAL  
BATANAGAR**

**1**

**2**

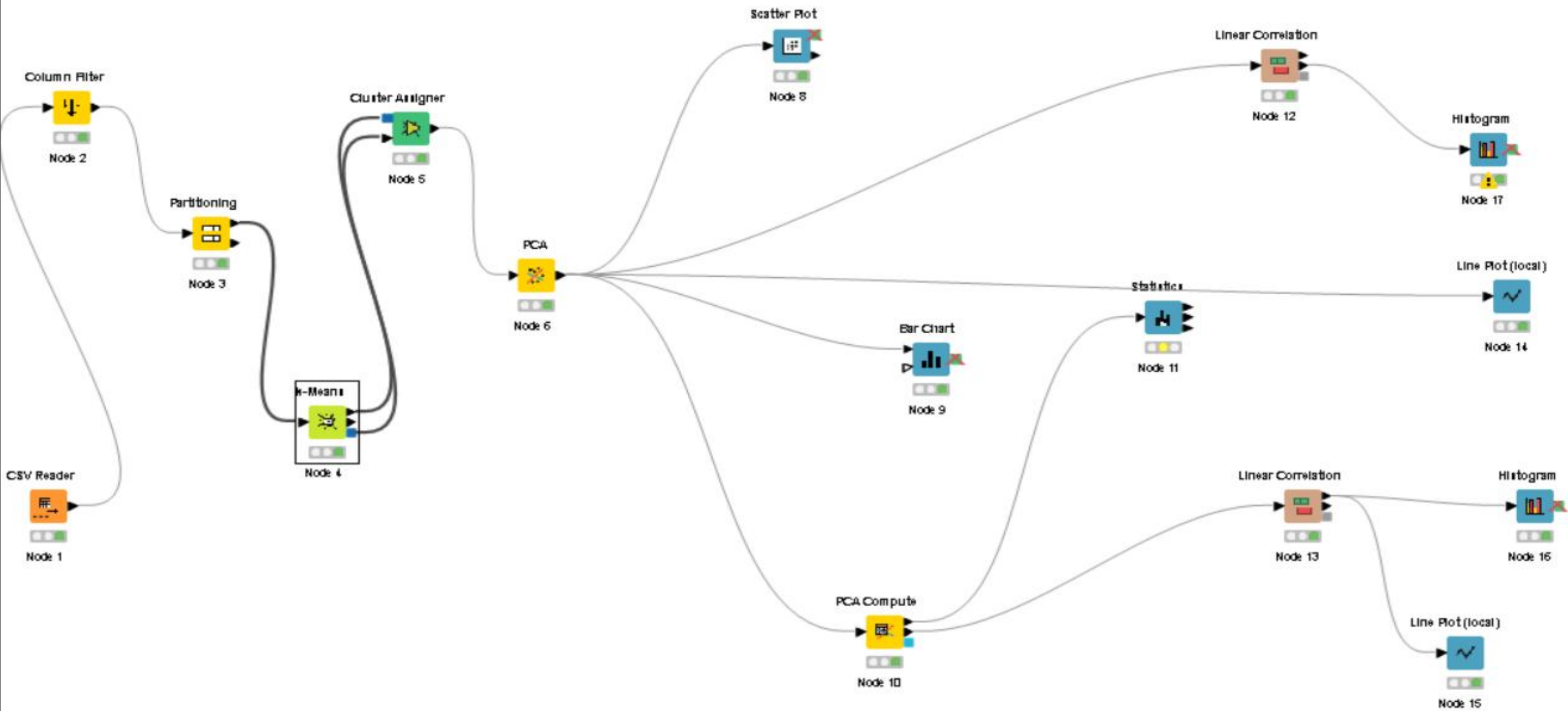
**3**

**4**

# motive



The loss ratio is constructed specifically to analyze the operation of an insurance company.



# At first we need to read the data set so here i used csv reader

The image shows a data science IDE interface. On the left, a workflow canvas contains a node labeled "Node 1" with an icon representing a CSV reader. A line connects this node to a large dialog box titled "Dialog - 3:1 - CSV Reader".

The dialog box has several tabs: "Settings" (selected), "Transformation", "Advanced Settings", "Limit Rows", "Encoding", "Flow Variables", "Job Manager Selection", and "Memory Policy".

**Input location**

- Read from: Local File System
- Mode: ☒ File ☐ Files in folder
- File: D:\MIOX\attendance system\insurance\_premiums.csv
- Browse...

**Reader options**

**Format**

- Autodetect format
- Column delimiter: ,
- Row delimiter: ☒ Line break ☐ Custom \r\n
- Quote char: "
- Quote escape char: "
- Comment char: #
- ☒ Has column header ☐ Has row ID
- ☐ Support short data rows ☐ Prepend file index to row ID

**Preview**

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	S State	D fatal_c...	I fatal_c...	I fatal_c...	I fatal_c...	I fatal_c...	D premiums	D insuran...	S Regi
Row0	Alabama	18.8	39	30	96	80	784.55	145.08	South
Row1	Alaska	18.1	41	25	90	94	1,053.48	133.93	West
Row2	Arizona	18.6	35	28	84	96	899.47	110.35	West
Row3	Arkansas	22.4	18	26	94	95	827.34	142.39	South
Row4	California	12	35	28	91	89	878.41	165.63	West
Row5	Colorado	13.6	37	28	79	95	835.5	139.91	West
Row6	Connecticut	10.8	46	36	87	82	1,068.73	167.02	Northeast
Row7	Delaware	16.2	38	30	87	99	1,137.87	151.48	South
Row8	District of C...	5.9	34	27	100	100	1,273.89	136.05	South
Row9	Florida	17.9	21	29	92	94	1,160.13	144.18	South
Row10	Georgia	15.6	19	25	95	93	913.15	142.8	South

Buttons at the bottom: OK, Apply, Cancel, ?

now here i have filtered the column accordingly and removed the unnecessary columns to the right side

The screenshot displays the KNIME Analytics Platform interface. A workflow is visible with two nodes: 'Node 2' (Column Filter) and 'Node 3' (Partitioning). The 'Column Filter' node is highlighted with a yellow box. A dialog box titled 'Dialog - 3:2 - Column Filter' is open, showing the configuration for the 'Column Filter' node. The dialog has tabs for 'Column Filter', 'Flow Variables', 'Job Manager Selection', and 'Memory Policy'. The 'Column Filter' tab is active, showing the 'Manual Selection' radio button selected. The 'Exclude' list contains the following columns: 'Fatal\_collisions', 'Fatal\_collisions\_speeding', 'Fatal\_collisions\_alc', 'Fatal\_collisions\_not\_distracted', and 'Fatal\_collisions\_no\_hist'. The 'Include' list contains the following columns: 'State', 'premiums', 'insurance\_losses', and 'Region'. The 'Enforce exclusion' radio button is selected. The dialog also has 'OK', 'Apply', and 'Cancel' buttons at the bottom.

Workflow components:

- Node 2:** Column Filter (Yellow node with a filter icon)
- Node 3:** Partitioning (Yellow node with a grid icon)

Dialog - 3:2 - Column Filter configuration:

- Tab:** Column Filter
- Selection Method:** Manual Selection (Selected)
- Exclude List:**
  - Fatal\_collisions
  - Fatal\_collisions\_speeding
  - Fatal\_collisions\_alc
  - Fatal\_collisions\_not\_distracted
  - Fatal\_collisions\_no\_hist
- Include List:**
  - State
  - premiums
  - insurance\_losses
  - Region
- Enforcement:** Enforce exclusion (Selected)

here's the filtered table

Filtered table - 3:2 - Column Filter

File Edit Hilite Navigation View

Table "default" - Rows: 51 Spec - Columns: 4 Properties Flow Variables

Row ID	S State	D premiums	D insuran...	S Region
Row0	Alabama	784.55	145.08	South
Row1	Alaska	1,053.48	133.93	West
Row2	Arizona	899.47	110.35	West
Row3	Arkansas	827.34	142.39	South
Row4	California	878.41	165.63	West
Row5	Colorado	835.5	139.91	West
Row6	Connecticut	1,068.73	167.02	Northeast
Row7	Delaware	1,137.87	151.48	South
Row8	District of C...	1,273.89	136.05	South
Row9	Florida	1,160.13	144.18	South
Row10	Georgia	913.15	142.8	South
Row11	Hawaii	861.18	120.92	West
Row12	Idaho	641.96	82.75	West
Row13	Illinois	803.11	139.15	Midwest
Row14	Indiana	710.46	108.92	Midwest
Row15	Iowa	649.06	114.47	Midwest
Row16	Kansas	780.45	133.8	Midwest
Row17	Kentucky	872.51	137.13	South
Row18	Louisiana	1,281.55	194.78	South
Row19	Maine	661.88	96.57	Northeast
Row20	Maryland	1,048.78	192.7	South
Row21	Massachusetts	1,011.14	135.63	Northeast
Row22	Michigan	1,110.61	152.26	Midwest
Row23	Minnesota	777.18	133.35	Midwest
Row24	Mississippi	896.07	155.77	South
Row25	Missouri	790.32	144.45	Midwest
Row26	Montana	816.21	85.15	West
Row27	Nebraska	732.28	114.82	Midwest
Row28	Nevada	1,029.87	138.71	West
Row29	New Hamps...	746.54	120.21	Northeast
Row30	New Jersey	1,301.52	159.85	Northeast
Row31	New Mexico	869.85	120.75	West
Row32	New York	1,234.31	150.01	Northeast
Row33	North Carolina	708.24	127.82	South
Row34	North Dakota	688.75	109.72	Midwest
Row35	Ohio	697.73	133.52	Midwest
Row36	Oklahoma	881.51	178.86	South
Row37	Oregon	804.71	104.61	West

3: insuranceWelcome to KNIME Analytics Platform

SV ReaderNode 1

Column FilterNode 2

PartitioningNode 3

Dialog - 3:3 - Partitioning

File

First partition

Flow Variables

Job Manager Selection

Memory Policy

Choose size of first partition

☐ Absolute

100

☒ Relative[%]

70

☐ Take from top

☐ Linear sampling

☒ Draw randomly

☐ Stratified sampling

☐ Use random seed

S Region

1,676,787,589,€

OK

Apply

Cancel

?

Linear CorrelationNode 12

StatisticsNode 11

Linear CorrelationNode 10

Node 9

Partitioning

The input table is split into two partitions (i.e. row-wise), e.g. train and test data. The two partitions are available at the two output ports. The following options are available in the dialog:

Dialog Options

Absolute

Specify the

Outline

Console

Node Monitor

KNIME Console



first partition

First partition (as defined in dialog) - 3:3 - Partitioning

Table "default" - Rows: 35 Spec - Columns: 4 Properties Flow Variables

Row ID	S State	D premiums	D insuran...	S Region
Row0	Alabama	784.55	145.08	South
Row3	Arkansas	827.34	142.39	South
Row5	Colorado	835.5	139.91	West
Row6	Connecticut	1,068.73	167.02	Northeast
Row7	Delaware	1,137.87	151.48	South
Row9	Florida	1,160.13	144.18	South
Row10	Georgia	913.15	142.8	South
Row11	Hawaii	861.18	120.92	West
Row14	Indiana	710.46	108.92	Midwest
Row15	Iowa	649.06	114.47	Midwest
Row16	Kansas	780.45	133.8	Midwest
Row19	Maine	661.88	96.57	Northeast
Row20	Maryland	1,048.78	192.7	South
Row22	Michigan	1,110.61	152.26	Midwest
Row23	Minnesota	777.18	133.35	Midwest
Row24	Mississippi	896.07	155.77	South
Row25	Missouri	790.32	144.45	Midwest
Row26	Montana	816.21	85.15	West
Row27	Nebraska	732.28	114.82	Midwest
Row29	New Hamps...	746.54	120.21	Northeast
Row30	New Jersey	1,301.52	159.85	Northeast
Row33	North Carolina	708.24	127.82	South
Row34	North Dakota	688.75	109.72	Midwest
Row35	Ohio	697.73	133.52	Midwest
Row36	Oklahoma	881.51	178.86	South
Row38	Pennsylvania	905.99	153.86	Northeast
Row39	Rhode Island	1,148.99	148.58	Northeast
Row41	South Dakota	669.31	96.87	Midwest
Row42	Tennessee	767.91	155.57	South
Row44	Utah	809.38	109.48	West
Row45	Vermont	716.2	109.61	Northeast
Row46	Virginia	768.95	153.72	South
Row47	Washington	890.03	111.62	West
Row49	Wisconsin	670.31	106.62	Midwest
Row50	Wyoming	791.14	122.04	West

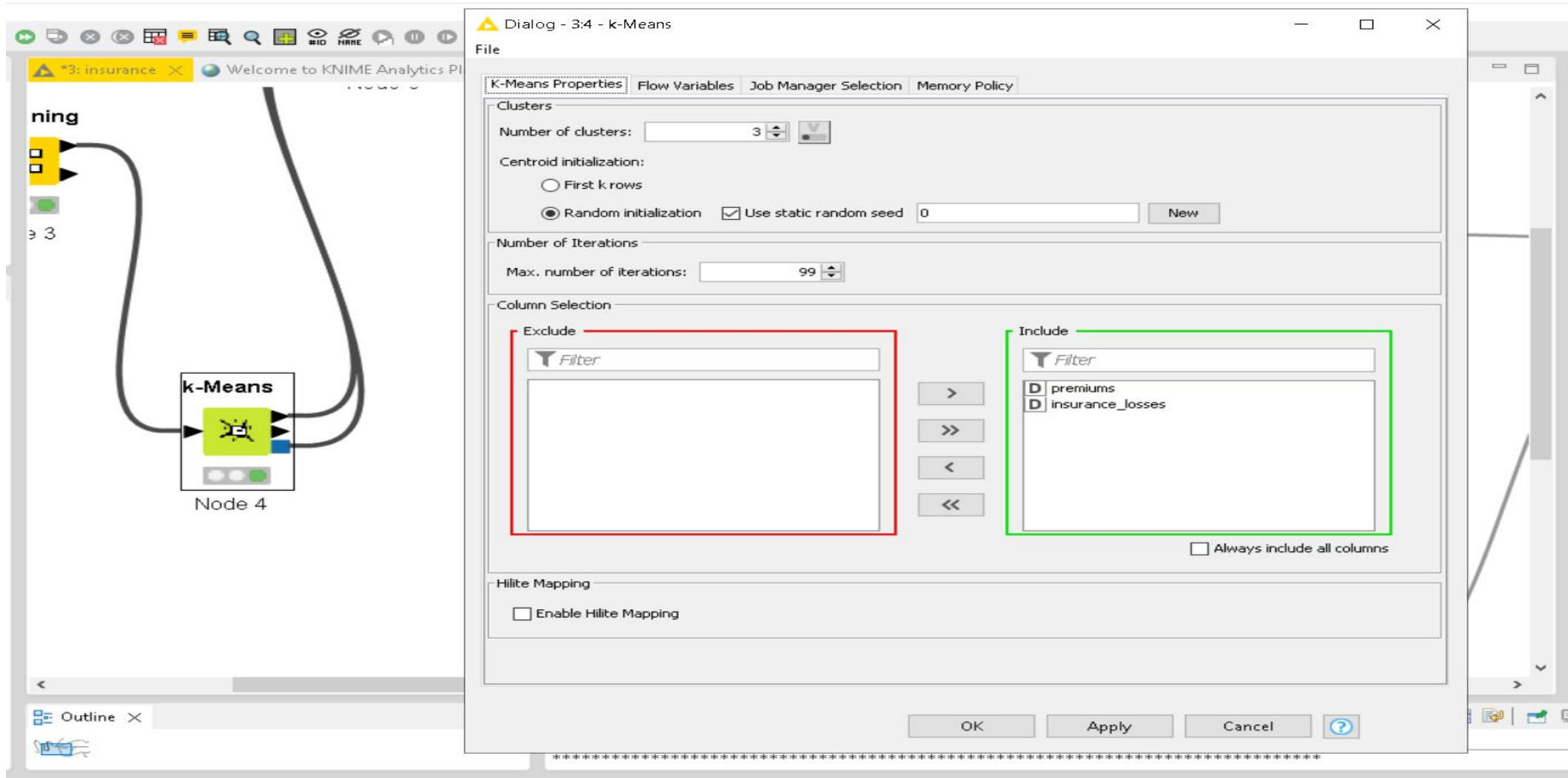
second partitionm

Second partition (remaining rows) - 3:3 - Partitioning

Table "default" - Rows: 16 Spec - Columns: 4 Properties Flow Variables

Row ID	S State	D premiums	D insuran...	S Region
Row1	Alaska	1,053.48	133.93	West
Row2	Arizona	899.47	110.35	West
Row4	California	878.41	165.63	West
Row8	District of C...	1,273.89	136.05	South
Row12	Idaho	641.96	82.75	West
Row13	Illinois	803.11	139.15	Midwest
Row17	Kentucky	872.51	137.13	South
Row18	Louisiana	1,281.55	194.78	South
Row21	Massachusetts	1,011.14	135.63	Northeast
Row28	Nevada	1,029.87	138.71	West
Row31	New Mexico	869.85	120.75	West
Row32	New York	1,234.31	150.01	Northeast
Row37	Oregon	804.71	104.61	West
Row40	South Carolina	858.97	116.29	South
Row43	Texas	1,004.75	156.83	South
Row48	West Virginia	992.61	152.56	South

# K-means for grouping the unlabeled dataset into different clusters.



here clusters are created we can see the coverage, premiums of the insurance , and the loss on the insurance after this i have used cluster assigner it assigns new data to an existing set of prototypes, which are obtained

#### Cluster View - 3:4 - k-Means

File Hilite

##### 3 Clusters

##### cluster\_0 (coverage: 17)

- premiums = 829.2270588235294
- insurance\_losses = 136.9864705882353

##### cluster\_1 (coverage: 7)

- premiums = 1139.5185714285712
- insurance\_losses = 159.4385714285714

##### cluster\_2 (coverage: 11)

- premiums = 695.5236363636363
- insurance\_losses = 112.65

KNIME Explorer

adult

countinue\_of1

insurance

jamisbond

movie\_decision

work1

Node Repository

IO

Manipulation

Views

Analytics

DB

Other Data Types

Structured Data

Scripting

Tools & Services

KNIME Labs

Workflow Control

Workflow Abstraction

Reporting

KNIME Analytics Platform

File Edit View Node Help

150%

Welcome to KNIME Analytics Platform

means

de 4

Cluster Assigner

Node 5

PCA Compute

Node 6

PCA Compute

Node 10

Statistics

Node 11

Linear Correlation

Node 12

Line Plot (local)

Node 14

Cluster Assigner

Dialog - 3:5 - Cluster Assigner

File

Flow Variables Job Manager Selection Memory Policy

Select memory policy for data output(s)

☒ Cache tables in memory.

☐ Write tables to disc.

OK Apply Cancel ?

Cluster Assigner

This node assigns new data to an existing set of prototypes, which are obtained e.g. by a k-means clustering. Each data point is assigned to its nearest prototype.

Ports

Input Ports

0

Prototyp model

1

DataTal containi the inpu data...

Outline

Console

Node Monitor

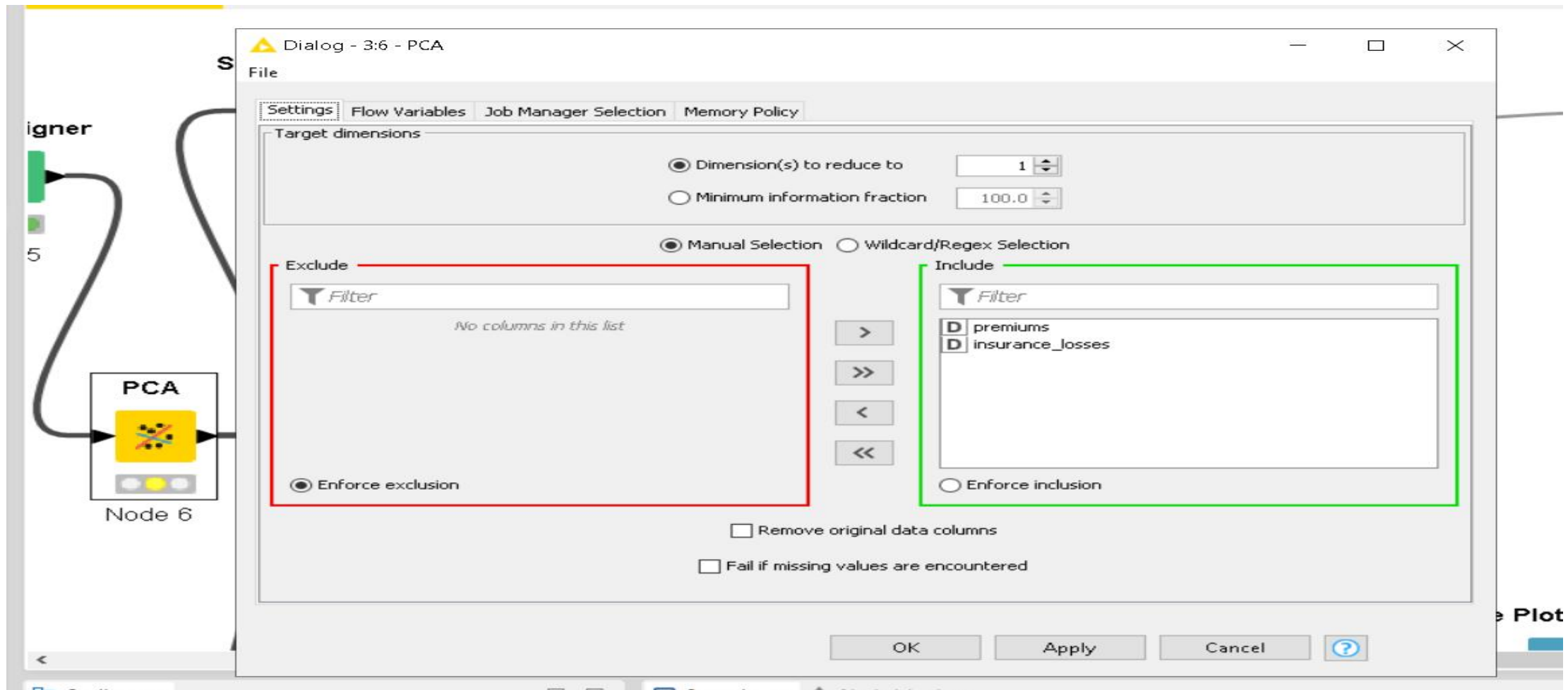
KNIME Console



the new assigned data

Assigned Data - 3:5 - Cluster Assigner						
File Edit Hilite Navigation View						
Table "default" - Rows: 35 Spec - Columns: 6 Properties Flow Variables						
Row ID	S State	D premiums	D insuran...	S Region	S Cluster	S Cluster ...
Row0	Alabama	784.55	145.08	South	cluster_0	cluster_0
Row3	Arkansas	827.34	142.39	South	cluster_0	cluster_0
Row5	Colorado	835.5	139.91	West	cluster_0	cluster_0
Row6	Connecticut	1,068.73	167.02	Northeast	cluster_1	cluster_1
Row7	Delaware	1,137.87	151.48	South	cluster_1	cluster_1
Row9	Florida	1,160.13	144.18	South	cluster_1	cluster_1
Row10	Georgia	913.15	142.8	South	cluster_0	cluster_0
Row11	Hawaii	861.18	120.92	West	cluster_0	cluster_0
Row14	Indiana	710.46	108.92	Midwest	cluster_2	cluster_2
Row15	Iowa	649.06	114.47	Midwest	cluster_2	cluster_2
Row16	Kansas	780.45	133.8	Midwest	cluster_0	cluster_0
Row19	Maine	661.88	96.57	Northeast	cluster_2	cluster_2
Row20	Maryland	1,048.78	192.7	South	cluster_1	cluster_1
Row22	Michigan	1,110.61	152.26	Midwest	cluster_1	cluster_1
Row23	Minnesota	777.18	133.35	Midwest	cluster_0	cluster_0
Row24	Mississippi	896.07	155.77	South	cluster_0	cluster_0
Row25	Missouri	790.32	144.45	Midwest	cluster_0	cluster_0
Row26	Montana	816.21	85.15	West	cluster_0	cluster_0
Row27	Nebraska	732.28	114.82	Midwest	cluster_2	cluster_2
Row29	New Hamps...	746.54	120.21	Northeast	cluster_2	cluster_2
Row30	New Jersey	1,301.52	159.85	Northeast	cluster_1	cluster_1
Row33	North Carolina	708.24	127.82	South	cluster_2	cluster_2
Row34	North Dakota	688.75	109.72	Midwest	cluster_2	cluster_2
Row35	Ohio	697.73	133.52	Midwest	cluster_2	cluster_2
Row36	Oklahoma	881.51	178.86	South	cluster_0	cluster_0
Row38	Pennsylvania	905.99	153.86	Northeast	cluster_0	cluster_0
Row39	Rhode Island	1,148.99	148.58	Northeast	cluster_1	cluster_1
Row41	South Dakota	669.31	96.87	Midwest	cluster_2	cluster_2
Row42	Tennessee	767.91	155.57	South	cluster_0	cluster_0
Row44	Utah	809.38	109.48	West	cluster_0	cluster_0
Row45	Vermont	716.2	109.61	Northeast	cluster_2	cluster_2
Row46	Virginia	768.95	153.72	South	cluster_0	cluster_0
Row47	Washington	890.03	111.62	West	cluster_0	cluster_0
Row49	Wisconsin	670.31	106.62	Midwest	cluster_2	cluster_2
Row50	Wyoming	791.14	122.04	West	cluster_0	cluster_0

This node performs a principal component analysis (PCA) on the given data. The input data is projected from its original feature space into a space of (possibly) lower dimension with a minimum of information loss.



# the transformed data

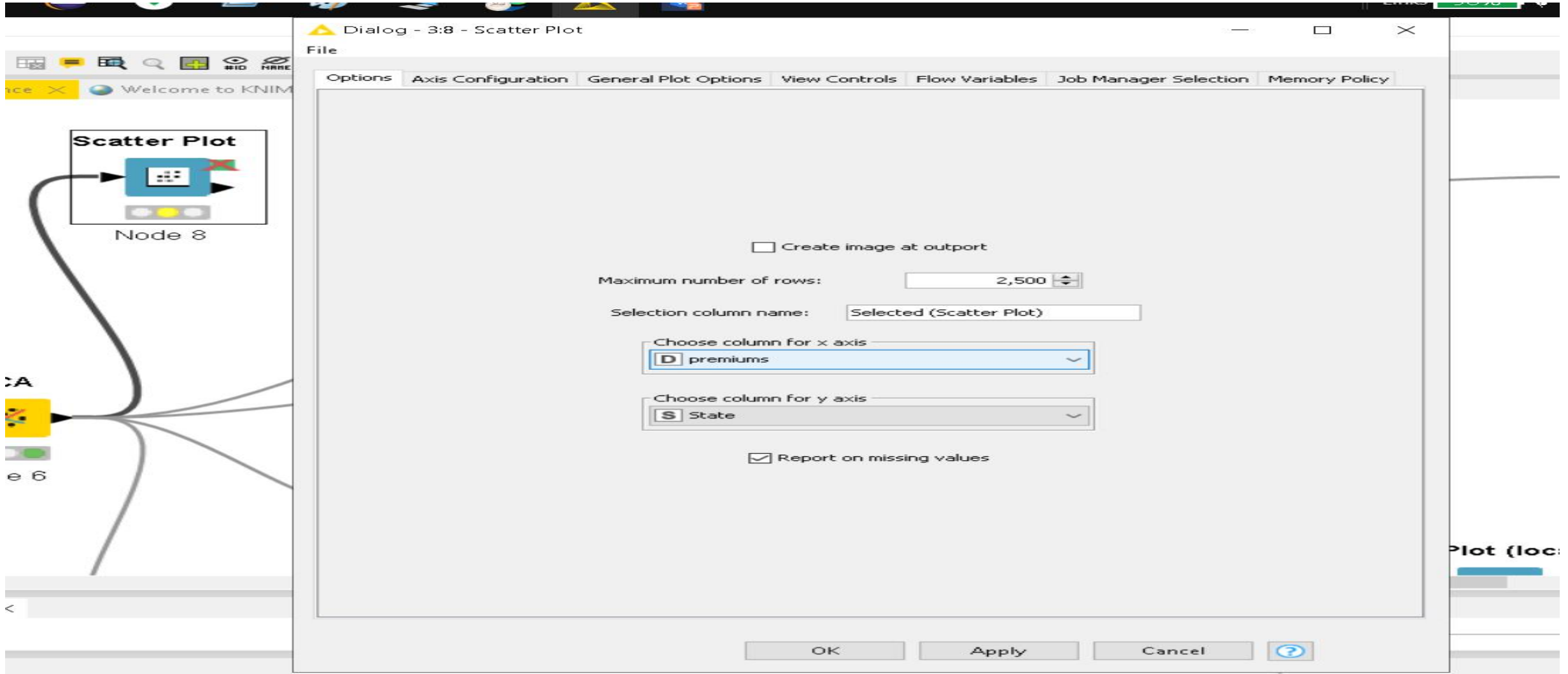
Transformed data - 3:6 - PCA

File Edit Hilite Navigation View

Table "default" - Rows: 35 Spec - Columns: 7 Properties Flow Variables

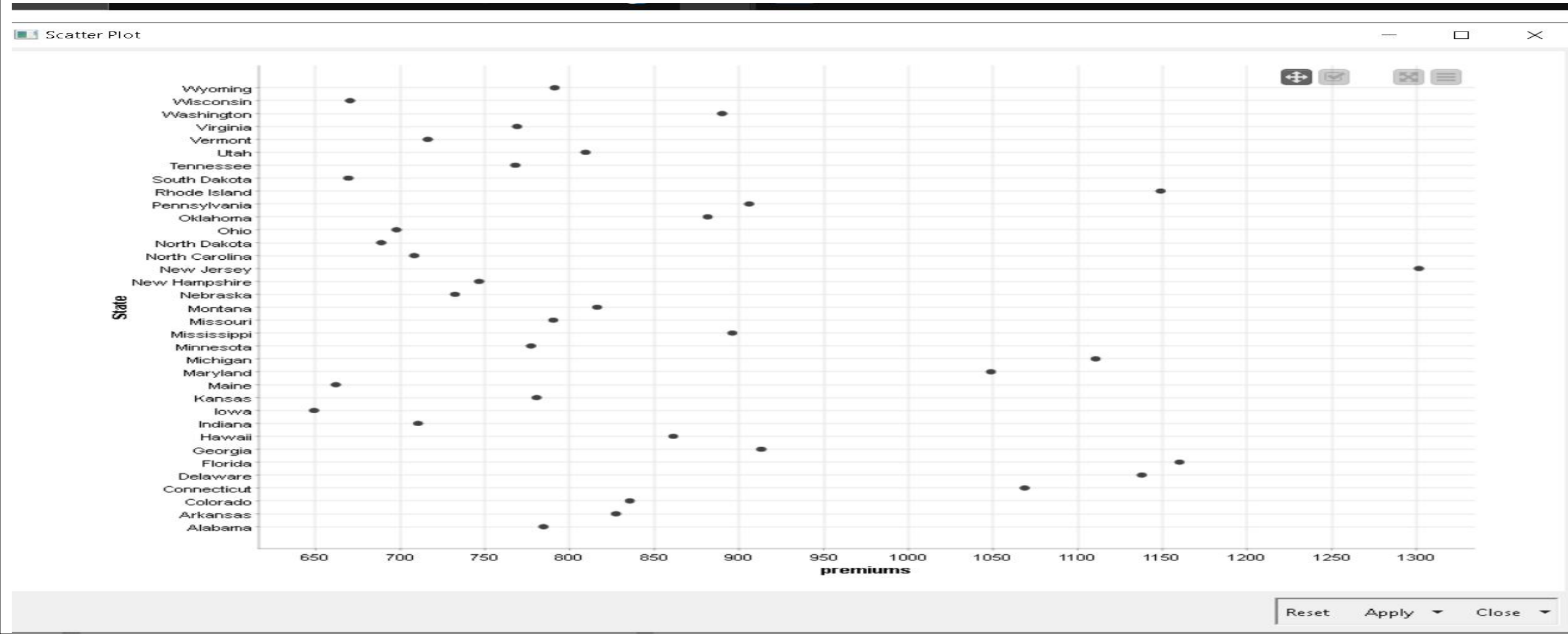
Row ID	\$ State	D premiums	D insuran...	\$ Region	\$ Cluster	\$ Cluster ...	D PCA di...
Row0	Alabama	784.55	145.08	South	cluster_0	cluster_0	-63.393
Row3	Arkansas	827.34	142.39	South	cluster_0	cluster_0	-21.037
Row5	Colorado	835.5	139.91	West	cluster_0	cluster_0	-13.142
Row6	Connecticut	1,068.73	167.02	Northeast	cluster_1	cluster_1	221.597
Row7	Delaware	1,137.87	151.48	South	cluster_1	cluster_1	289
Row9	Florida	1,160.13	144.18	South	cluster_1	cluster_1	310.488
Row10	Georgia	913.15	142.8	South	cluster_0	cluster_0	64.442
Row11	Hawaii	861.18	120.92	West	cluster_0	cluster_0	10.668
Row14	Indiana	710.46	108.92	Midwest	cluster_2	cluster_2	-140.515
Row15	Iowa	649.06	114.47	Midwest	cluster_2	cluster_2	-201.137
Row16	Kansas	780.45	133.8	Midwest	cluster_0	cluster_0	-68.521
Row19	Maine	661.88	96.57	Northeast	cluster_2	cluster_2	-190.03
Row20	Maryland	1,048.78	192.7	South	cluster_1	cluster_1	204.112
Row22	Michigan	1,110.61	152.26	Midwest	cluster_1	cluster_1	261.929
Row23	Minnesota	777.18	133.35	Midwest	cluster_0	cluster_0	-71.819
Row24	Mississippi	896.07	155.77	South	cluster_0	cluster_0	48.637
Row25	Missouri	790.32	144.45	Midwest	cluster_0	cluster_0	-57.707
Row26	Montana	816.21	85.15	West	cluster_0	cluster_0	-37.422
Row27	Nebraska	732.28	114.82	Midwest	cluster_2	cluster_2	-118.242
Row29	New Hamps...	746.54	120.21	Northeast	cluster_2	cluster_2	-103.544
Row30	New Jersey	1,301.52	159.85	Northeast	cluster_1	cluster_1	452.721
Row33	North Carolina	708.24	127.82	South	cluster_2	cluster_2	-140.974
Row34	North Dakota	688.75	109.72	Midwest	cluster_2	cluster_2	-162.058
Row35	Ohio	697.73	133.52	Midwest	cluster_2	cluster_2	-150.911
Row36	Oklahoma	881.51	178.86	South	cluster_0	cluster_0	36.279
Row38	Pennsylvania	905.99	153.86	Northeast	cluster_0	cluster_0	58.338
Row39	Rhode Island	1,148.99	148.58	Northeast	cluster_1	cluster_1	299.803
Row41	South Dakota	669.31	96.87	Midwest	cluster_2	cluster_2	-182.605
Row42	Tennessee	767.91	155.57	South	cluster_0	cluster_0	-78.99
Row44	Utah	809.38	109.48	West	cluster_0	cluster_0	-41.969
Row45	Vermont	716.2	109.61	Northeast	cluster_2	cluster_2	-134.736
Row46	Virginia	768.95	153.72	South	cluster_0	cluster_0	-78.126
Row47	Washington	890.03	111.62	West	cluster_0	cluster_0	38.533
Row49	Wisconsin	670.31	106.62	Midwest	cluster_2	cluster_2	-180.705
Row50	Wyoming	791.14	122.04	West	cluster_0	cluster_0	-58.966

after that we plotting the data in x & y axis

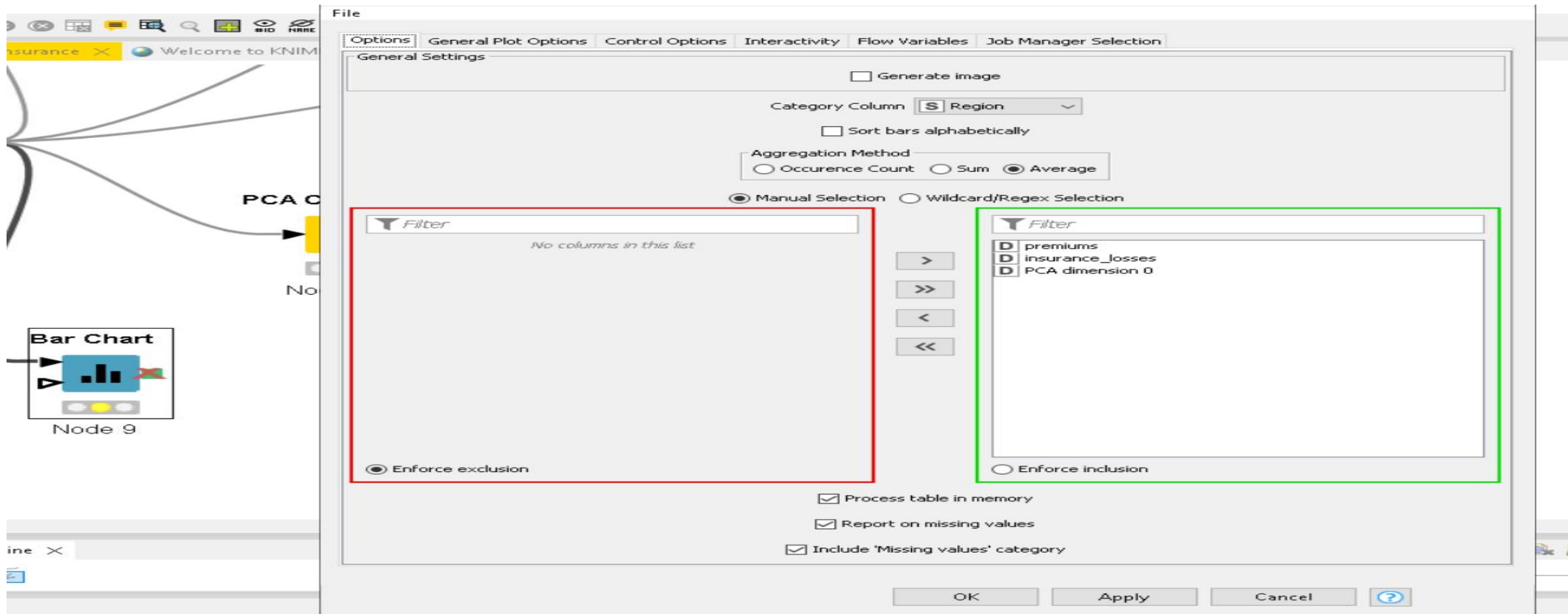




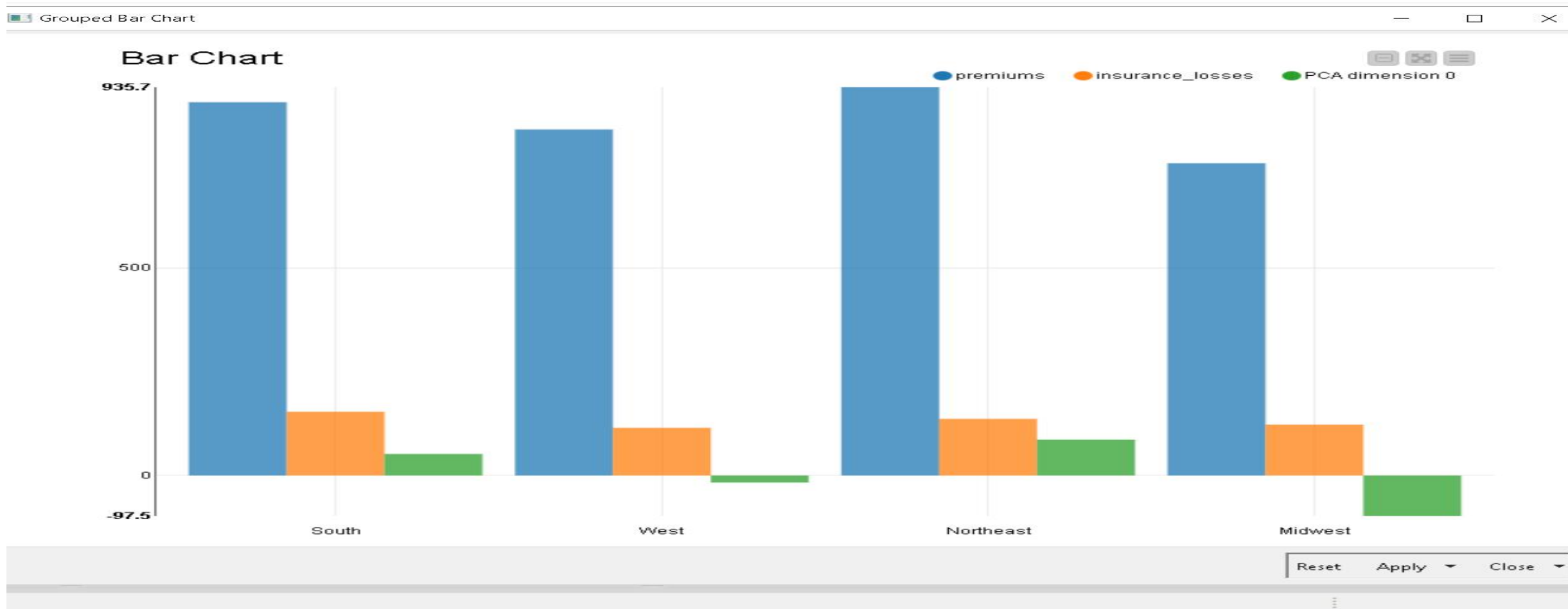
the states which have maximum premiums holder are from 750 -800



also in bar chart plotting is done



in midwest the coverage of insurance is going in negative  
same in west  
and northeast has the most premiums



Calculates for each pair of selected columns a correlation coefficient, i.e. a measure of the correlation of the two variables.

The screenshot displays the KNIME Analytics Platform interface with three nodes visible in the workflow: **Statistics** (Node 11), **Linear Correlation** (Node 12), and **Line Plot (1d)** (Node 14). The **Linear Correlation** node is selected, and its configuration dialog, titled "Dialog - 3:12 - Linear Correlation", is open.

The dialog features several tabs: **Options** (selected), **Flow Variables**, **Job Manager Selection**, and **Memory Policy**. Under the **Options** tab, the **Manual Selection** radio button is chosen over **Wildcard/Regex Selection**.

The dialog is divided into two main sections for column selection:

- Exclude** (outlined in red): Contains a "Filter" input field and the text "No columns in this list". The **Enforce exclusion** radio button is selected.
- Include** (outlined in green): Contains a "Filter" input field and a list of available columns: State, premiums, insurance\_losses, Region, Cluster, Cluster (#1), and PCA dimension 0. The **Enforce inclusion** radio button is unselected.

Between these sections are navigation buttons: >, >>, <, and <<.

Below the selection sections, the **Output column pairs** section has three radio buttons:

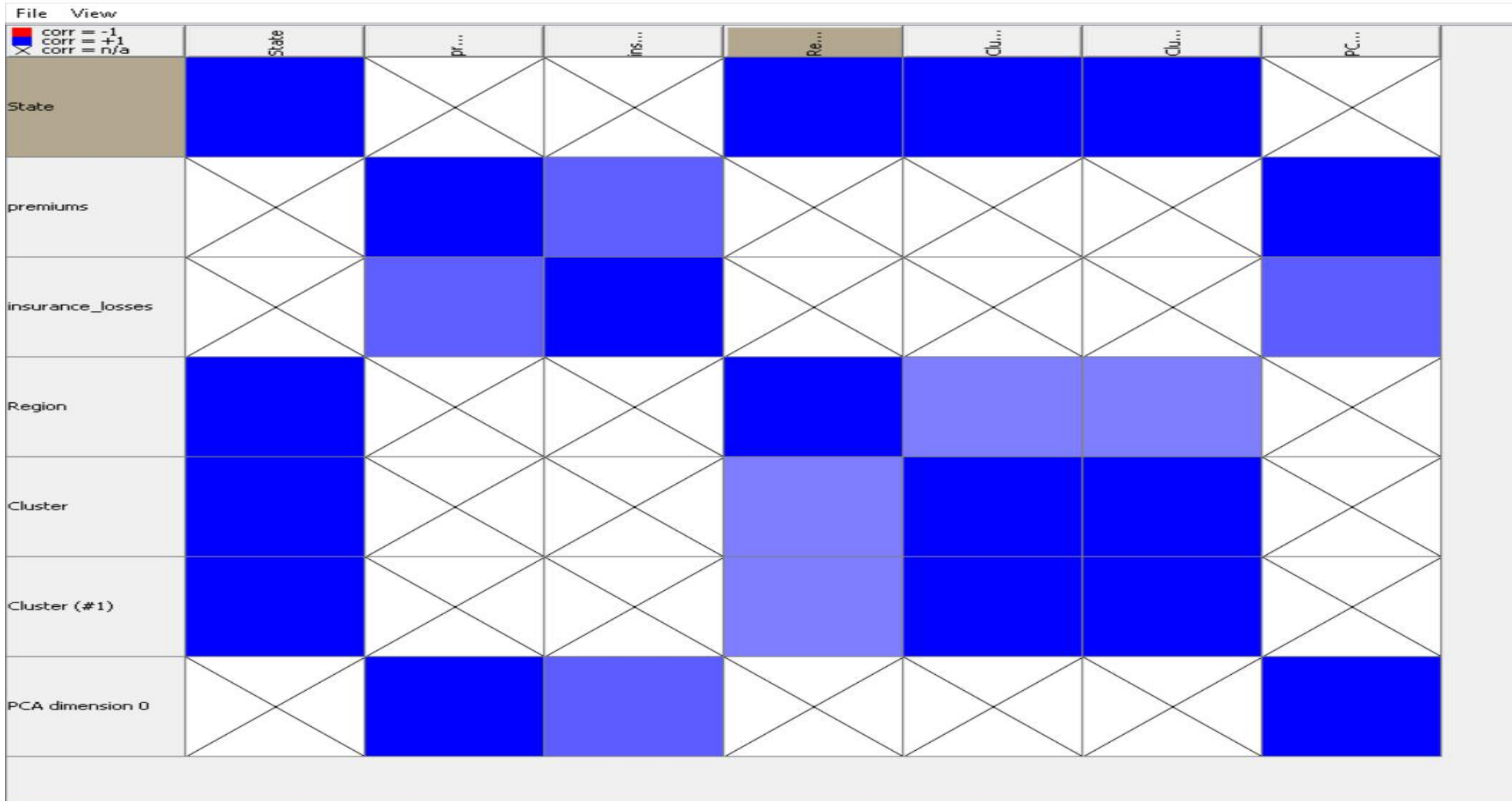
- ☐ Include all column pairs
- ☒ Include only column pairs of compatible columns
- ☐ Include only column pairs with a valid correlation

The **Possible Values Count** is set to 50. The **p-value** section has three radio buttons:

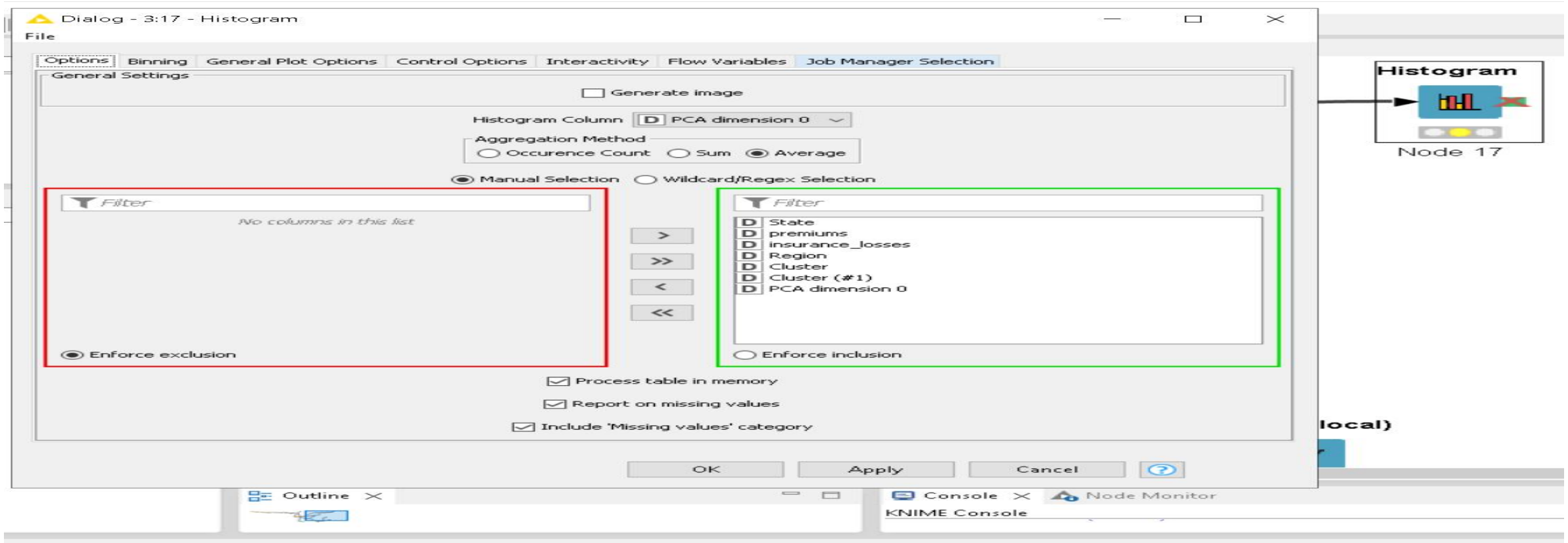
- ☒ two-sided
- ☐ one-sided (right)
- ☐ one-sided (left)

At the bottom of the dialog are the **OK**, **Apply**, **Cancel**, and a help icon buttons.

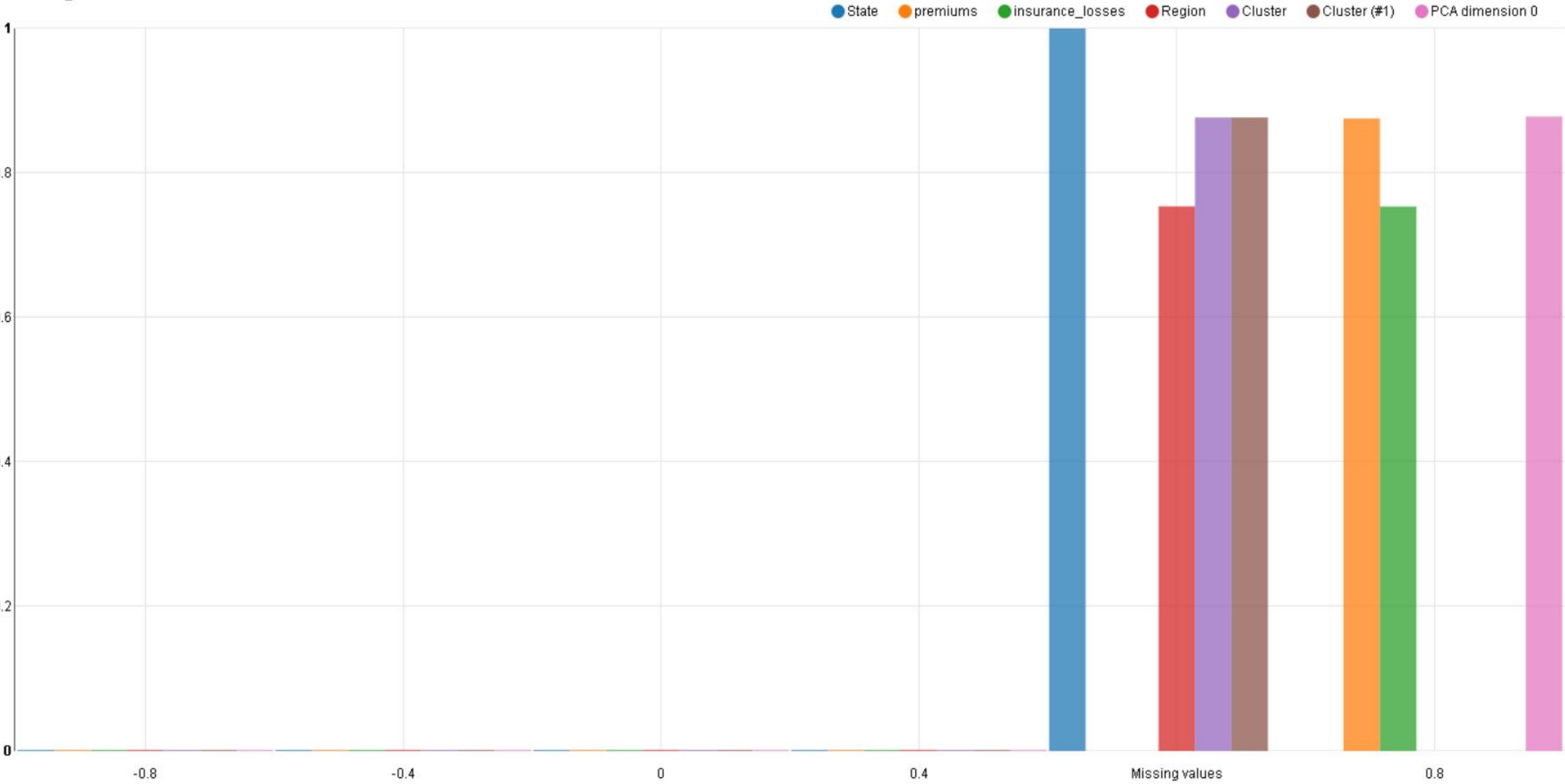
Correlation Matrix - 3:12 - Linear Correlation



after selection put CSS rules into a single string and set it as a flow variable 'customCSS' so i have taken this columns

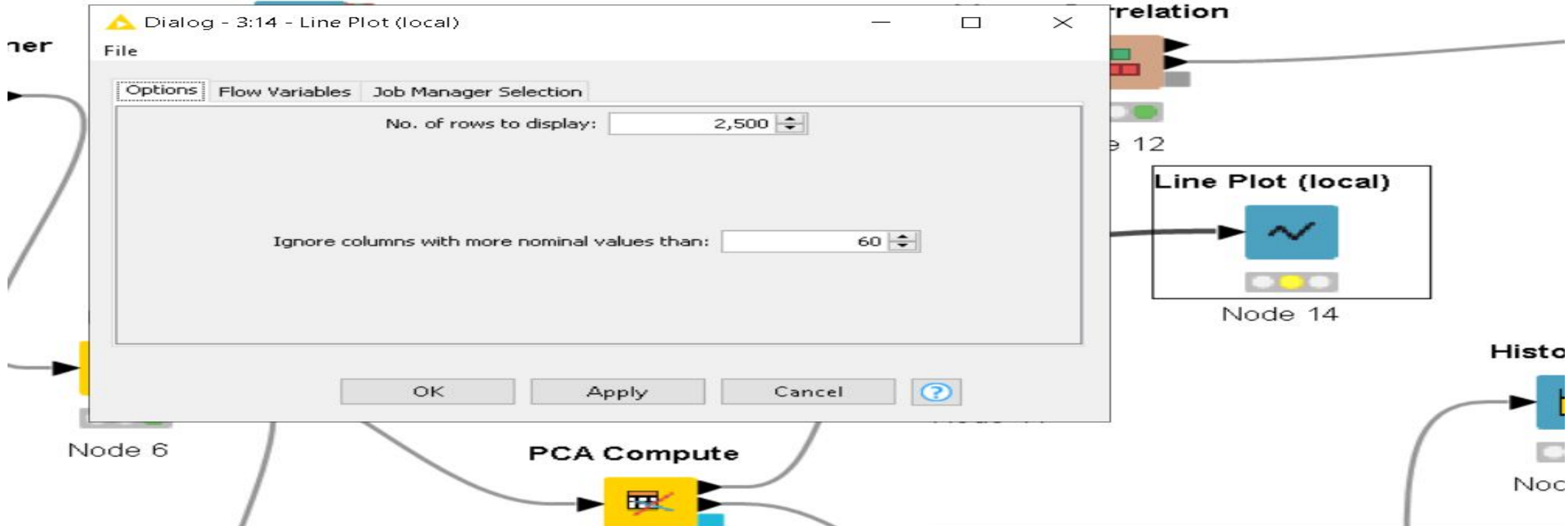


Histogram



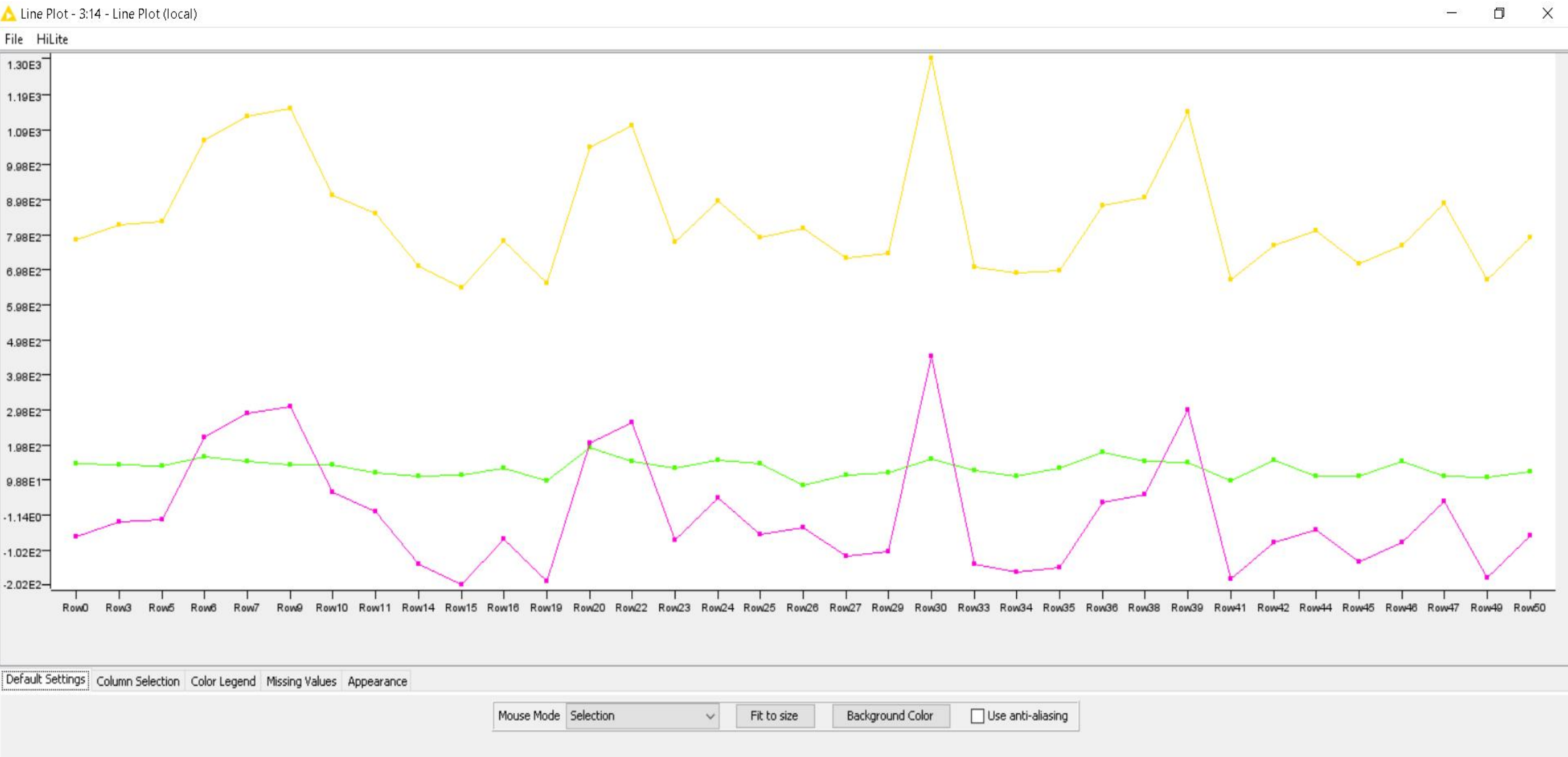
also line plot is necessary to understand the story

### Scatter Plot





so company seems to be in heavy loss lets further investigate the data



after watching all plots now connect pca with this node performs a principal component analysis (PCA) on the given input data. The directions of maximal variance (the principal components) are extracted and can be used in the PCA Apply node to project the input into a space of lower dimension while preserving a maximum of information.

3: insurance X

Welcome to KNIME Analytics Platform

Line Plot (local)

Dialog - 3:10 - PCA Compute

File

Settings

Flow Variables

Job Manager Selection

Memory Policy

Manual Selection

Wildcard/Regex Selection

Exclude

Filter

No columns in this list

Enforce exclusion

Include

Filter

premiums

insurance\_losses

PCA dimension 0

Enforce inclusion

>

>>

<

<<

Force the set of excluded columns to stay the same.

Fail if missing values are encountered

OK

Apply

Cancel

?

PCA Compute

Node 10

Node 13

Transformation model - 3:10 - PCA Compute

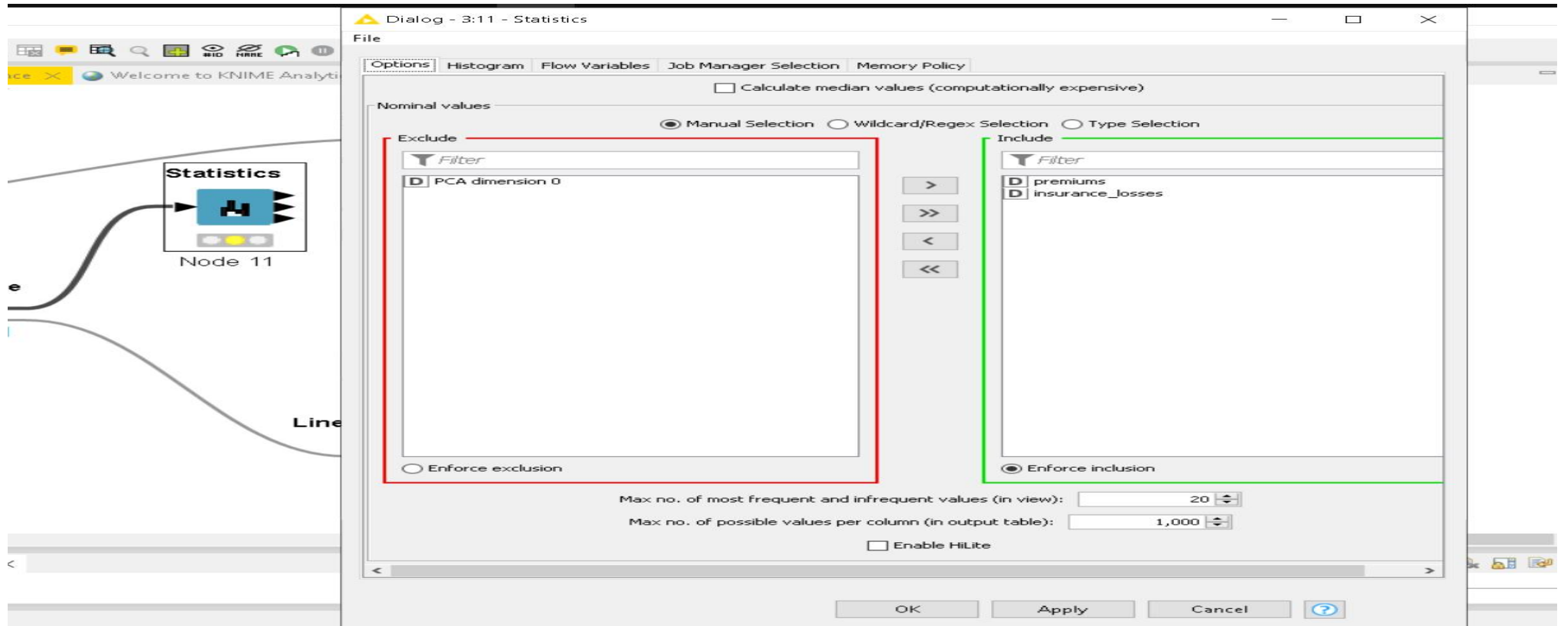
File

Model Content Model Content Spec Flow Variables

Model Content




- max\_dim\_to\_reduce\_to [xint] -> 3
- eigenvalues
  - array-size [xint] -> 3
  - 0 [xdouble] -> 56344.06026054872
  - 1 [xdouble] -> 362.18122720463464
  - 2 [xdouble] -> 0.0
- center
  - array-size [xint] -> 3
  - 0 [xdouble] -> 849.2642857142854
  - 1 [xdouble] -> 133.8282857142857
  - 2 [xdouble] -> 3.199662756969701E-13
- number\_rows [xint] -> 3
- eigenvector
  - row\_0
    - array-size [xint] -> 3
    - 0 [xdouble] -> 0.704064965981454
    - 1 [xdouble] -> 0.06551735401811051
    - 2 [xdouble] -> 0.7071067811865475
  - row\_1
    - array-size [xint] -> 3
    - 0 [xdouble] -> 0.09265553062320986
    - 1 [xdouble] -> -0.9956982236827239
    - 2 [xdouble] -> 1.617900691891495E-15
  - row\_2
    - array-size [xint] -> 3
    - 0 [xdouble] -> -0.7040649659814538
    - 1 [xdouble] -> -0.06551735401810845
    - 2 [xdouble] -> 0.7071067811865477

now i have used statistics to calculate statistical moments such as minimum, maximum, mean, standard deviation, variance, median, overall sum, number of missing values and row count across all numeric columns, and counts all nominal values together with their occurrences. The dialog offers two options for choosing the median and/or nominal values calculations:



File

Numeric Nominal Top/bottom

Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	No. Missing	No. + $\infty$	No. - $\infty$	Histogram
premiums	2,565.6518	19,516.5911	?	28,050.8404	14,680.0618	-1.7319	?	0	0	0	
insurance_losses	600.9301	1,925.6254	?	2,610.2944	1,147.4369	-1.7291	?	0	0	0	
PCA dimension 0	2,610.2944	19,611.055	?	28,172.0301	14,723.2152	-1.7319	?	0	0	0	

# again finding the correlations

The image shows the KNIME Analytics Platform interface. On the left, a workflow diagram includes a node labeled "Linear Correlation" (Node 13). A red box highlights this node, and a red arrow points from it to the "Dialog - 3:13 - Linear Correlation" configuration window on the right.

The configuration window has several tabs: "Options", "Flow Variables", "Job Manager Selection", and "Memory Policy". The "Options" tab is active.

Under the "Options" tab, there are two main sections: "Exclude" (outlined in red) and "Include" (outlined in green). Both sections have a "Filter" input field and a list of columns. The "Exclude" section is currently empty, with the text "No columns in this list". The "Include" section contains a list of columns: "eigenvalue", "premiums", "insurance\_losses", and "PCA dimension 0".

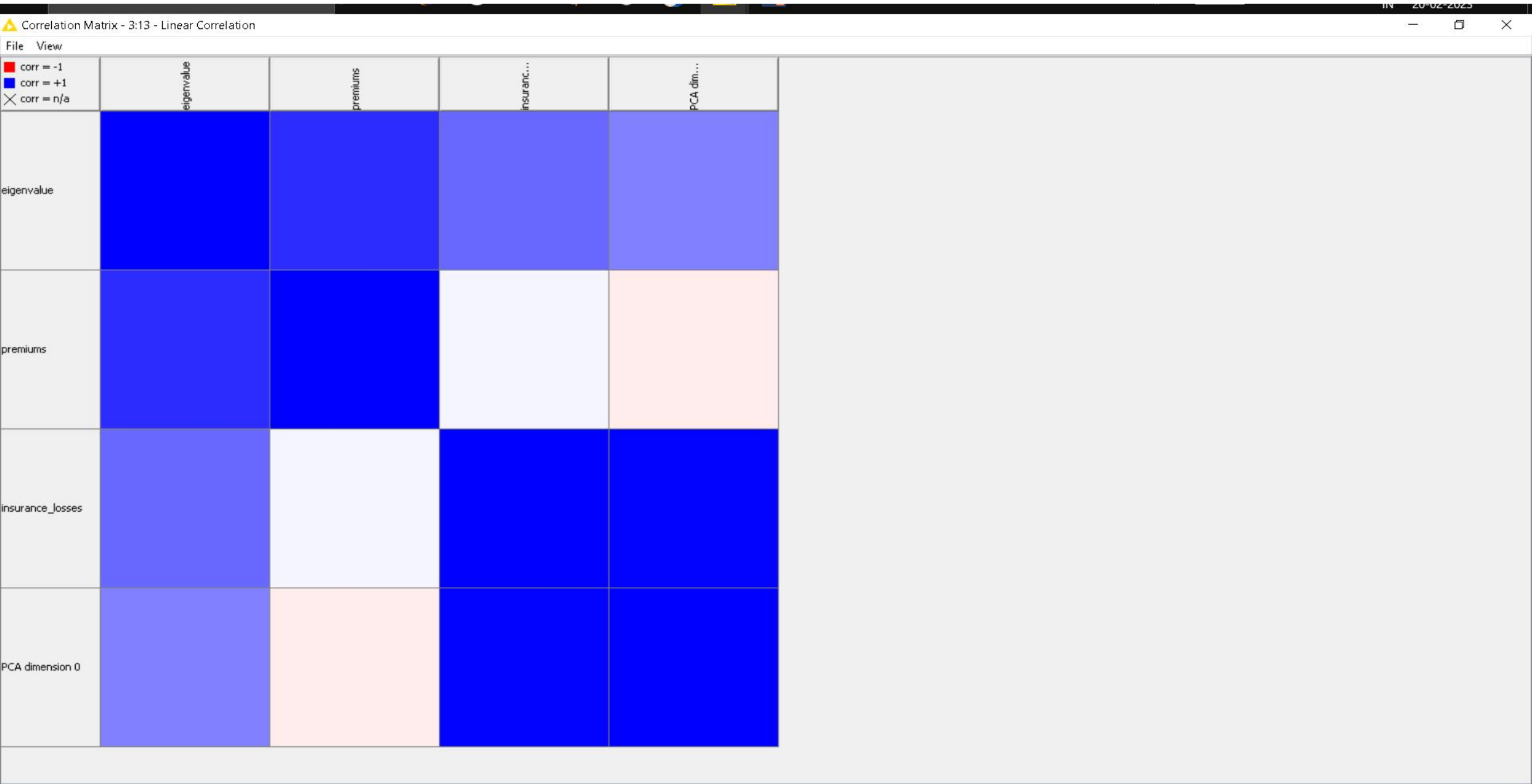
Below these sections, there are radio buttons for "Enforce exclusion" (selected) and "Enforce inclusion".

At the bottom, there are options for "Output column pairs": "Include all column pairs", "Include only column pairs of compatible columns" (selected), and "Include only column pairs with a valid correlation".

Below the output options, there is a "Possible Values Count" field set to "50".

At the very bottom, there are radio buttons for "p-value": "two-sided" (selected), "one-sided (right)", and "one-sided (left)".

At the bottom right of the dialog, there are buttons for "OK", "Apply", "Cancel", and a help icon.



i have taken degree of freedom as column

Dialog - 3:16 - Histogram

File

Options

Binning

General Plot Options

Control Options

Interactivity

Flow Variables

Job Manager Selection

General Settings

☐ Generate image

Histogram Column

I

Degrees of freedom

Aggregation Method

☐ Occurence Count

☐ Sum

☒ Average

☒ Manual Selection

☐ Wildcard/Regex Selection

Filter

No columns in this list

>

>>

<

<<

☒ Enforce exclusion

D

Correlation value

D

p value

I

Degrees of freedom

☐ Enforce inclusion

☒ Process table in memory

☒ Report on missing values

☒ Include 'Missing values' category


OK

Apply

Cancel


?

Histogram



Node 16

Line Plot (local)



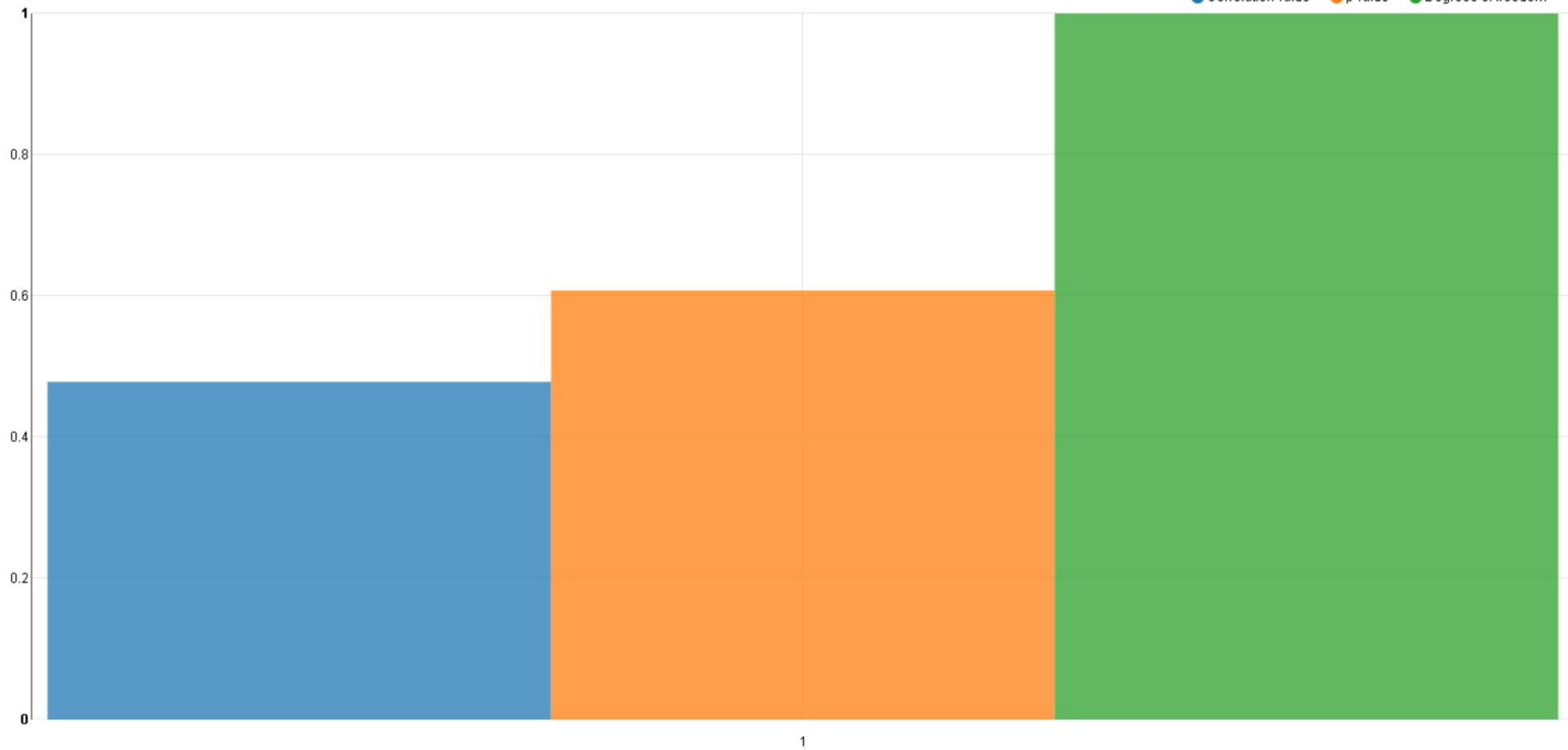
Node 15



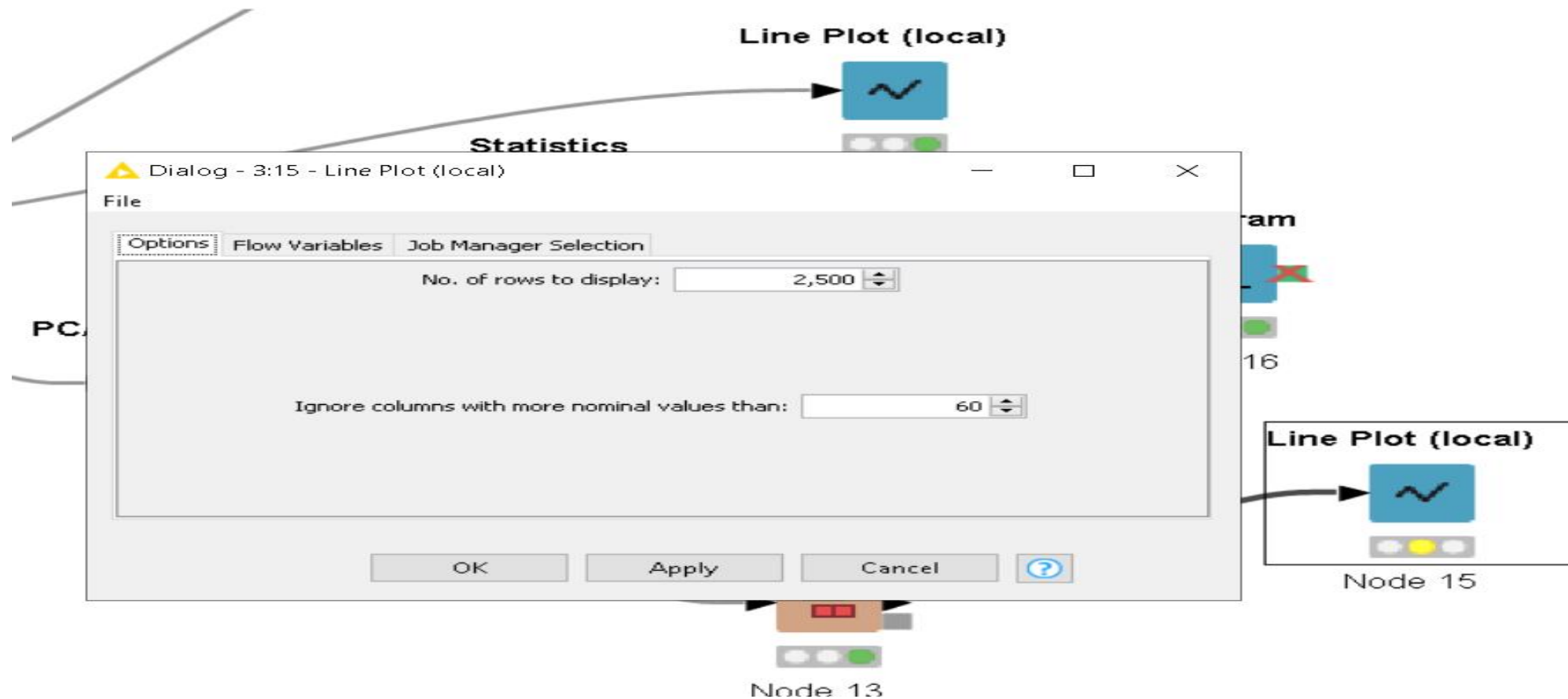
# Histogram



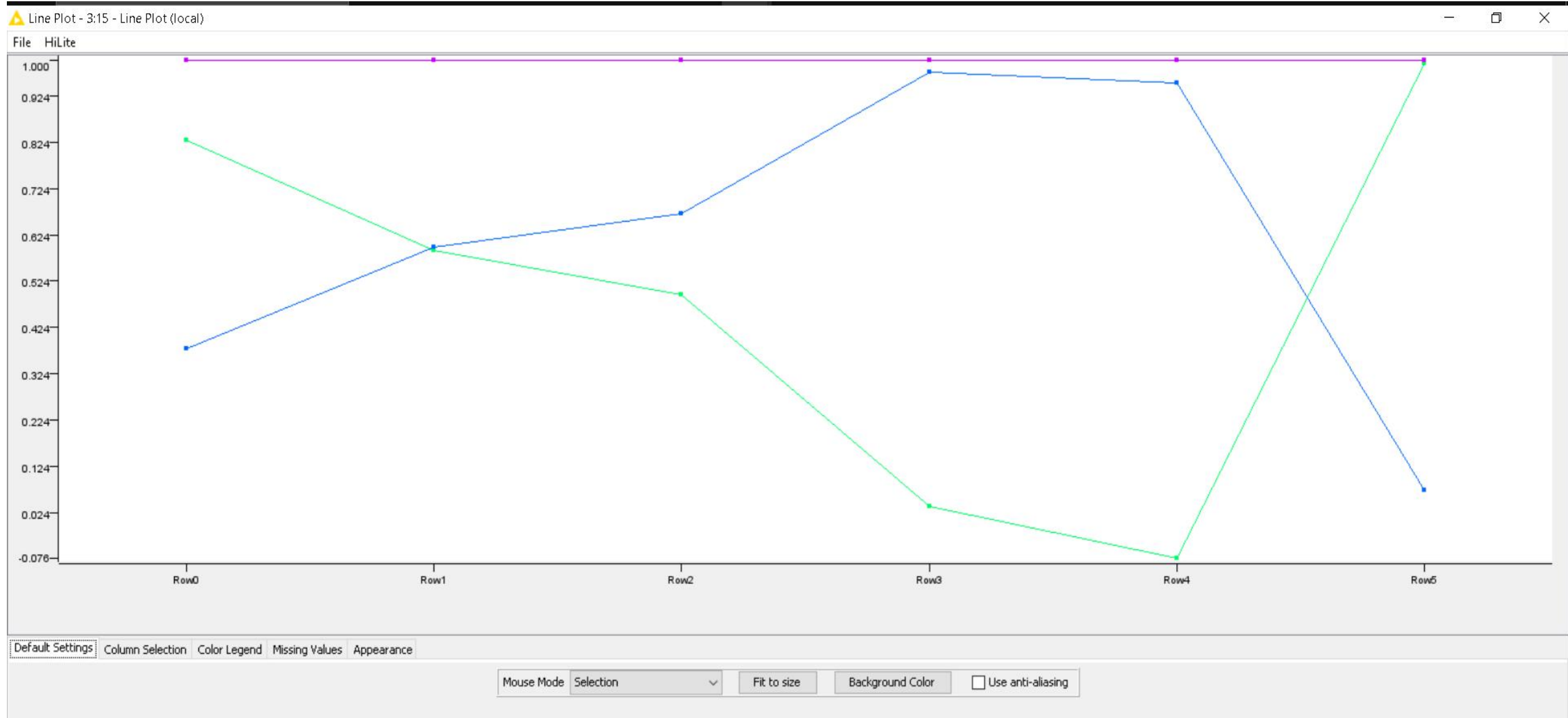
● Correlation value ● p value ● Degrees of freedom



now line plot



but as we can see there is a huge growth no loss in the line plot



Thanks for watching

