

# User Manual

Jacob Stolk PhD Simone Stolk MPH July 2021

# Contents

Introduction1
Start DIONE
Load Data3
Missing Values
Complete and Incomplete Cases7
Explore Data9
One Variable9
Time Series Data11
Two Variables13
Inspect Results15
All Variables
Copy and Save Results20
Recode Data23
Recode Data
Numerical to Categorical23
Numerical to Categorical23 Categorical to Categorical
Numerical to Categorical
Numerical to Categorical23Categorical to Categorical27Learn Bayesian Networks30Learn Network Structure30
Numerical to Categorical23Categorical to Categorical27Learn Bayesian Networks30Learn Network Structure30Edit Network Structure33
Numerical to Categorical23Categorical to Categorical.27Learn Bayesian Networks30Learn Network Structure30Edit Network Structure33Fit Parameters37
Numerical to Categorical23Categorical to Categorical.27Learn Bayesian Networks30Learn Network Structure30Edit Network Structure33Fit Parameters.37Infer Probabilities.39
Numerical to Categorical23Categorical to Categorical.27Learn Bayesian Networks30Learn Network Structure30Edit Network Structure33Fit Parameters37Infer Probabilities.39Copy and Save Results44

## Introduction

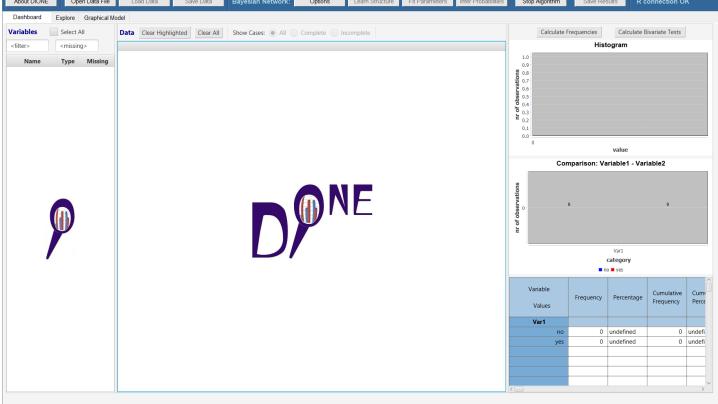
Data analysis with **DIONE** is typically done in four main steps:

- Load the data to be analysed. In the current version data can be loaded from CSV files and from common statistics data file formats. When loading data from a CSV file, the data file is assumed to have a header record containing variable names and any number of data records containing values of the variables. Variable names and values are separated by commas (,), tabs (\t), semicolons (;), colons(:) or vertical slashes (|) and optionally enclosed in double quotes ("). Each record is delimited by a line break (CRLF).
- Explore the data using standard statistical tools to get a better intuition of data characteristics, trends, relationships, etc. Each variable can easily be inspected individually, showing plots of its distribution and statistical information. Pairs of variables can easily be compared for a first intuition of possible relationships between variables.
- 3. Recode data, if needed: for example, transform numerical data to categorical data or group categories of categorical data to new categories.
- 4. Analyse the data using sophisticated data analysis algorithms. At present multivariate analysis is implemented by learning a Bayesian network from the data values of the variables of interest, which will be represented by nodes in the network. The learned network can then be used for inference and for estimating effects of interventions.
- Save data and/or analysis results. After selecting and/or recoding data for your analysis, you may wish to keep these data to continue your analysis at another time.
   DIONE makes it easy to do this. It is also easy to save analysis results as tables, graphs and charts, ready for your publications.

This guide shows and explains these steps with a simple demonstration data set with fictitious data on variables such as having visited a country with a high incidence of tuberculosis, smoking, age, having the diseases of tuberculosis or lung cancer, and X-ray results. You can of course do similar analyses of your own data.

## **Start DIONE**

To launch DIONE, double-click the will desktop icon or select DIONE in the Windows Start Menu A The DIONE screen now looks like this:



At the top of the screen there is a toolbar with buttons for loading and saving data and for data analysis. To the right of the buttons there is an area displaying messages about what the program is doing (in light blue), including possible error messages (in red):

DIONE - Decision Intelligence for Organizations in Network Environments	- 🗆 X
About DIONE Open Data File Load Data Save Data Bayesian Network: Options Learn Structure Fit Parameters Infer Probabilities Stop Algorithm Save Results 0%	R connection OK
The main area of the screen has three tabs: Dashboard , Explore and	
Graphical Model . Initially the Dashboard tab is selected.	
In the Dashboard tab there are two tables: after loading data the Variables table	
shows the variable names and the <b>Data</b> table shows data of selected variables. At the top	
right-hand side there are visual representations of selected data, and at the bottom right-	
hand side there is a table with frequency distributions and test results of selected data. In the	
Explore tab a more detailed view of exploratory data analysis results is shown and	
there are options to recode data. The Graphical Model tab will show results of Bayesian	
network learning and inference algorithms.	

## Load Data

Data can be loaded from text (.csv) files and from common statistics data file formats, including data files with extension .sav, .dta or .sas7bdat. Column headers of a text (.csv) data file must be variable names. In this guide the text file healthdemo.csv is used. To open a data file, press the Open Data File button to bring up the Open Data File dialog. Now select the data file to analyse and press the Open button:

		Open Data File					×
		$\leftarrow \rightarrow \land \uparrow$	« Docu	ments > Dione > data > v (	ט		ta
							?
DIONE - Dec	sision Intelligence	e for Organizations ir	n Networ	Name	Dat	e modified	Туре
About DION	NE Open	Rata File Loa	d Data	Simone	6/04	4/2019 6:48 AM	File folder
Dashboard	Explore	Graphical Model	Geneti	🔊 alarm.csv	8/04	4/2019 3:37 PM	Microsoft Excel Co
	Explore	· ·		ChronicIIInessCaregivingData.csv	5/02	2/2018 11:19 A	Microsoft Excel Co
/ariables		Data	Cl	ChronicIIInessCaregiving-Health.csv	10/0	)1/2020 4:58 A	Microsoft Excel Co
<filter></filter>	<missing></missing>			ChroniclIInessCaregiving-HealthRecoded	10/0	01/2020 5:02 A	Microsoft Excel Co
		JU Objects		🛯 healthdemo - Copy.csv	24/0	04/2019 8:12 A	Microsoft Excel Co
		Desktop		💶 healthdemo.csv	10/0	)5/2019 9:10 A	Microsoft Excel Co
		Documents		healthdemo+recoded.csv	20/0	)7/2018 7:53 A	Microsoft Excel Co
		Downloads		healthdemo1.csv	4/07	7/2018 8:28 AM	Microsoft Excel Co
		Music		healthdemomissing.csv	8/04	4/2019 3:39 PM	Microsoft Excel Co
		Pictures		healthdemoRecoded.csv	18/0	)7/2018 9:25 A	Microsoft Excel Co
		Pictures	~ <				>
			File <u>n</u> ame:	healthdemo.csv	$\sim$	Data Files (*.csv	) ~
						Open N	Cancel

Now you see the variable names (column headers in the text file) in the **Variables** table at the left of the screen (data values will be loaded in the next step):

DIONE - Decision I	ntellige	nce for Organiz	ations in N	letwork	Environments : data	file C:\Users\hj_vv\Doc
About DIONE	Open	Data File	Load Da	ta	Save Data	Bayesian Network
Dashboard Exp	lore	Graphical Mode				
Variables	Select A	I	Data	Clear	Highlighted Clea	r All Show Cases:
<filter></filter>	<m< th=""><th>issing&gt;</th><th></th><th></th><th></th><th></th></m<>	issing>				
Name	Туре	Missing				
Smoker		<missing></missing>				
LungCancer	N	<missing></missing>				
VisitToTubia	R	<missing></missing>				
Tuberculosis		<missing></missing>				
TuberculosisOrCan		<missing></missing>				
X_ray		<missing></missing>				
Bronchitis		<missing></missing>				
Dyspnea		<missing></missing>				
Age		<missing></missing>				
NrSmoked		<missing></missing>				

The second column in the **Variables** table is the variables' data type (categorical, ordinal or numerical), to be shown after loading the data values. The third column shows the code for missing values in the data (default is <missing>, but this can be changed if needed).

Select the variables you want to analyse (possibly all) by clicking on their names. You can select variables one at a time, use Shift-click or Ctrl-click for multiple selection or check Select All to select all variables.

If there are many variables, you can easily locate desired variables by typing part of their names in the <filter> text box (clearing the <filter> text box shows all variables again). When you are satisfied with your selection of variables, press the Load Data button to load the data for the selected variables. You can add more variables and press Load Data again, if desired. You can also press the Enter key or double-click to load variables. You can now see data corresponding to the selected variables in the middle of the screen, for example with 4 variables selected and loaded here:

DIONE - Decision I	ntellige	nce for Organiz	ations in N	letwor	k Environments	: data file C:\U	sers\hj_vv\Doc	uments\Di
About DIONE	Open	Data File	Load Da	ta	Save Dat	a Bayes	sian Network	c: Alg
Dashboard Exp	olore	Graphical Model						
Variables	Select A	11	Data	Clear	Highlighted	Clear All	Show Cases: (	IIA (
<filter></filter>	<m< th=""><th>issing&gt;</th><th>Smol</th><th>ker</th><th>LungCancer</th><th>VisitToTubia</th><th>Tuberculosis</th><th>;</th></m<>	issing>	Smol	ker	LungCancer	VisitToTubia	Tuberculosis	;
Name	Туре	Missing	yes		yes	no	no	
Smoker		<missing></missing>	no		yes	no	no	
LungCancer		<missing></missing>	yes		yes	no	no	
VisitToTubia		<missing></missing>	yes		yes	no	no	
Tuberculosis	С	<missing></missing>	yes		yes	no	no	
TuberculosisOrCan		<missing></missing>	yes		yes	no	no	
X_ray		<missing></missing>	yes		yes	no	no	
Bronchitis		<missing></missing>	yes		yes	no	no	
Dyspnea		<missing></missing>	no		yes	no	no	
Age		<missing></missing>	yes		yes	no	no	
NrSmoked		<missing></missing>	yes		yes	no	no	

#### Here all variables have been selected and loaded:

About DIONE	Open	Data File	Load Data	Save Dat	a Bayes	an Network	Algorithm	ns Lea	rn Structure	Fit Parameters	Infer Proba	bilities Sto
Dashboard Exp	olore	Graphical Mode	ł									
Variables	Select A	All	Data Clea	ar Highlighted	Clear All	Show Cases:	All 🔵 Con	nplete 🔵 In	complete 12	99 cases for 10 va	riables	
<filter></filter>	< <b>n</b>	nissing>	Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis	X_ray	Bronchitis	Dyspnea	Age	NrSmoked
Name	Туре	Missing	billokei	Lungeuneer	VISICIOTUSIA	Tuberculosis	OrCancer	N_INJ	bronemas	Dyspired	Age	Monokeu
Smoker	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	53	9
ungCancer		<missing></missing>	no	yes	no	no	yes	yes	no	no	62	0
/isitToTubia		<missing></missing>	yes	yes	no	no	yes	yes	no	yes	62	8
uberculosis		<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	74	10
uberculosisOrCan		<missing></missing>	yes	yes	no	no	yes	yes	no	yes	47	6
ray		<missing></missing>	yes	yes	no	no	yes	yes	yes	no	57	4
ronchitis		<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	71	10
yspnea		<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	61	6
lge		<missing></missing>	no	yes	no	no	yes	yes	no	yes	53	0
- IrSmoked		<missing></missing>	yes	yes	no	no	yes	yes	no	yes	48	4
			yes	yes	no	no	yes	yes	yes	yes	72	9
			yes	yes	no	no	yes	yes	yes	yes	46	8

The column **Type** in the **Variables** table shows the data type of the variables, categorical (C), ordinal (O) or numerical (N). **DIONE** infers the data type from the loaded data. If the data for a variable have text values such as yes or no, the variable is categorical. If the data of a variable consist of numbers, the variable is numerical, unless the number of distinct number values is less than 12, in which case it is assumed the numbers are codes for different levels and the variable is categorical. The data type of a variable can be changed by clicking in the type cell of the variable and selecting a different type, for example making variable **NrSmoked** numerical:

TUDETCUIUSIS	C	<ul><li><iiissing< li=""></iiissing<></li></ul>	,	,	,	,		
TuberculosisOr	С	<missing></missing>	yes	yes	no	yes	47	6
X_ray	С	<missing></missing>	yes	yes	yes	no	57	4
Bronchitis	С	<missing></missing>	yes	yes	yes	yes	71	10
Dyspnea	С	<missing></missing>	yes	yes	yes	yes	61	6
Age	N	<missing></missing>	yes	yes	no	yes	53	0
NrSmoked	<b>√</b>	C hissing>	yes	yes	no	yes	48	4
	(	0	yes	yes	yes	yes	72	9
		N	yes	yes	yes	yes	46	8
	- hg		yes	yes	yes	yes	61	7

If you want to remove variables from the analysis, right-click on the column headers of these variables to highlight them in red. If you change your mind about removing a variable, right-click again on the selected column header. Press button Clear Highlighted to remove the variables highlighted in red:

DIONE - Decision I										
About DIONE	Open	Data File	Load Data	Save Dat	a Bayes	sian Network:	Algorithm	ns Le	earn Structure	Fit Parameters
Dashboard Exp	olore	Graphical Mod	el							
Variables 🗸	Select A	di	Data Clea	rHighlighted	Clear All	Show Cases:	All O Con	nplete 🔵 I	Incomplete 12	99 cases for 10 va
<filter></filter>	<m< th=""><th>issing&gt;</th><th>Smoker</th><th>LungCancer</th><th>VisitToTubia</th><th>Tuberculosis</th><th>Tuberculosis</th><th>X_ray</th><th>Bronchitis</th><th>Dyspnea</th></m<>	issing>	Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis	X_ray	Bronchitis	Dyspnea
Name	Туре	Missing					OrCancer	1014		-,-,
Smoker	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes
LungCancer	С	<missing></missing>	no	yes	no	no	yes	yes	no	no
VisitToTubia	С	<missing></missing>	yes	yes	no	no	yes	yes	no	yes
Tuberculosis	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes
TuberculosisOrCan	С	<missing></missing>	yes	yes	no	no	yes	yes	no	yes
X_ray	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	no
Bronchitis	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes
Dyspnea	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes
Age	N	<missing></missing>	no	yes	no	no	yes	yes	no	yes
NrSmoked	С	<missing></missing>	yes	yes	no	no	yes	yes	no	yes
			yes	yes	no	no	yes	yes	yes	yes
			yes	yes	no	no	yes	yes	yes	yes

These variables are removed from the **Data** table and will not be used for the Bayesian network analysis but are still loaded. They can be added again to the **Data** table, if desired, by selecting them in the **Variables** table and pressing the Enter key. You can also press button **Clear All** to remove all variables from the **Data** table and start again selecting variables form the **Variables** table.

You can also remove variables from the **Variables** table by selecting them and pressing the Delete key to unload them. This way it is easier to remove many variables at a time, but be aware you can <u>no longer access</u> these variables in the current session.

## **Missing Values**

The default code for missing values is <missing>. If your data have a special code for missing values, you can set that code in the <missing> text box by typing the desired code, pressing the Enter key and confirming the recoding in the next dialog, as in this example with data from file healthdemomissing1.csv:

DIONE - Decision	Intellige	nce for Organi	zations in N	etwo	rk Environments	: data i	file C:\U	sers\hj_vv\Docu	uments\Dione\c		
About DIONE	Open	Data File	Load Dat	oad Data Save Data		a	Bayes	Recode Missing			
						Dujot	Are you sure you want to continue?				
Variables 🗸	Select A		Data	Clea	r Highlighted	Clear	All	This will recode these	e values to <missing>: # Yes</missing>		
<filter></filter>	#NU	JLL! I	Smok	er	LungCancer	VisitT	oTubia	Tuberculosis	Iuperculosis		
Name	Туре	Missing							OrCancer		
Smoker	С	<missing></missing>	yes		yes	no		no	yes		
LungCancer	С	<missing></missing>	no		no	no		no	no		
VisitToTubia	С	<missing></missing>	#NULL!		no	no		no	no		
Tuberculosis	С	<missing></missing>	no		no	no		no	no		
TuberculosisOrCan	С	<missing></missing>	yes		no	#NULI	.!	no	no		
X-rav	C	<missina></missina>	yes		no	no		no	no		

For some data more than one value is considered to represent a missing value. You can set several values as missing value by entering them in the <missing> field, separated by commas:

DIONE - Decisio	n Intellige	nce for Orgai	niza	itions in N	etwor	k Environment	
About DIONE	Open	Data File	Load Dat	ta	Save Da		
Dashboard E	Explore	Graphical Mo	del				
Variables 🗸	Select A		Data	Clear	r Highlighted		
<filter></filter>	#NU	μιι , ΝιήΙ		Smol	(er	LungCancer	
Name	Туре	Missing					
Smoker	C	<missina></missina>		yes		yes	

You can also set a code for a specific variable in the **Variables** table row for that variable, for example NIL for variable **LungCancer** instead of #NULL! for all other variables here:

DIONE - Decision I	ntellige	nce for Organiza	tions in N	letwor	k Environments	: : data file C:\L	Jsers\hj_v\
About DIONE	Open	Data File	Load Da	ta	Save Dat	a Baye	sian Net
Dashboard Exp	lore	Graphical Model					
Variables 🗸	Select A	II	Data	Clea	r Highlighted	Clear All	Show Ca
<filter></filter>	er> #NULL!		Smol	(or	LungCancer	VisitToTubia	Tuberc
Name	Туре	Missing	Smor	(er	Lungcancer	VISICIOTUDIa	Tuberc
Smoker	С	#NULL!	yes		yes	no	no
LungCancer	С	ліц т	no		no	no	no
Mainta Tabia	C		#NULL!		no	no	no
VisitToTubia	~	#NULL!	no		no	no	no
Tuberculosis	С	#NULL!	yes		no	#NULL!	no

#### **Complete and Incomplete Cases**

Each row of loaded data contains a specific combination of values for the variables that are selected for analysis. We call such a combination of values a case. For many analyses, including common Bayesian network structure learning algorithms, it is necessary to have only complete cases. i.e. cases with no missing values for any of the selected variables. In **DIONE** it is easy to inspect complete and incomplete cases for a set of selected variables.

For example, with the data from file healthdemomissing.csv and missing value code set to #NULL!, check radio button Incomplete :

About DIONE	Open	Data File	Load Data	Save Da	ita Baye	esian Networ	k: Algorith	ms L	earn Structure	Fit Paramete
Dashboard Exp	olore	Graphical Mod	el							
Variables 🗸 🗸	Select A	П	Data Clea	r Highlighted	Clear All	Show Cases:	All Com	nplete 💿 l	ncomplete 344	cases for 8 va
<filter></filter>	#NU	JLL!	Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis OrCancer	X-ray	Bronchitis	Dyspnea
Name	Туре	Missing		7			oreancer			
Smoker	С	#NULL!	<missing></missing>	no	no	no	no	no	yes	yes
LungCancer	С	#NULL!	no	no	no	no	no	<missing></missing>	yes	yes
VisitToTubia	С	#NULL!	yes	no	<missing></missing>	no	no	no	yes	no
Tuberculosis	С	#NULL!	no	no	no	no	<missing></missing>	no	no	no
TuberculosisOrCan	С	#NULL!	yes	no	<missing></missing>	no	no	no	yes	yes
X-ray	с	#NULL!	no	<missing></missing>	no	no	no	no	no	yes
Bronchitis	с	#NULL!	yes	no	no	no	<missing></missing>	no	no	no
Dyspnea	С	#NULL!	yes	<missing></missing>	no	no	no	no	yes	yes
bysprice	~	WINDEL:	no	no	no	no	no	no	no	<missina></missina>

The **Data** table now shows the cases with missing values, with missing value codes replaced by the standard code <missing> used by **DIONE**. To give more insight in numbers of missing values of all variables, the variables in the **Variables** table are colour-coded in shades of red: darker red means a higher percentage of missing values, so in this example variable **VisitToTubia** has the highest proportion of missing values. Values of a variable can be inspected by double-clicking in the variable's data column in the **Data** table, with radio button **All** selected to see all values:

<b>ata</b> Clear	Highlighted	Clear All S	how Cases: 🜘	) All 🔵 Com	plete 🔵
Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis OrCancer	X-ra
s	yes	no	no	yes	yes
)	no	no	no	no	yes
nissing>	no	no	no	no	no
)	no	no 🧟	no	no	<missing< td=""></missing<>
s	no	<missing></missing>	no	no	no
s	no	no	no	no	no
)	no	no	no	<missing></missing>	no

Variable Values	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage	
VisitToTubia					
no	60	15.96%	60	15.96%	
yes	23	6.12%	83	22.07%	
<missing></missing>	293	77.93%	376	100.00%	
Total	376	100.00%			
<					

To show complete cases, select radio button O Complete :

About DIONE	Ope	n Data File	Load Data	Save Da	ita Bay	esian Networ	k: Algorith	nms Le	arn Structure	Fit Paramete	
Dashboard Exp	plore	Graphical Model									
Variables 🗸	Select A	AII	Data Clea	ar Highlighted	Clear All	Show Cases:	All 💽 Con	nplete 🔵 In	complete 32	cases for 8 va	
<filter></filter>	#N	ULL!	Smoker	LungCancer	VisitToTubi	a Tuberculosis	Tuberculosis OrCancer	X-ray	Bronchitis	Dyspnea	
Name	Туре	Missing					oreancer				
Smoker	С	#NULL!	yes	yes	no	no	yes	yes	yes	yes	
LungCancer	С	#NULL!	no	no	no	no	no	yes	yes	yes	
VisitToTubia	С	#NULL!	yes	no	no	no	no	no	yes	yes	
Tuberculosis	С	#NULL!	no	no	no	no	no	no	yes	yes	
TuberculosisOrCan	С	#NULL!	yes	no	no	no	no	no	yes	yes	
X-ray	С	#NULL!	yes	no	no	no	no	no	yes	yes	
Bronchitis	С	#NULL!	yes	no	no	no	no	no	no	no	
Dyspnea	С	#NULL!	no	no	no	no	no	no	yes	yes	
			no	no	no	no	no	no	no	no	
			ves	no	no	no	no	no	no	no	

You can now also see details on variables for complete cases only by double-clicking in the variable's data column in the **Data** table with the **Complete** radio button selected:

Clear	Highlighted	Clear All S	how Cases:	) All 💿 Comp	olete 🔵 Inco	mplete
oker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis OrCancer	X-ray	Bronc
	yes	no	no	yes	yes	yes
	no	no	no	no	yes	yes
	no	no 🦻	no	no	no	yes
	no	no	no	no	no	yes
	no	no	no	no	no	yes
	no	no	no	no	no	yes
	no	no	no	no	no	no

Variable Values	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage	
VisitToTubia					
no	32	8.51%	32	8.51%	
Total	32	8.51%			
					~
					>

Please note that, for numerical variables that have some nonnumerical values, the nonnumerical values will de recoded to missing when selecting radio button Complete or

Incomplete

## **Explore Data**

Before proceeding to sophisticated data analysis, it is usually a good idea to explore the data with relatively simple tools to develop a better understanding and intuition of their characteristics. With **DIONE** it is simple to explore data by looking at one variable at a time or by comparing two variables (univariate and bivariate methods): frequency tables, bar charts, histograms, scatter plots and calculation of statistics such as mean, median, mode, standard deviation, chi-square tests and risk ratios with p-values, as well as Bayes factors and correlations. To make it easy to decide which variables are most interesting for multivariate analysis, it is also possible to obtain bivariate comparisons between a variable of interest and all other variables and see in one table which variables are likely to be related to the variable of interest. The variable of interest will in many cases be an outcome variable for which you would like to know causes.

#### **One Variable**

Select a variable in the **Data** table, by clicking anywhere in its data column, and press button **Calculate Frequencies**, press the Enter key or double-click to see a bar chart and/or histogram and a frequency table of this variable's data at the right of the screen, for example variable **Smoker** with possible values yes and no:

About DIONE	Open	Data File	Load Data	Save Da	a Bayes	aian Network	: Options		Learn Structure	Fit Parameters	Infer Pr	obabilities Stop A	lgorithm	Save Results	frequenc	cy calculation	ns co
Dashboard Exp	plore	Graphical Mode	1														
ariables 🗸 S	elect All		Data Clear	r Highlighted	Clear All S	how Cases: 🜘	All 🔵 Com	olete 🔵	Incomplete 129	ecases for 10 vari	iables	Cale	ulate Frequenc	ies Calc	ulate Bivariate 1	lests	
<filter></filter>	missing>		Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis OrCancer	X_ra	ay Bronchitis	Dyspnea	Age	1.0	Va	riable: Smol	(er		
Name	Туре	Missing									~	0.9					
moker	С	<missing></missing>	yes	yes		no		yes	yes	yes		0.8 0.7 0.6 0.5 0.5					
ungCancer	С	<missing></missing>	no	yes		no		yes	no	no		0.0 Jan					
/isitToTubia	С	<missing></missing>	yes	yes		no		yes	no	yes		80.5					
uberculosis	С	<missing></missing>	yes	yes		no		yes	yes	yes		0.00					
uberculosisOrCan	С	<missing></missing>	yes	yes		no		yes	no	yes		E 0.2					
<_ray	С	<missing></missing>	yes	yes		no		yes	yes	no		0.1					
Bronchitis	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes			0.2 0.3		0.6 0.7	0.8 0.9	1.0
Dyspnea	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes				value			
Age	Ν	<missing></missing>	no	yes	no	no	yes	yes	no	yes				riable: Smol	er		
NrSmoked	С	<missing></missing>	yes	yes	no	no	yes	yes	no	yes		600	000			646	
			yes	yes	no	no	yes	yes	yes	yes		SE 500					
			yes	yes	no	no	yes	yes	yes	yes		005 005 005 007 005 005 005 005					
			yes	yes	no	no	yes	yes	yes	yes		300 Bet					
			yes	yes	no	yes	yes	yes	no	yes		<b>6</b> 200					
			no	yes	no	no	yes	yes	no	yes		≥ <sub>100</sub>					
			yes	yes	no	no	yes	yes	no	yes		0					-
			no	yes	no	no	yes	yes	no	no				Smoker categor			
			yes	yes	no	no	yes	yes	yes	yes				no yes	У		
			yes	yes	no	no	yes	yes	yes	yes							
			yes	yes	no	no	yes	yes	no	yes		Variable			Cumulative	Cumulative	
			yes	yes	no	no	yes	yes	yes	yes		Values	Frequency	Percentage	Frequency	Percentage	
			yes	yes	no	no	yes	yes	no	yes		Values					
			yes	yes	no	no	yes	yes	yes	yes		Smoker					
			yes	yes	no	no	yes	yes	yes	yes		no	653	50.27%	653	50.27%	
			yes	yes	no	no	yes	yes	yes	no		yes	646	49.73%	1,299	100.00%	1
			yes	yes	no	no		yes	no	yes		Tit	1.000	100.0001			+
			yes	yes		no		yes	yes	yes		Total	1,299	100.00%			+
			yes	yes		no		yes	no	yes							+
			20	,			y-2	,			× ×						

For categorical or ordinal variables, a bar chart is shown. For numerical variables a histogram is shown and under the frequency data in the frequency table their mean, median and standard deviation are shown.

This example from the same data set also shows a histogram for a numerical variable: **Age** (a bar chart is also shown to visualize the raw data):

About DIONE	Open	Data File	Load Data	Save Data	Bayesia	n Network:	Options	Learn S	tructure F	Fit Parameters	Infer Pro	obabilities Stop A	Algorithm	Save Results	frequen	cy calculatio	ns c
Dealthead						-										- -	
		Graphical Mode															
ariables 🗸 🤄	Select All		Data Clear Hig	ghlighted Cl	ear All Sho	w Cases: 💿 A	II 🔵 Compl	lete 🔵 Incomp	olete 1299 o	cases for 10 variabl	les	Cal	culate Frequenc	ies Calc	ulate Bivariate	Tests	
<filter></filter>	missing	•	1	Minist - Tubin	Technologia	Tuberculosis	¥	Duomobilitio	Dummer	0	Netwo			Variable: Age	Ð		
Name	Туре	Missing	LungCancer	VisitToTubia	Tuberculosis	OrCancer	X_ray	Bronchitis	Dyspnea	Age	NrSm	350		350			
noker	С	<missing></missing>	yes	no	no	yes	yes	yes	yes	53	$\hat{}$	S 300 -	2	273	49		
ngCancer	С	<missing></missing>	yes	no	no	yes	yes	no	no	62		250 - 200 -					
sitToTubia	С	<missing></missing>	yes	no	no	yes	yes	no	yes	62							
berculosis	С	<missing></missing>	yes	no	no	yes	yes	yes	yes	74		Ö 150	142		137		
berculosisOrCan	С	<missing></missing>	yes	no	no	yes	yes	no	yes	47		E 100	50		6	8	
ray	С	<missing></missing>	yes	no	no	yes	yes	yes	no	57		50 11				18	
onchitis	С	<missing></missing>	yes	no	no	yes	yes	yes	yes	71		5 10 1	5 20 25 30	35 40 45	50 55 60	65 70 75	80
spnea	С	<missing></missing>	yes	no	no	yes	yes	yes	yes	61				value			
le	N	<missing></missing>	yes	no	no	yes	yes	no	yes	53			,	Variable: Ag	Ð		
Smoked	с	<missing></missing>	yes	no	no	yes	yes	no	yes	48		350		550			
			yes	no	no	yes	yes	yes	yes	72		SU 300	2	.73 2	49		
			yes	no	no	yes	yes	yes	yes	46		250 - 200 -					
			yes	no	no	yes	yes	yes	yes	61		<b>SQ</b> 150	142		137		
			yes	no	yes	yes	yes	no	yes	62		5 100			6	8	
			yes	no	no	yes	yes	no	yes	41		E 50 11	30			18	1
			yes	no	no	yes	yes	no	yes	35		0		Age			_
			yes	no	no	yes	yes	no	no	40				categor	У		
			yes	no	no	yes	yes	yes	yes	56		<b>[</b> 6.0 - 13.7) <b>[</b> 13.7 - 2.			[36.8 - 44.5) 🔳 [4	14.5 - 52.2) = [52	.2 - 5
			yes	no	no	yes	yes	yes	yes	37		<b>[</b> 59.9 - 67.6] <b>[</b> 67.6 - 3	75.3) = [75.3 - 83.0				_
			yes	no	no	yes	yes	no	yes	62		Variable					
			yes	no		yes	yes	yes	yes	60			Frequency	Percentage	Cumulative Frequency	Cumulative Percentage	
			yes	no	no	yes	yes	no	yes	58		Values			requency	recentage	
			yes	no		yes	yes	yes	yes	41		Age					f
			yes	no	no	yes	yes	yes	yes	57		[6.0 - 13.7)	11	0.85%	11	0.85%	T
			yes	no		yes	yes	yes	no	52		[13.7 - 21.4)	50	3.85%	61	4.70%	
			ves	no	no	yes	yes	no	yes	65		[21.4 - 29.1)	142	10.93%	203	15.63%	-
										60		[29.1 - 36.8)	273	21.02%	476	36.64%	-
			yes	no	no	yes	yes	yes	yes	54	- 1	[36.8 - 44.5]	350	26.94%	826	63.59%	-
			yes	no	no	yes	yes	no	yes	54	$\sim$	[44.5 - 52.2)	249	19.17%	1,075	82.76%	

#### **Time Series Data**

It is also possible to analyse time series data, including smoothing of a time series by calculating a moving average and decomposition of a time series in trend, periodic and random components. As an example we analyse these data on the age of death of successive kings of England by selecting variable **Age Death Kings** in the **Data** table and clicking button **Analyse Time Series**:

DIONE - Decision	n Intelligen	ce for Organiz	ations in Networl	k Environments : da	ta file C:\Users\hj_vv\Docun	nents\Dione\data\ki	ngsmissing.csv						- 🗆 X
About DIONE	Open E	Data File	Load Data	Save Data	Bayesian Network:	Algorithms	Learn Structure	Fit Parar	meters	Infer Probabilities	Stop Algorithm	Save Results	data loaded: 42 new data for 2
Dashboard E	xplore G	Fraphical Model											
Variables 🗸	Select All		Data Clear	Highlighted	ear All Show Cases:	All Complete	Incomplete	42 case		Calculate Frequer	Calcula	te Bivariate Tests	Analyse Time Series
<filter></filter>	<mis< th=""><th>ssing&gt;</th><th>Nr</th><th>Age Death</th><th></th><th></th><th></th><th></th><th></th><th></th><th>Hi</th><th>stogram</th><th></th></mis<>	ssing>	Nr	Age Death							Hi	stogram	
Name	Туре	Missing	1	60				Â	1.0				
Nr	N		2	43					8.0 tions				
Age Death Kings	N		3	67					0.7 Adti				
			4	50					10.5 0.4				
			5	56					9 0.4 0.3				
			6	42					È 0.2				
			7	50					0.1				
			8	65					0.0				

This shows the following dialog. To smooth the time series, set the Smoothing Span parameter, which is the number of values to use for a moving average calculation, form

example 7 here:
-----------------

Time Series Options	×
Smoothing	Decomposition
Smoothing Span 1	Frequency
Smooth Time Series	Decompose Time Series

The result is shown here, with data points as dots, the moving average as a blue line and a regression line as a black line:

DIONE - Decision	Intellige	nce fo	r Organizations	in Network Env	ironments : da	ta file C:\Users\hj	vv\Docume	nts\Dione\data\l	kingsmissing.csv	v								-		k
About DIONE	Oper	n Data	File Loa	id Data	Save Data	Bayesian	letwork:	Algorithms	Learn Str	ructure F	Fit Paran	neters	Infer Probabilities	Stop Algori	thm	Save Result	s a	nalysis comp	leted	
Dashboard Ex	plore	Graph	ical Model																	
Variables 🗸	Select A	di	Data Clea	r Highlighted	Clear All	Show Cases:		mplete 🔵 Inco	omplete 42 o	cases for 4 va	a		Calculate Frequen	cies Cal	ulate Biva	ariate Tests	Analys	e Time Series		
<filter></filter>	<missi< td=""><td>ng&gt;</td><td></td><td></td><td>Nr</td><td>Age Death</td><td></td><td></td><td></td><td></td><td></td><td></td><td>Age</td><td>Death Kings</td><td>_timeste</td><td>ep - Age Dea</td><td>th Kings</td><td></td><td></td><td></td></missi<>	ng>			Nr	Age Death							Age	Death Kings	_timeste	ep - Age Dea	th Kings			
Name	Туре	М	Kings_smo	Kings_time	INI	Kings						80				•				
Age Death Kings_ti	N		NaN	1		60					- Â	<u>دم</u> 70				•			$\overline{}$	l
Age Death Kings_s	N		NaN	2	1	2 43					, in the second s	s 70 · •	· · · · · · · · · · · · · · · · · · ·	·		/	$\sim$			l
Nr	N		NaN	3	3	3 67					4	50	• •			· ./	•••			l
Age Death Kings	N		NaN	4	4	4 50						2 40	•	• `	$\sim$	·/				l
			NaN	5	1	5 56					4	<b>8</b> 30								l
			NaN	6	(	5 42						20			•					l
			52.5714285		1	7 50						10	4.1 8.2	12.3 16.4	20.5	24.6 2	28.7 3	2.8 36.9	41	Ì
			53.2857142			B 65								Age	Death K	ings_timestep				
			56.8571428	. 9	9	68									Bar Ch	nart				
			56.7857142			o <missing></missing>						1								l
			58.9285714	. 11	1	1 65														l
			55.7857142			2 34						suor								l
												2								4

To decompose a time series, set the Frequency parameter, which is the number of values assumed for one period of the time series, for example 12 here for a dataset of monthly births in New York:

Time Series Options	×
Smoothing	Decomposition
Smoothing Span	Frequency 12
Smooth Time Series	Decompose Time Series

The result is shown here, with the top chart showing again data points as dots, the trend component as a blue line and a regression line as a black line, and the bottom chart showing the random component as dots and the periodic component as a blue line:

About DIONE	Open Da	ata File	Load Data	Sa	ave Data	Bayesian	Network:	Algorithms	Learn S	tructure	Fit Parameters Inf	er Probabilities	Stop Algorit	thm Save Results	analysis comple	ted
Dashboard [	Explore Gra	phical Mode	1			I								l	I	
Variables	Select All	Data	Clear Highlig	ted Cl	lear All	Show Cases:	All 🔵 Con	nplete 🔵 Incom	nplete 1	58 cases f	Calc	ulate Frequencies	Calcula	te Bivariate Tests Ana	lyse Time Series	
<filter></filter>	<missing></missing>	births_	rand births_	peri bir	rths_trend0	births_time	births						births_ti	mestep - births		
Name	Туре	M NaN	-0.677	1947 Na	aN	1	26.663			ĉ	30 - 29 -					3
pirths_timestep	N	NaN	-2.082	9607 Na	aN	2	23.598				28				· · ··································	
pirths_random0	N	NaN	0.8625	2323 Na	aN	3	26.931				27 · • • 26 · •		•	· · · · · · · · · · · · · · · · · · ·		
oirths_periodic0	N	NaN	-0.801	5786 Na	aN	4	24.74				se 26 25 24		متغنيها		ts Analyse Time Series the Analyse Time Series Analyse Time Series A	
pirths_trend0	N	NaN	0.2516	5144 Na	aN	5	25.806				24	·······				
pirths	N	NaN	-0.153	2556 Na	aN	6					22 21		• •			
		-0.963	3790 1.4560	1567 23	3.9843333	7	24477				20					
		-0.925	7187 1.164	59375 23	3.6621250	8	23.901				0 16.7	33.4 50.1			133.6 150.3	167
		-0.939	9495 0.6916	1618 23	3.42333333	9	23.175						b	pirths_timestep		
		-0.709	3693 0.7752	1439	23.161125	10	23.227					bi	rths_timest	ep - births_periodic0		
		-0.082	48471.109	7652 22	2.8642500	11	21.672				5-					
		-0.298	38860.376	3197 22	2.5452083	12	21.87				e 3					
		-0.237	30520.677	1947 22	2.3534999	13	21.439						N N		. N. +N N	N
		0.8632	52402.082	9607 22	2.3087083	14	21.089				ber of the ber	the full t	17.1041			4
		0.5438	9342 0.8625	2323 22	2.3025833	15	23.709				string 2	$\gamma$ , $\gamma$ , $\gamma$ , $\gamma$ , $\gamma$	W W	N. W. W. W.	, M. M. M.	• V
		0.1758	87010.801	5786 22	2.2947916	16	21.669				ig -3					
		-0.793	1931 0.2516	5144 22	2.2935416	17	21.752				-4-					
		-1.391	36930.153	2556	22.305625	18	20.761				0 16.7	33.4 50.1	66.8	83.5 100.2 116.9	133.6 150.3	167
		-0.311	8790 1.4560	4567 22	2.3348333	19	23.479							irths_timestep		
		0.3477	3958 1.164	59375 22	2.3116666	20	23.824									
		0.1505	9214 0.6916	1618 22	2.2627916	21	23.105				Variable	births_timester				
		0.0767	9727 0.7752	1439 22	2.2579583	22	23.11				Values	births_timestep	)			
		0.5910	98551.109	7652 22	2.27766666	23	21.759									
		0.0958	19710.376	3197	22.354	24	22.073				births_random0 [-1.81.5)	[1.0 - 1]	0.00%	[17.7 - 34.4) 0 0.00%		0.00%
		0.1838	19710.677	1947	22.430375	25	21.937				[-1.81.5]	0	0.00%			0.00%
		-0.318	70592.082	9607 22	2.4366666	26	20.035				[-1.10.8]	4	23.53%			0.00%
		0.3402	6842 0.8625	2323 22	2.3872083	27	23.59				[-0.80.4)	1	5.88%	1 5.88%	5 2	9.419
		0.1212	62010.801	5786 22	2.3524166	28	21.672				[-0.40.1)	2	11.76%	3 17.65%	2 1	1.76%
											[-0.1 - 0.2]	2	11.76%	7 41.18%		5.299

#### **Two Variables**

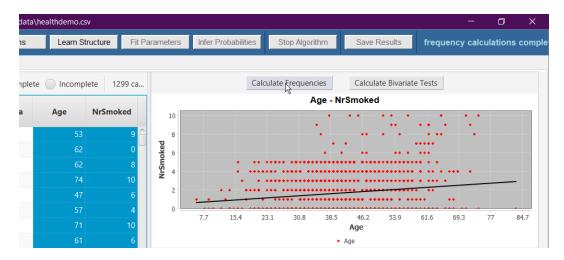
To compare two variables, select any pair of variables by clicking in the first variable's data column and using Ctrl-click anywhere in the second variable's data column: here **Smoker** (yes/no) and **LungCancer** (yes/no). After pressing button
Calculate Frequencies
or the Enter key, results are shown in the bar chart and in a two-way frequency table, with
a preview in the
Dashboard
tab and a more detailed view in the
Explore
tab.

In the frequency table, the chi square p-value indicates the significance of the relation between the variables (the lower the p-value, the more significant). The risk ratio is also calculated with corresponding confidence interval and p-values. The risk ratio (or relative risk) compares the risks for two groups by dividing the incidence in the exposed group by the incidence in the unexposed group. Thus, in the example, the percentage of smokers with lung cancer is 12.13 times the percentage of non-smokers with lung cancer. The risk ratio p-values show that this risk ratio value is significant, and the 95 % confidence interval means that the value is between 4.90 and 30.01 with 95 % probability.

About DIONE	Oper	n Data File	Load Data	Save Dat	a Bayes	ian Network:	Algorithm	ns Lea	arn Structure	Fit Parameters In	fer Probabilities	Stop Algorit	hm Sa	ive Results	frequency c	alculations cor
Dashboard Ex	plore	Graphical Mode	4													
Variables 🗸	Select /	All	Data Clea	ar Highlighted	Clear All	Show Cases:	All 🔵 Com	nplete 🔵 In	complete		Calcul	ate Frequencies	Calcul	late Bivariate T	ests	
<filter></filter>	<n< th=""><th>nissing&gt;</th><th></th><th></th><th></th><th></th><th>Tuberculosis</th><th></th><th></th><th></th><th></th><th>ŀ</th><th>listogram</th><th></th><th></th><th></th></n<>	nissing>					Tuberculosis					ŀ	listogram			
Name	Туре	Missing	Smoker	LungCancer	VisitToTubia	Tuperculosis	OrCancer	X_ray	Bronchitis	1.0						
Smoker	С	<missing></missing>	yes		no	no	yes	yes	yes	2 0.8						
LungCancer	С	<missing></missing>	no		no	no	yes	yes	no	0.8 0.7 0.6 0.5 0.5						
VisitToTubia	С	<missing></missing>	yes		no	no	yes	yes	no	0.5						
Tuberculosis	С	<missing></missing>	yes		no	no	yes	yes	yes	<b>10</b> 0.4 <b>10</b> 0.3						
TuberculosisOrCan	С	<missing></missing>	yes		no	no	yes	yes	no	E 0.2						
X_ray	С	<missing></missing>	yes		no	no	yes	yes	yes	0.1						
Bronchitis	С	<missing></missing>	yes		no	no	yes	yes	yes	0.0						
Dyspnea	С	<missing></missing>	yes		no	no	yes	yes	yes				value			
Age	N	<missing></missing>	no		no	no	yes	yes	no		c	omparison:	Smoker -> L	ungCancer		
NrSmoked	с	<missing></missing>	yes	yes	no	no	yes	yes	no		40			586		
			yes	yes	no	no	yes	yes	yes	600 ·						
			yes		no	no	yes	yes	yes	500 500 500 500 500 500 500 500 500 500						
			yes		no	no	yes	yes	yes	<b>S</b> 300						
			yes		no	yes	yes	yes	no	<b>jo</b> 200						
			no		no	no	yes	yes	no	È 100 -						60
			yes		no	no	yes	yes	no	0		5				
			no		no	no	yes	yes	no		no				yes	
			yes		no	no	yes	yes	yes			LungCon	Smoker ter no = LungCan	COT 1/05		
			yes		no	no	yes	yes	yes				co no = cungCdh	nur yes		
			yes		no	no	yes	yes	no	Variable						
			yes		no	no	yes	yes	yes		Smoker					
			yes		no	no	yes	yes	no	Values						
			yes		no	no	yes	yes	yes	LungCancer	n	0	ye	es	Tot	al
			yes		no	no	yes	yes	yes	no		99.23%	586	90.71%	1,234	95.00%
			yes		no	no	yes	yes	yes	yes	5	0.77%	60	9.29%	65	5.00%
			yes		no	no	yes	yes	no	Total	653	100.00%	646	100.00%	1,299	100.00%
			yes		no	no	yes	yes	yes	lotal	653	100.00%	646	100.00%	1,299	100.00%
			yes	yes	no	no	yes	yes	no			prol	hability that ve	riables are rela	ted	
<[		)>	<						>	<		pro				_

out DIONE	Open Data File	Load Data	Save Data	Bayesian Network:	Algorithms	Learn Structu	re Fit Par	ameters In	fer Probabilities	Stop Algo	rithm	Save Results	frequency	calculations
shboard	Explore Graphical Mo	del											1	
	45	Histogram			Variable									
					Vanable	Smoker								
I.9 ·					Values									
.8 -					LungCancer	no	)	у	es	То	tal			
0.7					no	648	99.23%	586	90.71%	1,234	95.00%			
0.6					yes	5	0.77%	60	9.29%	65	5.00%			
0.7 - 0.6 - 0.5 -					Total	653	100.00%	646	100.00%	1,299	100.00%			
1.4					Total	653	100.00%	040	100.00%	1,299	100.00%			
1.3							pro	bability that va	ariables are relat	ed		probability of	data if variables	are unrelated
.2						related with			th probability	Bayes	factor		ni square p-valu	
.1							1.000		0.000	2	.8659E10			0.000
									0504 6.1					
0						LungCance	risk ratio r % level yes /	% level no	95% confide lower	upper	midp.exact	risk ratio p-valu fisher.exact	e chi.square	
		value			Smoker level no	Lungcanee	r /o lever yes /	1.000	undefined	undefined	undefined	undefined	undefined	
odi Nur	nber of		Change	Recode Varia	Smoker level yes			12.130	4.903	30.009	0.000	0.000	0.000	
	Compariso	n: Smoker -> Lun	gCancer											
50 6	48		<u> </u>				risk ratio		95% confide			risk ratio p-valu		
i00 -			586		Smoker level no	LungCance	r % level no / 9	% level yes 1.000	lower undefined	upper undefined	midp.exact undefined	fisher.exact undefined	chi.square undefined	
i50 ·					Smoker level yes			0.914	0.891	0.938	0.000	0.000	0.000	
500														
150 · 1600 · 150 ·														
50														
00 - 50 -														
.00 -														
150 -														
50				60										
0	5													
	no	Smoker	ye	s										
	Luna	Smoker Cancer no E LungCancer	ves											

Frequencies can also be calculated for two numerical variables, for example **Age** and **NrSmoked**, after setting **NrSmoked** to **Type** numerical (N) in the **Variables** table. As shown here, they are compared in a scatter plot, including a linear line of best fit (regression line):

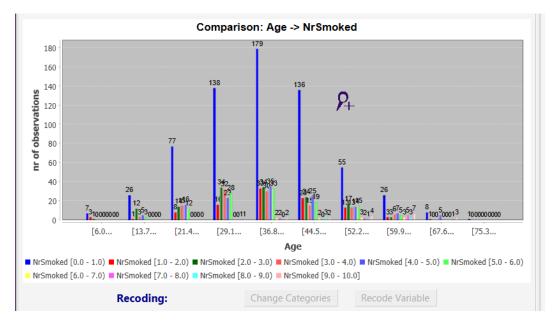


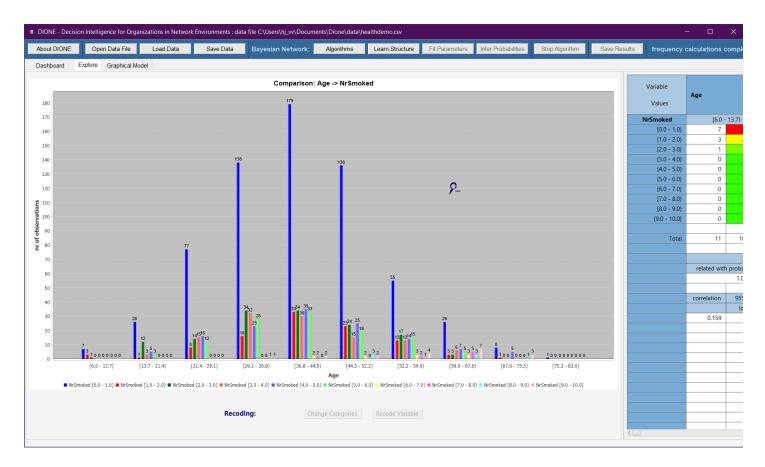
Their correlation coefficient is shown in the results table after the frequency data. Correlation coefficient 1 means perfect correlation and 0 means no correlation, so in the example there is a weak correlation, as is also clear from the scatter plot. Regression coefficients are also shown.



#### **Inspect Results**

A preview of results is shown in the Dashboard tab and a more detailed view in the **Explore** tab. Sometimes a variable has many levels and/or long texts as values, so the chart become hard to read. In the **Explore** tab you can zoom in on a chart by hovering over it with the mouse to show an hourglass with a + sign, then clicking to zoom in:





The hourglass now has a – sign to zoom out again.

At the right-hand side of the **Explore** tab, there is a table with detailed frequency data and other analysis results. For numerical variables, frequencies of ranges of values are shown. If desired, you can get more details about individual values of a variable by rightclicking on one of the value ranges in the left-hand column and choosing menu item **Inspect Data** to see the distinct values of that variable:

Variable Values		Frequer	су	Percentage	Cumulative Frequency	Cumulative Percentage	
Age							
[6.0 - 1	13.7)		11	0.85%	11	0.85%	
[13.7 - 2	21.4)		50	3.85%	61	4.70%	
[21.4 - 1	10 41		42	10.93%	203	15.63%	
[29.1 ·	Copy Se	election	73	21.02%	476	36.64%	
[36.8 -	Save Ta	ble	50	26.94%	826	63.59%	
[44.5 ·	Inspect	Data	49	19.17%	1,075	82.76%	
150.0	10.00		37	10.55%	1,212	93.30%	
[59.9 - 6	67.6)		68	5.23%	1,280	98.54%	
[67.6 - 7	75.3)		18	1.39%	1,298	99.92%	
[75.3 - 8	83.0]		1	0.08%	1,299	100.00%	
٦	Total	1,2	299	100.00%			

Variable Values	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage		Distinct Values	Î
Age						R	
[6.0 - 13.7)	11	0.85%	11	0.85%		28	
[13.7 - 21.4)	50	3.85%	61	4.70%		29	
[21.4 - 29.1)	142	10.93%	203	15.63%		24	
[29.1 - 36.8)	273	21.02%	476	36.64%		25	
[36.8 - 44.5)	350	26.94%	826	63.59%		23	
[44.5 - 52.2)	249	19.17%	1,075	82.76%		22	
[52.2 - 59.9)	137	10.55%	1,212	93.30%		26	
[59.9 - 67.6)	68	5.23%	1,280	98.54%		27	
[67.6 - 75.3)	18	1.39%	1,298	99.92%			
[75.3 - 83.0]	1	0.08%	1,299	100.00%			
Total	1,299	100.00%					
	me	ean					
		40.945					
	me	dian					
		40.000					
	ma	ode					
		35					
	standard	deviation					
		11.769					

#### **All Variables**

To obtain bivariate comparisons of one variable with other variables, start by selecting, in the **Dashboard** tab, one variable as the outcome variable of interest. In this example the outcome variable is **LungCancer**, with which all other loaded variables will be compared:

About DIONE	Open [	Data File	Load Data	Save Da	a Bayes	sian Network:	Options	; l	Learn Structure	Fit P	arameters Infer I	Probabilities	Stop Algorithm	Save Res	sults f	requency calculations co
Dashboard Ex	plore C	Graphical Mode	1													
Variables 🗸 🤄	Select All		Data Clear	Highlighted	Clear All	Show Cases: 🖲	All 🔵 Com	olete 🔵	Incomplete	1299 ca		Calcula	te Frequencies		Bivariate Te	ests
<filter> &lt;</filter>	missing>	Missing	Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis OrCancer	X_ray	/ Bronchi	tis D	1.0		Variable	LungCance	er <sup>°</sup>	
Smoker		<missing></missing>	yes	yes	no	no	yes	yes	yes	y∈≏	0.9 2 0.8					
ungCancer	c	<missing></missing>	no	yes	no	no	yes	yes	no	n	0.7					
/isitToTubia	c	<missing></missing>	yes	yes	no	no	yes	yes	no	ye	8.0 0.7 0.6 0.5 0.5 0.4					
uberculosis	c	<missing></missing>	yes	yes	no	no	yes	yes	yes	ye	0.4					
uberculosisOrCan		<missing></missing>	yes	yes	no	no	yes	yes	no	ye	0.3 0.3 0.2					
(_ray	C	<missing></missing>	yes	yes	no	no	yes	yes	yes	n	0.1					
ronchitis	c	<missing></missing>	yes	yes	no	no	yes	yes	yes	y€	0.0 0.05 0.	0 0.15 0.20 0.25	0.30 0.35 0.40 0.4	5 0.50 0.55 0.60	0 0.65 0.70	0.75 0.80 0.85 0.90 0.95 1.00
lyspnea	с	<missing></missing>	yes	yes	no	no	yes	yes	yes	y€				value		
ge	N	<missing></missing>	no	yes	no	no	yes	yes	no	ye			Variable	LungCance	er	
rSmoked	N	<missing></missing>	yes	yes	no	no	yes	yes	no	ye	1,200	1,234		-		
		-	yes	yes	no	no	yes	yes	yes	ye						
			yes	yes	no	no	yes	yes	yes	ує	008 Aatio					
			yes	yes	no	no	yes	yes	yes	y€	opservations 008 extra 000 -					
			yes	yes	no	yes	yes	yes	no	y€	<b>10 5</b> 400					
			no	yes	no	no	yes	yes	no	y€	È 200					65
			yes	yes	no	no	yes	yes	no	y€	0					65
			no	yes	no	no	yes	yes	no	n				LungCancer		
			yes	yes	no	no	yes	yes	yes	ye				category no sys		
			yes	yes	no	no	yes	yes	yes	ye						
			yes	yes	no	no	yes	yes	no	ye	Variable			Cumulative	Cumulativ	(A)
			yes	yes	no	no	yes	yes	yes	y€	Values	Frequency	Percentage	Frequency	Percentag	
			yes	yes	no	no	yes	yes	no	y€	values					
			yes	yes	no	no	yes	yes	yes	y€	LungCancer					
			yes	yes	no	no	yes	yes	yes	y€		1,234		1,234	95.00	
			yes	yes	no	no	yes	yes	yes	n	У	es 65	5.00%	1,299	100.00	%
			yes	yes	no	no	yes	yes	no	ye	To	al 1,299	100.00%			
			yes	yes	no	no	yes	yes	yes	ye	10	.,255	10010070			
			yes	yes	no	no	yes	yes	no	ye 🗸						

Press button Calculate Bivariate Tests above the plots (histogram and bar chart) at the right. Go to the Explore tab and scroll to the right in the results table to see the p-values of chi square tests of the selected outcome variable compared with all other loaded variables, ordered from most to least significant association. This allows you to see quickly which variables have the most significant association with the selected outcome variable and so are likely to be of interest for further analysis. In the example all variables except **Bronchitis** and **VisitToTubia** have significant associations at the 1 % (so also of course at the 5 %) significance level with LungCancer (p < 0.01 and p < 0.05).

Options	Learn Struc	ture Fit Pa	arameters	Infer Probabilitie:	s Stop Al	aorithm	Save Results	bivariat	te tests calcu	la
Variable Values	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage		of	variate Tests LungCancer d other variables	Chi Square p-value	Correlation	
LungCancer										
no	1,234	95.00%	1,234	95.00%		Tube	rculosisOrCancer	0.000	undefined	
yes	65	5.00%	1,299	100.00%			X_ray	0.000	undefined	
							Age	0.000	undefined	
Total	1,299	100.00%					NrSmoked	0.000	undefined	
							Smoker	0.000	undefined	
							Dyspnea	0.000	undefined	
							Tuberculosis	0.005	undefined	
	ma	ode					Bronchitis	0.381	undefined	
		no					VisitToTubia	0.595	undefined	

For numerical data correlations of all variables with the outcome variable are also calculated. In this case results are ordered first according to increasing p-value of the chi square tests, and variables with approximately equal p-values (i.e. with difference in p-values less than 0.01) are ordered according to decreasing correlation (i.e. higher to lower absolute value of the correlation coefficient). This example shows results for variable **NrSmoked**:

	Options	Learn Struc	ture Fit P	arameters	Infer Probabilities	Stop Algorithm	Save Results	bivaria	te tests calcı	ula
V	/ariable Values	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage		Bivariate Tests of NrSmoked and other variables	Chi Square p-value	Correlation	
Nr	Smoked									t
	[0.0 - 1.0)	653	50.27%	653	50.27%		Smoker	0.000	undefined	Т
	[1.0 - 2.0)	101	7.78%	754	58.04%		LungCancer	0.000	undefined	
	[2.0 - 3.0)	139	10.70%	893	68.75%	T	uberculosisOrCancer	0.000	undefined	
	[3.0 - 4.0)	114	8.78%	1,007	77.52%		X_ray	0.000	undefined	
	[4.0 - 5.0)	130	10.01%	1,137	87.53%		Bronchitis	0.000	undefined	
	[5.0 - 6.0)	115	8.85%	1,252	96.38%		Age	0.000	0.159	
	[6.0 - 7.0)	10	0.77%	1,262	97.15%		Dyspnea	0.000	undefined	
	[7.0 - 8.0)	9	0.69%	1,271	97.84%		Tuberculosis	0.002	undefined	
	[8.0 - 9.0)	9	0.69%	1,280	98.54%		VisitToTubia	0.778	undefined	
	[9.0 - 10.0]	19	1.46%	1,299	100.00%					+
	Total	1,299	100.00%							+
		m	ean							
			1.687							+
		me	dian							t
			0.000							t
										+
		m	ode 0							+
		standard	deviation							
			2.158							

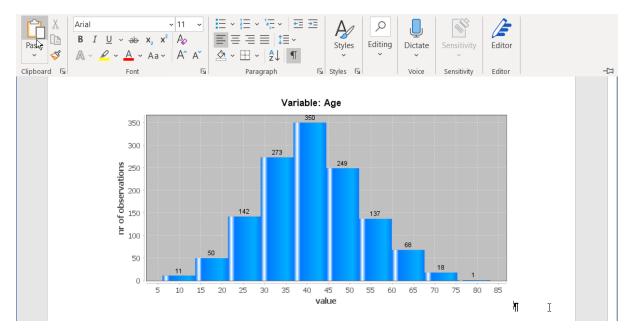
## **Copy and Save Results**

You can keep your results by copying charts and tables to another application, for example to insert them in a word processing document, or by saving them to a file.

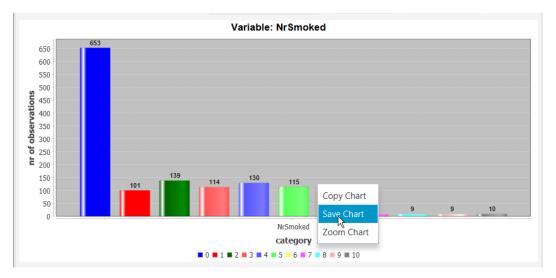
To copy a bar chart or a histogram, right-click on it and select **Copy Chart** from the pop-up menu to put a picture of the chart on your computer's clipboard:



You can now paste the picture in a Word document, for example:



To save a chart to a file, right-click on it and select **Save Chart** from the pop-up menu to open a save dialog:



Export to PNG							×
← → • ↑ <mark> </mark>	« Dione → results	~	Ö	Q	Search re	sults	
Organize 🔹 Ne	w folder					<b>•</b>	?
<ul> <li>Dione</li> <li>aktiviteiter</li> </ul>	1	^		No ite	ems match y	our search.	
		~					_
File name:	NrSmoked.png						~
Save as type:	Portable Network Graphics (*.png)						$\sim$
∧ Hide Folders					Save	Cancel	

You now have a Portable Network Graphics (PNG) picture file that you van open in any image processing program or insert in a document.

There is another option, **Zoom Chart**, in the pop-up menu that you can use to zoom in or out a chart as with the zoom hourglass, if you prefer.

You can copy results in the frequency table by right-clicking on a selection of data you want to copy and choosing Copy Selection in the pop-up menu:

k:	Options	Learn Stru	ucture Fit Pa	rameters	Infer Pr	robabilities	Sto	o Algo	rithm	ave Results	fre
			Variable Values	Frequenc	y Pe	ercentage	Cumula Freque		Cumulative Percentage		
			Age			Copy	election				
			[6.0 - 13.7)	1	1	Save Ta		11	0.85%		
			[13.7 - 21.4)	5	60			61	4.70%		
			[21.4 - 29.1)	14	12	Inspect	Data	203	15.63%		
			[29.1 - 36.8)	27	3	21.02%		476	36.64%		
			[36.8 - 44.5)	35	0			826	63.59%		
			[44.5 - 52.2)	24	9	19.17%	1	,075	82.76%		
			[52.2 - 59.9)	13	37	10.55%	1	,212	93.30%		
			[59.9 - 67.6)	6	58	5.23%	1	,280	98.54%		
18	1		[67.6 - 75.3)	1	8	1.39%	1	,298	99.92%		
70		85	[75.3 - 83.0]		1	0.08%	1	,299	100.00%		
70	75 80	80									
			Total	1,29	9	100.00%					

Copied data can be pasted in a Word document, for example, in the same way as described for a chart.

You can save the whole table in a text file by right-clicking anywhere in the table and choosing Save Table in the pop-up menu to bring up a save dialog:

		Variable Values	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage	
		Age					
		[6.0 - 13.7)	11	0.85%	11	0.85%	
		[13.7 - 21.4)	50	3.85%	61	4.70%	
		[21.4 - 29.1)	142	10.93%	203	15.63%	
Save Results File		(20.4 2C.0)	070	×	476	36.64%	
Save Results The				^ <mark>⊧%</mark>	826	63.59%	
$\leftarrow$ $\rightarrow$ $\checkmark$ $\uparrow$ $\blacksquare$ $\ll$ Dione $>$ results		・ じ	rch results	1%	1,075	82.76%	
				5%	1,212	93.30%	
Organize 🔻 New folder				- 170	1,280	98.54%	
1.0.1	^			1%	1,298	99.92%	
📌 Quick access				8%	1,299	100.00%	
🤜 This PC							
🧊 3D Objects				)%	C	py Selection	
E Desktop		allResults.txt	OECDonev	ar.txt			
Documents		unicourterett	02000000			ve Table	
Downloads					Ins	spect Data	
Music							
	*						
File name: Age				$\sim$			
Save as type: Result Files (*.txt)				$\sim$			
∧ Hide Folders		Save	e Ca	ncel:			

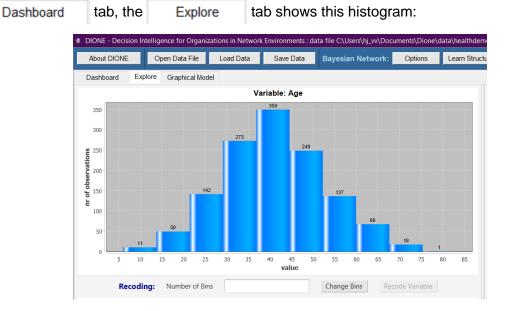
## **Recode Data**

When a variable has numerical data, sometimes more meaningful results can be obtained by recoding it to categorical data. An example is variable **Age** in the second example of the **Explore Data** section. You can recode such variables to categorical variables in the

**Explore** tab. You can also recode categorical variables to new categorical variables with different categories, typically to reduce the number of categories or to give categories meaningful names. Recoding a variable does not replace the existing variable but creates a new variable with a new name.

#### Numerical to Categorical

For example, we would like to recode the variable **Age**, which has values from 6 to 83 years old, to a variable with two categories (age groups). When **Age** is selected in the



Type the value 2 in the **Number of Bins** text field and press Enter to organize the **Age** data in two categories. The cut-off point between the two categories is automatically made:

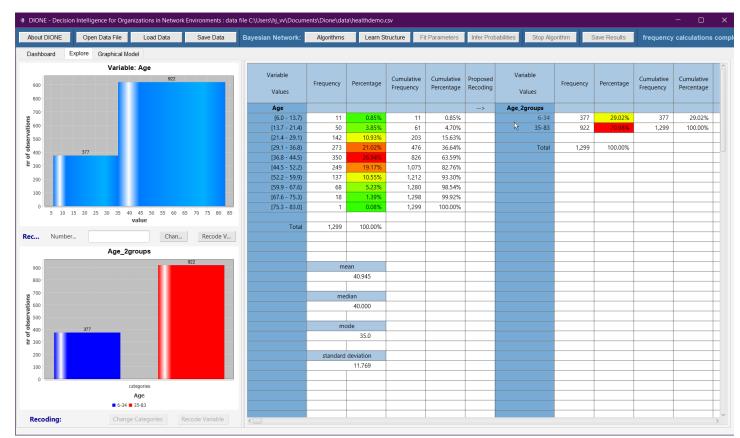


Category borders can be changed by pressing button Change Bins to open the recoding dialog, in which the recoded variable name and its category names can also be changed:

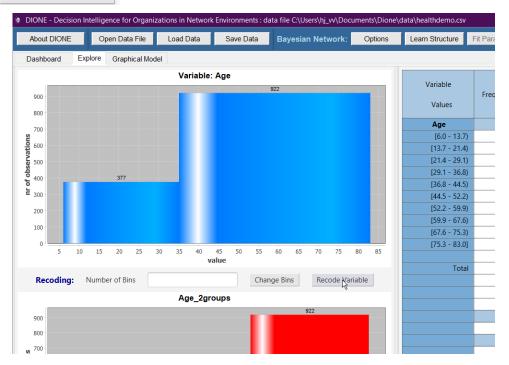
Recode Numerical Va	ariable				×				
Variable Name	Recoded Variable	Name							
Age	Age_recoded1								
Number of Bins	2	2							
Minimum	6.0	)							
Maximum	83.(	)							
Bins	>= start border	< er	nd border	cate	gory				
0110	6.0	44.5		6-44					
	44.5	83.0		45-83					
			🔞 Recode I	Numerical V	ariable				
			Variable	Name	Recoded V	ariable Nar	ne		
			Age		Age_2grou	DS			
End border values are	in next bin. Only ma	ximum v				P			
End numbers in propo	osed category names	are app	Number	- 6 Din -		2			
			Number	OT BINS		2			
			Minimum	n		6.0			
						83.0			
			Maximun	n		00.0			
			Bins		>= start bo	order	< end border	category	
			Diris		6.0	35		6-34.999	
					35	83.0	)	35-82.999	
					-				
									_
			End borde	r values ar	e in next bin. Oi	nly maximu	ım value is alwa	ays in last bin.	
			End numbe	ers in prop	osed category	names are	approximate.		
								Cance	
								Carrie	

In the **Recoded Variable Name** field a name is proposed for the new variable, which you can accept or change to a name you prefer. The new name cannot be the name of an existing variable; if the name is already in use, an error message is shown. In the **end border** column of the **Bins** table, type the desired cut-off value(s) (the cut-off value itself will be in the next higher bin), then press OK. Values in the table are easily edited by using arrow keys to change rows: for example, immediately after editing a cell, use the Down Arrow key to edit the cell below.

The proposed recoding preview (histogram, bar chart and frequency table) will now look like this:



When you are satisfied with the new categories, perform the actual recoding by pressing button Recode Variable :



This will create a new variable with the new name, with two categories (<35 and  $\geq$ 35) and show it in the **Dashboard** tab, where the new variable appears at the top of the **Variables** table and in the first column of the **Data** table.

Press button Calculate Frequencies , press the Enter key or double-click in the new variable column to see its data:

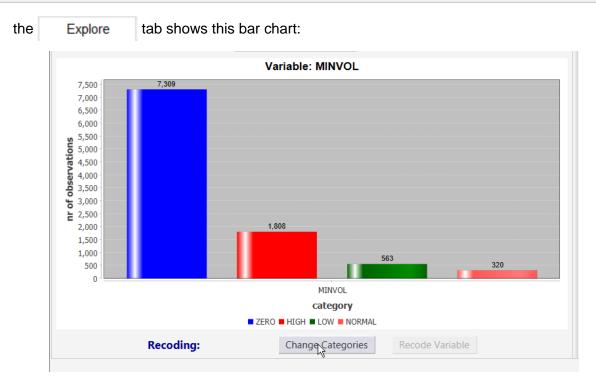
About DIONE		Data File	Load Data	Save Da		sers\hj_vv\Docu sian Network			Structure	Fit Parar	ifer Probab	ilitian	p Algorithm	Save Resul	10 Error	iencv calcula	
				Save Da	Bayes	sian Network	Algorithn	is Learr	Structure	Fit Paran	Iter Probat	Sto	p Algorithm	Save Resul	is frequ	iency calcula	itions con
		Graphical Mode															
Variables 🗸	Select A	1	Data Clear	Highlighted	Clear All	Show Cases: (	All Ocon	nplete 🔵 Inco	omplete	1299 cas		Calculate	Frequencies	Calculate	Bivariate Tests		
<filter></filter>	<m< th=""><th>issing&gt;</th><th>Age_2groups</th><th>Smoker</th><th>LungCancer</th><th>VisitToTubia</th><th>Tuberculosis</th><th>Tuberculosis OrCancer</th><th>X_ray</th><th>Bron</th><th>1.0</th><th></th><th>His</th><th>togram</th><th></th><th></th><th></th></m<>	issing>	Age_2groups	Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis OrCancer	X_ray	Bron	1.0		His	togram			
Name	Туре	Missing	35-83	yes	yes	no	no	yes	yes	yes 🄶	0.9						
Age_2groups	С	<missing></missing>	35-83	no	yes	no	no	yes	yes	no	0.8 0.7 0.6 0.5 0.5						
Smoker	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	no	0.6						
LungCancer	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes	80.5						
VisitToTubia	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	no							
Tuberculosis	С	<missing></missing>	35-83			no					<b>0.2</b>						
TuberculosisOrCan		<missing></missing>	35-83	yes yes	yes	no	no	yes	yes	yes	0.1						
X_ray	С	<missing></missing>	35-83		yes			yes	yes	yes	0			value			
Bronchitis	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes							
Dyspnea	C	<missing></missing>	35-83	no	yes	no	no	yes	yes	no			Variable:	Age_2group	S		
Age	N	<missing></missing>		yes	yes	no	no	yes	yes	no	900 -					022	
NrSmoked	N	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes	800 - 000 -						
			35-83	yes	yes	no	no	yes	yes	yes	600 ·						
			35-83	yes	yes	no	no	yes	yes	yes	500 400	377					
			35-83	yes	yes	no	yes	yes	yes	no	<b>5</b> 300						
			35-83	no	yes	no	no	yes	yes	no	≥ 200 100						
			35-83	yes	yes	no	no	yes	yes	no	0			Age_2groups			
			35-83	no	yes	no	no	yes	yes	no				Age_2groups			
			35-83	yes	yes	no	no	yes	yes	yes			6-3	34 = 35-83			
			35-83	yes	yes	no	no	yes	yes	yes							
			35-83	yes	yes	no	no	yes	yes	no	Variable	-		Cumulative	Cumulative		
			35-83	yes	yes	no	no	yes	yes	yes	Values	Frequency	Percentage	Frequency	Percentage		
			35-83	yes	yes	no	no	yes	yes	no							
			35-83	yes	yes	no	no	yes	yes	yes	Age_2groups						
			35-83	yes	yes	no	no	yes	yes	yes	6-34	377	29.02%	377	29.02%		
			35-83	yes	yes	no	no	yes	yes	yes	35-83	922	70.98%	1,299	100.00%		+
			35-83	yes	yes	no	no	yes	yes	no	Total	1,299	100.00%				+
			35-83	yes	yes	no	no	yes	yes	yes		1,255					+
			35-83	yes	yes	no	no	yes	yes	no							<u> </u>

### **Categorical to Categorical**

An example of recoding a categorical variable to a new categorical variable with different categories is variable **MINVOL** from data file alarm.csv. When **MINVOL** is selected in the

Dashboard tab and its frequencies are displayed ( Calculate Frequencies , Enter key or double-click),

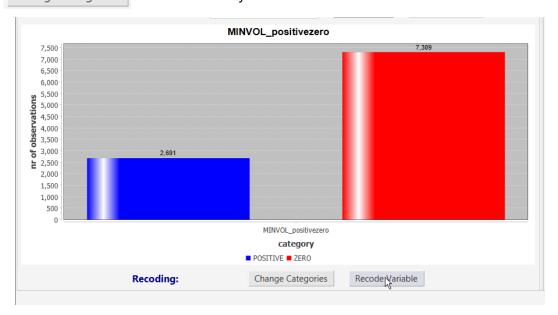
About DIONE		Open Dat	a File 🛛 L	bad Data	Save Data	Bayesian Netw	ork: Option	s Learn S	tructure	Fit Parameters	Infer Probabilities	Stop Algorithm	Save Re	sults	<b>100%</b> f	requency ca	lculations
Dashboard	Explor	re Grapi	nical Model														
/ariables	✓ Sele	ct All	Data Clea	r Highlighted	Clear All	Show Cases: 💿 A	II 🔵 Complete	Incomplete	e 10000	cases for 37 var		Calculate	eFrequencies		Calculate B	variate Tests	]
<filter></filter>	<mis< th=""><th>sing&gt;</th><th>/LA TI</th><th>PR EXPC</th><th></th><th></th><th>FIO2</th><th>PVSAT</th><th>SAO2</th><th>PAP</th><th></th><th></th><th>Variable</th><th>MINVOL</th><th></th><th></th><th></th></mis<>	sing>	/LA TI	PR EXPC			FIO2	PVSAT	SAO2	PAP			Variable	MINVOL			
name	type	missin	NORM	AL LOW	FALSE	HIGH	NORMAL	HIGH	HIGH		1.0						
ISTORY		<missi ^<="" td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>0.8 0.7 0.6</td><td></td><td></td><td></td><td></td><td></td><td></td></missi>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	0.8 0.7 0.6						
VP	С	<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>нібн</td><td>NORMAL</td><td>HIGH</td><td>HIGH</td><td>NORMAL</td><td>10.7 0.6</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	HIGH	LOW	FALSE	нібн	NORMAL	HIGH	HIGH	NORMAL	10.7 0.6						
CWP	С	<missi< td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>LOW</td><td>as 0.5 0.4</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	LOW	as 0.5 0.4						
YPOVOLEMIA		<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>LOW</td><td>5 o 2</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	HIGH	LOW	FALSE	NORMAL	NORMAL	LOW	LOW	LOW	5 o 2						
/EDVOLUME	С	<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>E 0.2</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	LOW	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	E 0.2						
/FAILURE		<missi< td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>0.1</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	0.1						
TROKEVOLUME	EC	<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>0.0 0.</td><td>.1 0.2 0.3</td><td>3 0.4</td><td>0.5 0.6</td><td>0.7</td><td>0.8 0.9</td><td>1.0</td></missi<>	LOW	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	0.0 0.	.1 0.2 0.3	3 0.4	0.5 0.6	0.7	0.8 0.9	1.0
RRLOWOUTP		<missi< td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>LOW</td><td></td><td></td><td></td><td>value</td><td></td><td></td><td></td></missi<>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	LOW				value			
RBP	С	<missi< td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td></td><td></td><td>Variable</td><td>: MINVOL</td><td></td><td></td><td></td></missi<>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL			Variable	: MINVOL			
REKG	С	<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>7,000</td><td>7,305</td><td></td><td></td><td></td><td></td><td></td></missi<>	HIGH	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	7,000	7,305					
RCAUTER		<missi< td=""><td>HIGH</td><td>NORMA</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>6,000</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	HIGH	NORMA	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	6,000						
RSAT	С	<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>нідн</td><td>NORMAL</td><td>HIGH</td><td>HIGH</td><td>NORMAL</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	LOW	LOW	FALSE	нідн	NORMAL	HIGH	HIGH	NORMAL							
ISUFFANESTH		<missi< td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>4,000</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	4,000						
APHYLAXIS		<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>NORMAL</td><td>NORMAL</td><td>NORMAL</td><td>0 3,000 0 2,000</td><td></td><td>1,808</td><td></td><td></td><td></td><td></td></missi<>	LOW	LOW	FALSE	ZERO	NORMAL	NORMAL	NORMAL	NORMAL	0 3,000 0 2,000		1,808				
PR	С	<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>≥ 1,000</td><td></td><td></td><td></td><td>563</td><td>320</td><td></td></missi<>	HIGH	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	≥ 1,000				563	320	
(PCO2	С	<missi< td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>0</td><td></td><td></td><td></td><td></td><td>520</td><td></td></missi<>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	0					520	
NKEDTUBE		<missi< td=""><td>NORM</td><td>AL LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td></td><td></td><td></td><td>MINVOL category</td><td></td><td></td><td></td></missi<>	NORM	AL LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL				MINVOL category			
INVOL	С	<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td></td><td></td><td>ZERO 📕 HIGH</td><td>LOW NORMA</td><td>L</td><td></td><td></td></missi<>	HIGH	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL			ZERO 📕 HIGH	LOW NORMA	L		
D2	С	<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>LOW</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	LOW	LOW	FALSE	LOW	NORMAL	LOW	LOW	NORMAL							
/SAT	С	<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>Variable</td><td></td><td></td><td>Cumulative</td><td>Cumulativ</td><td></td><td></td></missi<>	HIGH	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	Variable			Cumulative	Cumulativ		
102	С	<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>нідн</td><td>NORMAL</td><td>HIGH</td><td>HIGH</td><td>NORMAL</td><td>Values</td><td>Frequency</td><td>Percentage</td><td>Frequency</td><td>Percentage</td><td></td><td></td></missi<>	LOW	LOW	FALSE	нідн	NORMAL	HIGH	HIGH	NORMAL	Values	Frequency	Percentage	Frequency	Percentage		
ĄΡ	С	<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>values</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	LOW	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	values						
ILMEMBOLUS		<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>MINVOL</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	HIGH	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	MINVOL						
IUNT	С	<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>нідн</td><td>NORMAL</td><td>HIGH</td><td>HIGH</td><td>NORMAL</td><td>ZE</td><td></td><td>73.09%</td><td>7,309</td><td>73.099</td><td>_</td><td>_</td></missi<>	HIGH	LOW	FALSE	нідн	NORMAL	HIGH	HIGH	NORMAL	ZE		73.09%	7,309	73.099	_	_
TUBATION	С	<missi< td=""><td>NORM</td><td></td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>HIC</td><td></td><td>18.08% 5.63%</td><td>9,117 9,680</td><td>91.179</td><td></td><td></td></missi<>	NORM		FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	HIC		18.08% 5.63%	9,117 9,680	91.179		
RESS	С	<missi< td=""><td>LOW</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td>NORM</td><td></td><td>3.20%</td><td>10,000</td><td>100.009</td><td>_</td><td></td></missi<>	LOW	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL	NORM		3.20%	10,000	100.009	_	
SCONNECT		<missi< td=""><td>HIGH</td><td>LOW</td><td>FALSE</td><td>ZERO</td><td>NORMAL</td><td>LOW</td><td>LOW</td><td>NORMAL</td><td></td><td>520</td><td>0.2070</td><td>10,000</td><td></td><td></td><td></td></missi<>	HIGH	LOW	FALSE	ZERO	NORMAL	LOW	LOW	NORMAL		520	0.2070	10,000			
INVOLSET	С	<missi td="" ~<=""><td>NORM</td><td></td><td>FALSE</td><td>HIGH</td><td>NORMAL</td><td>HIGH</td><td>HIGH</td><td></td><td>To</td><td>tal 10,000</td><td>100.00%</td><td></td><td></td><td></td><td></td></missi>	NORM		FALSE	HIGH	NORMAL	HIGH	HIGH		To	tal 10,000	100.00%				



Now press button Change Categories to map the existing categories of the variable to the desired new categories and optionally give a new name for the recoded variable:

	Deced by the	- Norma			
/ariable Name	Recoded Variabl	e Name			
MINVOL	MINVOL_recoded	<b>11</b>			
old categor		v category order			
HIGH	HIGH	🔞 Recode Categorical Varia	able		
LOW	LOW				
NORMAL	NORMAL	Variable Name	Recoded Variable Nam	e	
ZERO	ZERO	-			
		MINVOL	MINVOL_positivezero		
		ald attended			order
		old category	new catego POSITIVE	ory 0	order
		LOW		2	
		2011	POSITIVE		
		NORMAL	POSITIVE	3	
n Rec	ode Categorical Variab	ZERO	POSITIVE ZERO	3 4 ×	
		ZERO		4	
Var	iable Name R	ZERO Ne Recoded Variable Name		4	
Var	iable Name R	ZERO		4	
Var	iable Name R	ZERO Ne Recoded Variable Name		4	
Var	iable Name R	ZERO Ne Recoded Variable Name		4	
Var	iable Name R	ZERO Ne Recoded Variable Name		4	Cancel
Var	iable Name R	ZERO Ne Recoded Variable Name		4	Cancel
Var	iable Name R	ZERO Recoded Variable Name MINVOL_positivezero	ZERO	4	Cancel
Var	iable Name R IVOL old category	ZERO  Lecoded Variable Name  MINVOL_positivezero  new category	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category	ZERO       Ile       Recoded Variable Name       MINVOL_positivezero       new category       POSITIVE	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO  Ale  Ale  Ale  Ale  Ale  Ale  Ale  Al	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO         Ile         Recoded Variable Name         MINUL_positivezero         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO         Ile         Recoded Variable Name         MINUL_positivezero         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO         Ile         Recoded Variable Name         MINUL_positivezero         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO         Ile         Recoded Variable Name         MINUL_positivezero         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO         Ile         Recoded Variable Name         MINUL_positivezero         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO         Ile         Recoded Variable Name         MINUL_positivezero         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE	ZERO	4	Cancel
Var MIN HIG LOV	iable Name R IVOL old category 6H W RMAL	ZERO         Ile         Recoded Variable Name         MINUL_positivezero         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE         POSITIVE	ZERO	4	Cancel

When you are satisfied with the proposed recoding, press button Recode Variable next to the Change Categories button to actually create the new recoded variable:



This will create a new variable with the new name, with two categories (POSITIVE and ZERO), and show it in the **Dashboard** tab, where it appears at the top of the **Variables** table and in the first column of the **Data** table. Press button **Calculate Frequencies**, the Enter key or double-click in the new variable column to see its data:

About DIONE		Open Dat	ta File Loa	ad Data	Save Data	Bayesian N	letwork:	Options	Learn Structure	Fit Parameters	Infer Probabilities	Stop Algori	ithm Sa	ve Results	100%	frequency calculations
Dashboard	Explo	re Grapi	hical Model													
Variables	✓ Sel	ect All	Data Clear	Highlighted	Clear All	Show Cases:	All     Com	plete 🔵 Inc	omplete 10		Calculate	Frequencies		Calculate Bive	ariate Tests	
<filter></filter>	<mi< th=""><th>ssing&gt;</th><th>MINVOL_p</th><th>HISTORY</th><th>CVP</th><th>PCWP</th><th>HYPOVOLE</th><th>LVEDVOLU</th><th> LVFAILURE</th><th></th><th></th><th>•</th><th>Histogra</th><th>m</th><th></th><th></th></mi<>	ssing>	MINVOL_p	HISTORY	CVP	PCWP	HYPOVOLE	LVEDVOLU	LVFAILURE			•	Histogra	m		
name	type	missin	POSITIVE	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	1.0						
MINVOL_positi	0	<missi td="" 🆳<=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>S 0.8</td><td></td><td></td><td></td><td></td><td></td><td></td></missi>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	S 0.8						
HISTORY		<missi< td=""><td>POSITIVE</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>0.8 0.7 0.6 0.5 0.4</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	POSITIVE	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	0.8 0.7 0.6 0.5 0.4						
CVP	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>0.5</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	0.5						
PCWP	С	<missi< td=""><td>POSITIVE</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>0.4 0.3</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	POSITIVE	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	0.4 0.3						
HYPOVOLEMIA		<missi< td=""><td>ZERO</td><td>TRUE</td><td>LOW</td><td>LOW</td><td>FALSE</td><td>LOW</td><td>TRUE</td><td>E 0.2</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	TRUE	LOW	LOW	FALSE	LOW	TRUE	E 0.2						
LVEDVOLUME	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>0.1</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	0.1						
LVFAILURE		<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>0.0</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	0.0						
STROKEVOLUM	EC	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td></td><td></td><td></td><td>val</td><td>ue</td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE				val	ue		
ERRLOWOUTP		<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td></td><td></td><td>Variab</td><td>le: MINVOL_</td><td>positivezero</td><td>)</td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE			Variab	le: MINVOL_	positivezero	)	
HRBP	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>7,000</td><td></td><td></td><td></td><td></td><td></td><td>7,508</td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	7,000						7,508
HREKG	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>6,000 · 5,000 ·</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	6,000 · 5,000 ·						
ERRCAUTER		<missi< td=""><td>POSITIVE</td><td>FALSE</td><td>NORMAL</td><td>LOW</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>5,000 ·</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	POSITIVE	FALSE	NORMAL	LOW	FALSE	NORMAL	FALSE	5,000 ·						
HRSAT	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>HIGH</td><td>HIGH</td><td>TRUE</td><td>HIGH</td><td>FALSE</td><td>4,000 ·</td><td>2,65</td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	HIGH	HIGH	TRUE	HIGH	FALSE	4,000 ·	2,65					
INSUFFANESTH		<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>5 2,000</td><td>2,00</td><td>1</td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	5 2,000	2,00	1				
ANAPHYLAXIS		<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>≥ 1,000</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	≥ 1,000						
TPR	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>HIGH</td><td>HIGH</td><td>TRUE</td><td>HIGH</td><td>FALSE</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	HIGH	HIGH	TRUE	HIGH	FALSE	0						
EXPCO2	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td></td><td></td><td></td><td>_</td><td>positivezero Egory</td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE				_	positivezero Egory		
KINKEDTUBE		<missi< td=""><td>ZERO</td><td>FALSE</td><td>HIGH</td><td>HIGH</td><td>TRUE</td><td>HIGH</td><td>FALSE</td><td></td><td></td><td></td><td>POSITIVE</td><td></td><td></td><td></td></missi<>	ZERO	FALSE	HIGH	HIGH	TRUE	HIGH	FALSE				POSITIVE			
MINVOL	С	<missi< td=""><td>POSITIVE</td><td>FALSE</td><td>NORMAL</td><td>LOW</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	POSITIVE	FALSE	NORMAL	LOW	FALSE	NORMAL	FALSE							
FIO2	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>HIGH</td><td>HIGH</td><td>TRUE</td><td>HIGH</td><td>FALSE</td><td>Variable</td><td>-</td><td></td><td>Cumulative</td><td>Cumulative</td><td></td><td></td></missi<>	ZERO	FALSE	HIGH	HIGH	TRUE	HIGH	FALSE	Variable	-		Cumulative	Cumulative		
PVSAT	С	<missi< td=""><td>POSITIVE</td><td>FALSE</td><td>HIGH</td><td>HIGH</td><td>TRUE</td><td>HIGH</td><td>FALSE</td><td>Values</td><td>Frequency</td><td>Percentage</td><td>Frequency</td><td>Percentage</td><td></td><td></td></missi<>	POSITIVE	FALSE	HIGH	HIGH	TRUE	HIGH	FALSE	Values	Frequency	Percentage	Frequency	Percentage		
SAO2	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE							
PAP	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>MINVOL_positivezo</td><td>er</td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	MINVOL_positivezo	er					
PULMEMBOLUS		<missi< td=""><td>POSITIVE</td><td>FALSE</td><td>NORMAL</td><td>HIGH</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>• POSITIVE</td><td>2.691</td><td>26.91%</td><td>2,691</td><td>26.91%</td><td></td><td></td></missi<>	POSITIVE	FALSE	NORMAL	HIGH	FALSE	NORMAL	FALSE	• POSITIVE	2.691	26.91%	2,691	26.91%		
SHUNT	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>ZERC</td><td></td><td>73.09%</td><td>10,000</td><td>100.00%</td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	ZERC		73.09%	10,000	100.00%		
INTUBATION	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>HIGH</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	HIGH	NORMAL	FALSE	NORMAL	FALSE							
PRESS	С	<missi< td=""><td>ZERO</td><td>FALSE</td><td>NORMAL</td><td>NORMAL</td><td>FALSE</td><td>NORMAL</td><td>FALSE</td><td>Tota</td><td>I 10,000</td><td>100.00%</td><td></td><td></td><td></td><td></td></missi<>	ZERO	FALSE	NORMAL	NORMAL	FALSE	NORMAL	FALSE	Tota	I 10,000	100.00%				
DISCONNECT		<missi td="" ~<=""><td>POSITIVE</td><td>FALSE</td><td>NORMAL</td><td>NORMAI</td><td>FALSE</td><td>NORMAL</td><td>FAISE ~</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></missi>	POSITIVE	FALSE	NORMAL	NORMAI	FALSE	NORMAL	FAISE ~							

## Learn Bayesian Networks

In the **Dashboard** tab, ensure that the variables you want to analyse are selected in the **Variables** table and, by pressing the **Load Data** button or the Enter key, loaded and shown in the **Data** table. If you want to change the variables to analyse, change the selected variables as explained under **Load Data**.

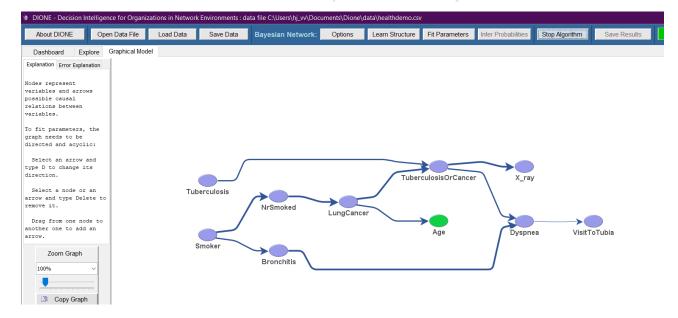
#### Learn Network Structure

Now press button Learn Structure to learn a Bayesian network structure from the selected data.

About DIONE		Open Data F	ile Load D	)ata Save	Data Bay	yesian Netwo	ork: Option	ns Learn	Structure F	it Parameters	Infer Probabilit	ies Stop Algor
Dashboard	Explo	re Graphica	I Model									
Variables	✓ Sel	ect All	Data Clea	r Highlighted	Clear All S	how Cases: 🧕	All Com	plete 🔵 Inco	omplete 129	9 cases for 10 v	ariables	
<filter></filter>	<m< th=""><th>issing&gt;</th><th>Smoker</th><th>LungConcer</th><th>VisitTaTubia</th><th>Tuborgulasis</th><th>Tuberculosis</th><th>V</th><th>Provobitio</th><th>Durannes</th><th>4.55</th><th>NrSmoked</th></m<>	issing>	Smoker	LungConcer	VisitTaTubia	Tuborgulasis	Tuberculosis	V	Provobitio	Durannes	4.55	NrSmoked
name	type	missing	Smoker	LungCancer	VisitToTubia	Tuberculosis	OrCancer	X_ray	Bronchitis	Dyspnea	Age	NrSmoked
Smoker	С	<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	53	9
LungCancer		<missing></missing>	no	yes	no	no	yes	yes	no	no	62	0
VisitToTubia		<missing></missing>	yes	yes	no	no	yes	yes	no	yes	62	8
Tuberculosis		<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	74	10
TuberculosisOr		<missing></missing>	yes	yes	no	no	yes	yes	no	yes	47	6
X_ray		<missing></missing>	yes	yes	no	no	yes	yes	yes	no	57	4
Bronchitis		<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	71	10
Dyspnea		<missing></missing>	yes	yes	no	no	yes	yes	yes	yes	61	6
Age		<missing></missing>	no	yes	no	no	yes	yes	no	yes	53	0
VrSmoked		<missing></missing>	yes	yes	no	no	yes	yes	no	yes	48	4
			yes	yes	no	no	yes	yes	yes	yes	72	9
			VOC	MOC	20	20	1005	Voc	Voc	Voc	46	Q

#### The result should look something like this in the

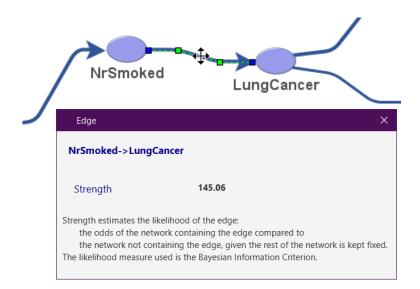
Graphical Model tab:



In the learned Bayesian network, nodes represent variables and arrows represent possible causal relations between variables. Directed lines, with arrows, represent possible causal relations including the direction of causality, while any undirected lines, without arrows, represent possible causal relations with unspecified causal direction. If there is no significant association between two variables, there is no line between their nodes. The default structure learning algorithm, tabu search, produces a directed graph with only arrows.

The algorithm also calculates edge strengths. The strength of an edge estimates the likelihood of the edge: the odds of the network containing the edge compared to the network not containing the edge, given the rest of the network is kept fixed. The likelihood measure used is the Bayesian Information Criterion.

Edge strengths are indicated by the thickness and colour of the arrows. The thicker an arrow, the stronger is the evidence for the causal connection. Edge strengths can be inspected by right-clicking on an arrow:



In this example the strength of the connection between **NrSmoked** and **LungCancer** is over 145, indication that the network with this connection is much more likely than a network without the connection and otherwise the same.

The default structure learning algorithm is tabu search, but other structure learning algorithms can be used by selecting one in the Set Algorithm Options dialog accessed by pressing the Algorithms button:

Set Algorithm Options				×
Selected Algorithm:	tabu - Tabu Se	arch		
Selected Score: bic	cg - Bayesian Informatio	on Criterion for Conditiona	al Gaussian Networks	
Type of Algorithm:	Score-based	Constraint-based	Local Discovery	Hybrid
Algorithm:	tabu 🔻	•	•	¥
Network Score:	bic-cg 👻			
Conditional Independence Test:		mi-cg 💌		
Probability of Incorrect Effect:		0.05		
Impute Missing Values:				
		J		
Sample Size:		100000		
				Cancel OK

DIONE uses the R package bnlearn for structure learning and parameter fitting. Supported structure learning algorithms are described in the <u>bnlearn manual</u> on pages 100 – 102 at bookmark structure-learning. Structure learning algorithms can be Score-based, Constraint-based, Local Discovery or Hybrid algorithms. To use a specific algorithm, select it in the Algorithm drop down list under the relevant category.

Score-based algorithms use a network score to compare solutions. Possible network scores are described in the <u>bnlearn manual</u> on pages 81 – 82 at bookmark network-scores. To use a specific score, select it in the Network Score drop down list.

Constraint-based algorithms and Local Discovery algorithms use a conditional independence test to find solutions. Possible tests are described in the <u>bnlearn manual</u> on pages 66 – 67 at bookmark independence-tests. To use a specific test, select it in the Conditional Independence Test drop down list.

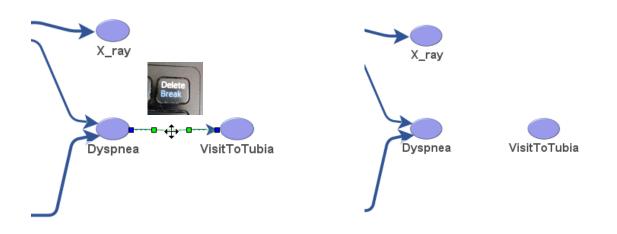
Hybrid algorithms use a network score as well as a conditional independence test. When a Hybrid algorithm is selected in the Algorithm drop down list under Hybrid, the algorithm uses the Network Score selected under Score-based and the Conditional Independence Test selected under Constraint-based. When the rsmax2 Hybrid algorithm is selected, it combines the algorithms selected under Score-based and under Constraint-based.

#### **Edit Network Structure**

The next step in the analysis will be the parameter fitting algorithm, which calculates detailed information for all nodes and arrows, such as conditional probabilities and/or regression coefficients quantifying the causal relationships between variables. Before continuing with parameter fitting, you may wish to edit the network structure to reflect knowledge or intuition you already have about causal connections between variables. You can move or delete nodes, add or delete arrows or reverse their direction.

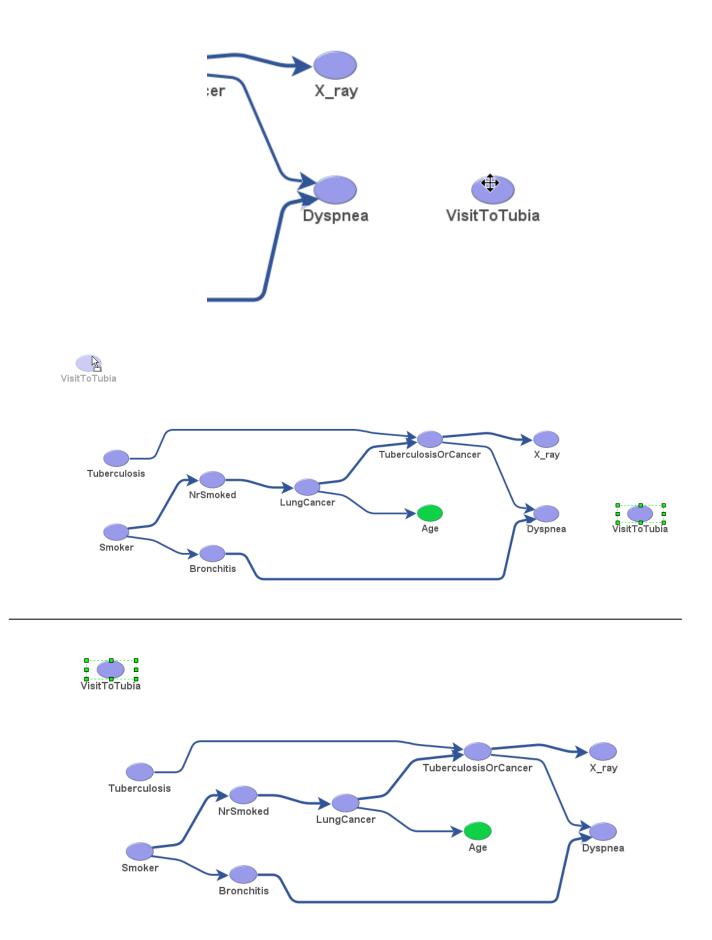
Adding an arrow or changing direction of an arrow can possibly result in a red arrow, meaning that a network with that connection is less likely than a network without it.

For example, in the learned network shown on page 20 the arrow between variables **Dyspnea** and **VisitToTubia** does not seem to reflect a likely causal connection. Let us say we want to remove it and instead postulate a causal connection between **VisitToTubia** and **Tuberculosis**.

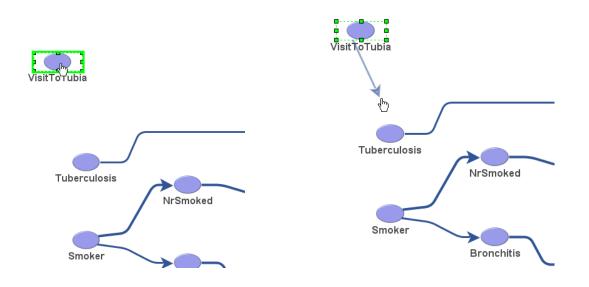


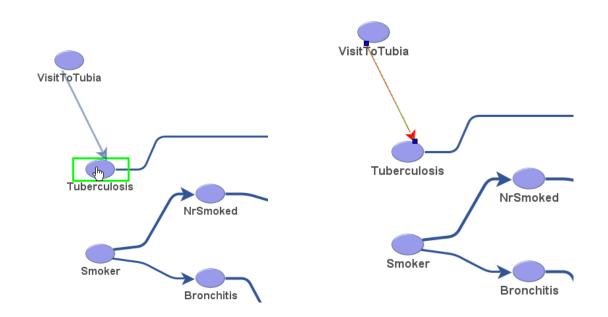
First, we delete the arrow between **Dyspnea** and **VisitToTubia** with the Delete key:

Now, to keep the diagram tidy, we move the node of variable **VisitToTubia** to a position near the **Tuberculosis** node. To move a node, hover over it with the mouse to show a black cross and drag it to the desired location:



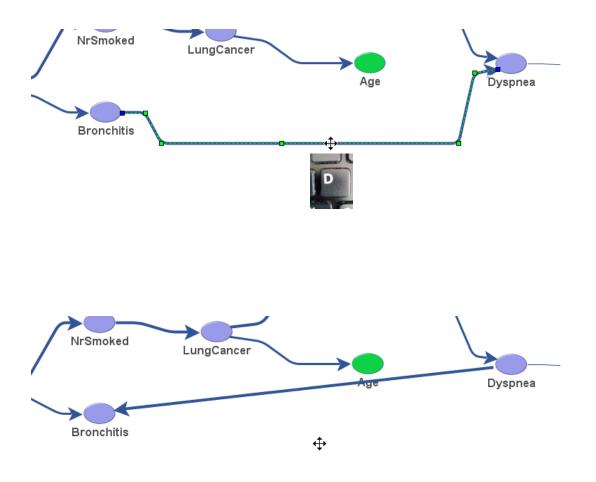
To add a new arrow between the two nodes **VisitToTubia** and **Tuberculosis**, first hover with the cursor over the node that is considered the cause (**VisitToTubia**) and where the arrow is to start, such that the cursor is a hand and there is a green frame around the node. Now press the left mouse button and, while holding it, move the mouse to draw an arrow to the node that is considered the effect (**Tuberculosis**) and where the arrow is to end. When there is a green frame around the effect node, release the mouse button. The new arrow has now been added:





Note that the arrow between nodes **VisitToTubia** and **Tuberculosis**, which has been manually added, is red, meaning that the causal relationship between variables **VisitToTubia** and **Tuberculosis** is not likely, given the data.

You can also reverse the direction of directed arrows if their direction does not make sense to you, or to eliminate cycles. To reverse the direction of an arrow, select it by clicking on the line of the arrow and push the **D** key on your keyboard:



If undirected arrows are present in the learned network, you can give them a direction in the same way, pushing the **D** key to give an arrow a direction and possibly pushing it again to reverse the direction.

#### **Fit Parameters**

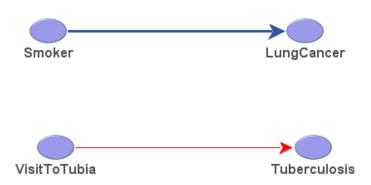
More information about causal relations between variables in the network can be gained by fitting parameters of the network to obtain conditional probability tables and regression coefficients, showing the (probabilities of) values of a node given values of its parent nodes. Parameters are estimated with a Maximum Likelihood algorithm.

When you are done editing network connections and the graph is completely directed, i.e. there are no undirected edges present, and the graph does not have cycles, you can calculate conditional probabilities, marginal probabilities and regression coefficients for numerical variables by pressing button Fit Parameters .

For clarity in the example shown here structure learning has been done with only four variables:

DIONE - Decisio	on Inte	Iligence for Org	ganizations i	n Network	Enviro	onments : data	file C:\Users\hj_	vv\Documents	Dione\dat	ta\healthdemo.csv
About DIONE		Open Data Fi	ile Loa	d Data	S	ave Data	Bayesian Ne	twork: O	otions	Learn Structure
Dashboard	Explo	re Graphical	Model							
Variables	🗸 Sel	ect All		Data	Clear	Highlighted	Clear All	Show Cases:		Complete 🔵 Incomp
<filter></filter>		<missing></missing>		Smo	ker	LungCancer	VisitToTubia	Tuberculosis		
name	type	missing		yes		yes	no	no		
Smoker		<missing></missing>		no		yes	no	no		
LungCancer		<missing></missing>		yes		yes	no	no		
VisitToTubia		<missing></missing>		yes		yes	no	no		
Tuberculosis		<missing></missing>		yes		yes	no	no		
TuberculosisOr	С	<missing></missing>		yes		yes	no	no		
X_ray	С	<missing></missing>		yes		yes	no	no		
Bronchitis	С	<missing></missing>		Ves		Ves	no	no		

After some editing, we have this simple network:



When the parameter fitting algorithm has finished, right-click on any node of a categorical variable (shown in blue) to see its conditional probability table, showing conditional probabilities given the values of its parents, along with 95 % confidence intervals for these conditional probabilities. At the bottom of the dialog marginal probabilities of the node are shown:

Variable						
LungCancer			Save	Result		
Conditional Pr	obabilities					
Parent Node		Cancer=no		Cancer=yes	Count	
	Cond Prop	95% CI	Cond Prop	95% CI		
Smoker=no	0.992	[0.982 , 0.997]	0.00766	[0.00327 , 0.0178]	653	
Smoker=yes	0.907	[0.882 , 0.927]	0.0929	[0.0728 , 0.118]	646	
Marginal Prob	abilities					
LungCan	cer=no	LungCancer=	yes Evic	lence Intervention		
		0.0500		0		

For a node of a numerical variable (shown in green), coefficients of regression equations are displayed, along with the standard deviation of the residuals. In this example, variable **NrSmoked** depends on the two variables **Age** and **Smoker**. **Smoker** has two possible values, so there are regression equations reflecting the dependence of **NrSmoked** on **Age** for each of the values of **Smoker**:

e	Variable	Hr smok				
	NrSmoked				Save Result	]
	Coefficients					
5	Parent Nod	e Intercept	Age		StDev	Count
oker	Smoker=no	0.000	)	0.00	0.00	653
	Smoker=yes	1.74	ŀ	0.0396	1.83	646
	Marginal Pro	obabilities				
	Mean	St Dev	Min	Max	k Evidence	Intervention
	1.69	2.16	0.000	0.000		

### **Infer Probabilities**

Now you can perform inference on the Bayesian network by setting evidence on one or more nodes, then pressing button Infer Probabilities to recompute marginal probabilities and see the effect of the evidence on the other nodes:

			Dione\da	ta\healthdemo.csv				
Variable			tions	Learn Structure	Fit Parameters	Infer Probabilities	Stop Algorithm	Sa
Smoker			s					
Conditional Pr	obabilities							
	Smok	er=no	Smol	ker=yes				
Parent Node	Cond Prop	95% CI	Cond Prop	95% CI	Count			
env=e	0.503	[0.476 , 0.530]	0.497	[0.470 , 0.524]	1299			
					Variable			Í
Marginal Prob	abilities							_
Smoke					LungCancer			Sa
		Smoker=ye	S EVIC	ence Intervent				
0.000	1.	000	S EVIU	ence Intervent	Conditional P			1
0.000	1.				Conditional P	LungCar		
0.000	1.				Rarent Node	LungCar Cond Prop	95% CI	Lun Cond Prop
0.000	1.				Rarent Node Smoker=no	LungCar Cond Prop 0.992	<b>95% Cl</b> [0.982 , 0.997]	Cond Prop 0.007
0.000	1.				Rarent Node	LungCar Cond Prop	95% CI	
0.000	1.				Rarent Node Smoker=no	LungCar Cond Prop 0.992	<b>95% Cl</b> [0.982 , 0.997]	Cond Prop 0.007
0.000	1.				Rarent Node Smoker=no	LungCar Cond Prop 0.992 0.907	<b>95% Cl</b> [0.982 , 0.997]	Cond Prop 0.007
0.000	1.				Rarent Node Smoker=no Smoker=yes Marginal Prot	LungCan Cond Prop 0.992 0.907	<b>95% CI</b> [0.982 , 0.997] [0.882 , 0.927]	Cond Prop 0.007( 0.09)
0.000	1.				<b>Rarent Node</b> Smoker=no Smoker=yes	LungCar Cond Prop 0.992 0.907 Dabilities	<b>95% CI</b> [0.982 , 0.997] [0.882 , 0.927] LungCancer=	Cond Proj 0.007 0.09
0.000	1.				Rarent Node Smoker=no Smoker=yes Marginal Prok	LungCar Cond Prop 0.992 0.907 Dabilities	<b>95% CI</b> [0.982 , 0.997] [0.882 , 0.927] LungCancer=	Cond Proj 0.007 0.09

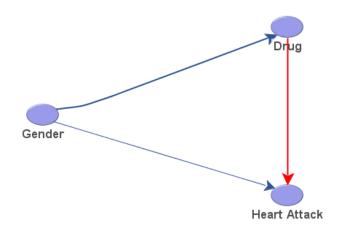
In this example the probability of lung cancer has increased from 5 % to over 9 % on the evidence that someone is a smoker.

We can also infer the effects of interventions, as illustrated by the following example related to the famous Simpson paradox. The Simpson paradox occurs in these fictitious data on gender, taking of a drug and risk of a heart attack (from Pearl & Mackenzie 2018):

	Control Grou	up (No Drug)	Treatment Gro	up (Took Drug)
	Heart Attack	No Heart Attack	Heart Attack	No Heart Attack
Female	1	19	3	37
Male	12	28	8	12
Total	13	47	11	49

From these data it seems that the drug increases the risk of a heart attack for women, as 3/40 (7.5%) > 1/20 (5%), it also increases the risk for men, as 8/20 (40%) > 12/40 (30%), yet it decreases the risk for the population as a whole, as 11/60 (18.33%) < 13/60 (21.67%).

This is shown in **DIONE** by the data in file simpson.csv: Learn Structure and add an arrow to the learned network



After **Fit Parameters** right-click on the **Heart Attack** node to see its conditional probabilities:

eart Attack	1-1141		Save Result		
onditional Proba	Heart Att	ack=no	Heart Att	tack=yes	Count
Parent Node	Cond Prop	95% CI	Cond Prop	95% CI	Count
Drug=no					
Gender=F	0.950	[0.764 , 0.991]	0.0500	[0.00888, 0.236]	20
Drug=no					
Gender=M	0.700	[0.546 , 0.819]	0.300	[0.181 , 0.454]	40
<ul> <li>Drug=yes</li> </ul>					
Gender=F	0.925	[0.801 , 0.974]	0.0750	[0.0258 , 0.199]	40
<ul> <li>Drug=yes</li> </ul>					
Gender=M	0.600	[0.387 , 0.781]	0.400	[0.219 , 0.613]	20

The conditional probabilities of getting a heart attack according to gender are the same as the percentages in the table. Yet, when comparing variables Drug and Heart Attack in the **Dashboard** tab, we see these percentages for the population:

Variable Values	Drug				
Heart Attack	n	0	у	es	
no	47	78.33%	49	81.67%	
yes	13	21.67%	11	18.33%	
Total	60	100.00%	60	100.00%	

The paradox is resolved by realizing that **Gender** is a confounding variable that influences both the risk of a heart attack and the probability of taking the drug: women have less risk of a heart attack, but also take the drug more often than men. To correctly estimate the effect of the drug on the risk of a heart attack, we must adjust for gender and estimate the risk of a heart attack when taking the drug by averaging the percentages for men and women.

The paradox is illustrated in **DIONE** by inferring the probability of a heart attack given observational *evidence* of taking the drug, compared to the effect of an *intervention* to take the drug. Setting evidence that the drug has been taken and inferring probabilities gives the following results, with a probability of a heart attack of 18.3% when taking the drug and 21.7% when not taking the drug:

	Save Data	Bayesian Network:	Options	Learn Structure	Fit Parameters	InfenProbabilities	Stop Algorithr	m Save
			Variable					,
			Drug			Save Result		
			Conditiona	l Probabilities				
				Dr	ug=no	Drug=	yes	
			Parent Noo	de Cond Prop	95% CI	Cond Prop	95% CI	Count
		Drug	Gender=F	0.333	[0.227 , 0.459]	0.667	[0.541 , 0.773]	60
Gen	nder		Gender=M	0.667	[0.541 , 0.773]	0.333	[0.227 , 0.459]	60
	ginal Probabiliti							
	art Attack=no	Heart Attack=yes Evider 0.183 <sub>k</sub>	nce Interventio	n 				
<b>Не</b> 0.81	aart Attack=no 7	Heart Attack=yes Evider 0.183	nce Interventio	n ] 				
He 0.81	art Attack=no	Heart Attack=yes Evider						
He 0.81 largir	a <b>rt Attack=no</b> 7 nal Probabilities	Heart Attack=yes Evider		n				
He 0.81	art Attack=no 7 nal Probabilities g=no Drug=	Heart Attack=yes Evider	tion Marginal Pro	bbabilities ►	ack=yes byidence			

The evidence means we have only observed that the drug has been taken without doing anything, and the inference will include the effect of the confounding variable **Gender**, so the incorrect conclusion could be drawn that the drug is beneficial for the population.

When setting an *intervention* to take the drug, evidence is also set that the drug has been taken, but now the effect of confounding is eliminated by disregarding the arrow from **Gender** to **Drug** in the network. We are assuming the drug is now given to people, without considering gender. This gives the following inference results:

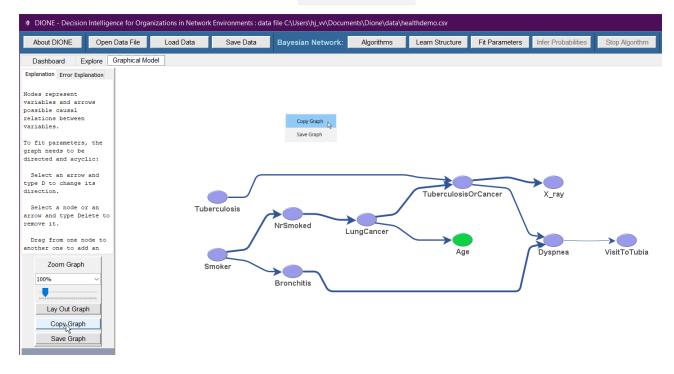
Drug=no	Drug=yes	Evidence	Intervention	Marginal Proba	abilities	$\frown$			
0.000	1.000			Heart Attack	<b>no Hear</b> 0.238	: Attack=yes	Evidence	Intervention	
Marginal Prol	oabilities								
Marginal Prol Drug=no	Dabilities Drug=yes	Evidence	Intervention	Marginal Pro	obabilities				

An intervention means we have given the drug to people regardless of gender, and the inference will eliminate the effect of the confounding variable **Gender**, so the correct conclusion can be drawn that the drug increases the risk of a heart attack for the population.

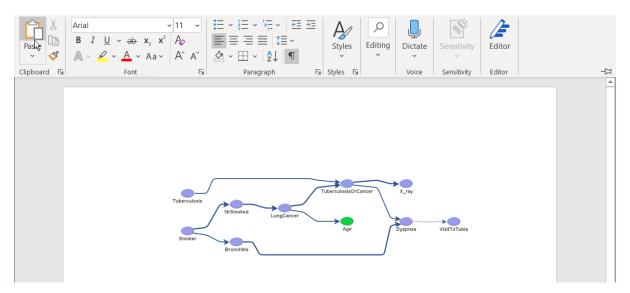
### **Copy and Save Results**

You can keep your results by copying the network graph to another application, for example to insert it in a word processing document, or by saving it to a file. After fitting parameters, you can also save numerical results to a text file.

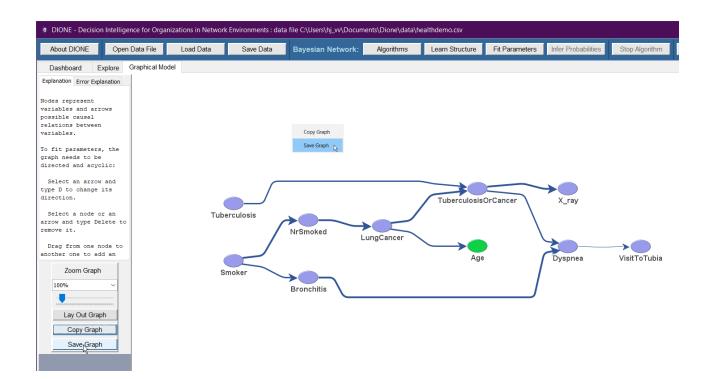
To copy the network graph, press button Copy Graph in the left-hand panel to put a picture of the chart on your computer's clipboard, or, if you prefer, you can right-click on the background of the graph and select Copy Graph from the pop-up menu:



You can now paste the picture in a Word document, for example:



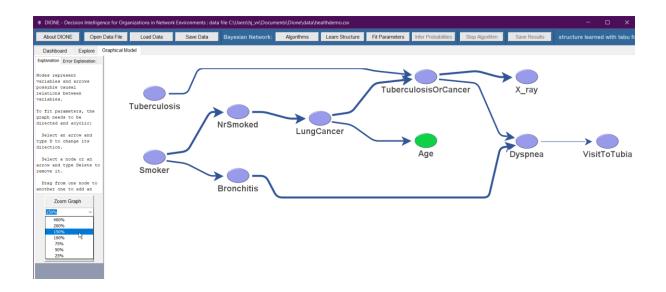
Press buttonSave Graphin the left-hand panel to save a network graph to a file,or right-click on the graph background and selectSave Graphfrom the pop-up menu:



Save Results File					×
$\leftarrow \rightarrow \checkmark \uparrow$	📙 « Documents > Dione > results		✓ Ŭ	Search results	
Organize 🔹 🛛	New folder			<b>•</b>	?
★ Quick access	5	~	healthdemo1.pn g	healthdemo2.pn g	
	healtdemo				~
	Portable Network Graphics (*.png)				$\sim$
∧ Hide Folders				Save	:I

You now have a Portable Network Graphics (PNG) picture file that you van open in any image processing program or insert in a document.

With optionZoom Graphin the left-hand panel you can zoom in or out the networkgraph:



With option Lay Out Graph you can redraw the graph with a nice layout, if it has become messy after a lot of manual editing, for example.

## Save Data

In the current version only data saving to CSV files is supported and in the saved CSV file column headers contain variable names, with values in the columns. The saved data file will have all data of variables shown in the **Variables** table, including any recoded data, for example the recoded **Age** variable here:

About DIONE	Open	Data File	Load Data	Save Da	ta N Bay	esian Networ	k: Algorith	ims Lea	arn Structure	Fit Paramete	rs Infer Proba	abilities	Stop Algorithm	Save R	esults re	ecoding com	pleted
Dashboard Exp	olore	Graphical Mode	el		\$												
Variables 🗸	Select A	II	Data Clear	Highlighted	Clear All	Show Cases:	All 🔵 Com	nplete 🔵 Inc	omplete 1	299 cases for 11 v	/ariables		Calculate	Frequencies	Calculate	Bivariate Tests	1
<filter></filter>	<mi< th=""><th>issing&gt;</th><th>Age_2groups</th><th>Smoker</th><th>LungCancer</th><th>VisitToTubia</th><th>Tuberculosis</th><th>Tuberculosis OrCancer</th><th>X_ray</th><th>Bronchitis</th><th>Dyspnea</th><th>4</th><th></th><th>Varia</th><th>ble: Age</th><th>922</th><th></th></mi<>	issing>	Age_2groups	Smoker	LungCancer	VisitToTubia	Tuberculosis	Tuberculosis OrCancer	X_ray	Bronchitis	Dyspnea	4		Varia	ble: Age	922	
Name	Туре	Missing	35-83	yes	1105	no	no		1100		1105	~	900 - 800 -				
Age_2groups	С	<missing></missing>	35-83		yes			yes	yes	yes	yes		700 - 600 -				
Smoker	С	<missing></missing>		no	yes	no	no	yes	yes	no	no	-	600 GOO				
ungCancer	С	<missing></missing>		yes	yes	no	no	yes	yes	no	yes		500	377			
/isitToTubia	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes	yes		5 300				
Fuberculosis	С	<missing></missing>		yes	yes	no	no	yes	yes	no	yes	_	È 200				
FuberculosisOrCan	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes	no		100				
X_ray	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes	yes			20 25 30 35	40 45 50 5	5 60 65 70	75 8
Bronchitis	С	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes	yes				value		
Dyspnea	С	<missing></missing>	35-83	no	yes	no	no	yes	yes	no	yes			Age	2groups		
Age	N	<missing></missing>	35-83	yes	yes	no	no	yes	yes	no	yes		900 -			022	
NrSmoked	с	<missing></missing>	35-83	yes	yes	no	no	yes	yes	yes	yes		2 800 ·				
		-	35-83	yes	yes	no	no	yes	yes	yes	yes		00 700 600				
			35-83	yes	yes	no	no	yes	yes	yes	yes		800 - 700 -	077			
			35-83	yes	yes	no	yes	yes	yes	no	yes		9 400 - 5 300 -	377			
			35-83	no	yes	no	no	yes	yes	no	yes		È 200				
			35-83	yes	yes	no	no	yes	yes	no	yes		100				
			35-83	no	yes	no	no	yes	yes	no	no		0		categories		
			35-83	yes	yes	no	no	yes	yes	yes	yes				category		
			35-83	yes	yes	no	no	yes	yes	yes	yes			<b>6</b> -	34 🔳 35-83		
			35-83	yes	ves	no	no			no	yes		Mariahia				
			35-83					yes	yes				Variable	Frequency	Percentage	Cumulative	Cu
				yes	yes	no	no	yes	yes	yes	yes		Values	. requeriey	seconage	Frequency	Per
			35-83	yes	yes	no	no	yes	yes	no	yes		•				-
			35-83	yes	yes	no	no	yes	yes	yes	yes		Age	11	0.85%	11	
			35-83	yes	yes	no	no	yes	yes	yes	yes		[6.0 - 13.7) [13.7 - 21.4)	50	0.85%	61	+
			35-83	yes	yes	no	no	yes	yes	yes	no		[13.7 - 21.4]	142	10.93%	203	-
			35-83	yes	yes	no	no	yes	yes	no	yes		[29.1 - 36.8]	273	21.02%	476	-
			35-83	yes	yes	no	no	yes	yes	yes	yes		[36.8 - 44.5)	350	26.94%	826	-
			35-83	yes	yes	no	no	yes	yes	no	yes		[44.5 - 52.2)	249	19.17%	1.075	

Note that the saved data will not include data for any variables you have removed from the **Variables** table. If radio button Complete or Incomplete has been selected, only complete or incomplete cases will be saved.

To save data to a CSV file, press button	Save Data	and in the	Save Data File	dialog
enter a name for the data file you wish to	o create:	1		•

Save Data File											×
← → ~ ↑ 🖡 « D	ocuments > Dione > data			~	Ü		0 5	earch da	ita		
Organize 🔹 New fold	er								4	•	?
🗢 This PC		^		Name			,	^			Da ^
🧊 3D Objects				📕 Si	mone						6/
📃 Desktop				🔊 al	arm.csv						8/
Documents				🔊 Cl	hronicIII	ness	Careg	ivingDat	a.csv		5/
Downloads				🔁 Cł	hronicIII	ness	Careg	iving-He	ealth.cs	/	10
Music				🔁 CI	hronicIII	ness	Careg	iving-He	althRee	coded	10
Pictures				🖾 he	ealthden	no -	Сору	.CSV			24
				🛂 he	ealthden	no.c	sv				10
🛃 Videos				🖾 he	ealthden	no+i	recod	ed.csv			20
💺 Windows (C:)				🔼 he	ealthden	no1.	CSV				4/ 🗸
Motwork		$\sim$	<								>
File <u>n</u> ame: heal	thdemo+recoded.csv	I									$\sim$
Save as type: Data	Files (*.csv)										$\sim$
<ul> <li>Hide Folders</li> </ul>							<u>S</u>	ave		Cancel	

The saved data file now looks like this:

	А	В	С	D	E	F	G	Н	1	J	K	L	
1	Age_recoded1	Smoker	LungCancer	VisitToTubia	Tuberculosis	TuberculosisOrCancer	X-ray	Bronchitis	Dyspnea	Age	NrSmoked		
2	31-57	yes	yes	no	no	yes	yes	yes	yes	53	9		
3	57-83	no	yes	no	no	yes	yes	no	no	62	0		
4	57-83	yes	yes	no	no	yes	yes	no	yes	62	8		
5	57-83	yes	yes	no	no	yes	yes	yes	yes	74	10		
6	31-57	yes	yes	no	no	yes	yes	no	yes	47	6		
7	31-57	yes	yes	no	no	yes	yes	yes	no	57	4		
8	57-83	yes	yes	no	no	yes	yes	yes	yes	71	10		
9	57-83	yes	yes	no	no	yes	yes	yes	yes	61	6		
10	31-57	no	yes	no	no	yes	yes	no	yes	53	0		
11	31-57	yes	yes	no	no	yes	yes	no	yes	48	4		
12	57-83	yes	yes	no	no	yes	yes	yes	yes	72	9		
13	31-57	yes	yes	no	no	yes	yes	yes	yes	46	8		
14	57-83	yes	yes	no	no	yes	yes	yes	yes	61	7		
15	57-83	yes	yes	no	yes	yes	yes	no	yes	62	9		
16	31-57	no	yes	no	no	yes	yes	no	yes	41	0		
17	31-57	yes	yes	no	no	yes	yes	no	yes	35	9		
18	31-57	no	yes	no	no	yes	yes	no	no	40	0		
19	31-57	yes	yes	no	no	yes	yes	yes	yes	56	5		
20	31-57	yes	yes	no	no	yes	yes	yes	yes	37	6		
21	57-83	yes	yes	no	no	yes	yes	no	yes	62	7		
22	57-83	yes	yes	no	no	yes	yes	yes	yes	60	9		
23	57-83	yes	yes	no	no	yes	yes	no	yes	58	5		
27	31-57	VAC	Vac	no	no	Vac	Vec	VAC	Vec	/1	6		
	<ul> <li>→ hea</li> </ul>	Ithdemo+	recoded	(+)								Þ	

# **Big Data – Performance**

**DIONE** uses efficient R algorithms to load, process and save data, which is quite fast for datasets of up to some 1 million rows of data (also known as cases or records), even on an average laptop PC.

However, when memory and processing power are limited, these tasks become more timeconsuming. Some typical times are given here:

computer	processor Intel i7, memory 4 GB					
number of variables	31					
number of cases	10 000 000					
loading data	2 minutes					
checking complete cases	3 minutes					
number of complete cases	1 500 000					
learning network structure	17 minutes					
fitting parameters	6 minutes					

## References

Judea Pearl & Dana Mackenzie 2018. *The Book of Why – The New Science of Cause and Effect*. Allen Lane, Great Britain.