SMARTVISION Europe

Premier Business Partner IBM

Authorized Software Value Plus
Business Analytics

Business Analytics Award
Winner 2012

Maximise Your Investment In SPSS

15th May 2015 – Royal Exchange, London

Agenda

0900	Welcome and Introduction To Smart Vision Europe
0915	What's available in each SPSS module? How could each be of value in your own organisation?
	What's new in SPSS v23.0 – an overview of the additional functionality included within the latest release of SPSS.
1015	Break for Tea & Coffee
1030	Getting more from SPSS – best practice, effective working and avoiding common pitfalls
	Automating and Extending capabilities within SPSS
	Training for SPSS – access to specialist support and training options available to both new and experienced users
1115	Summary, Q&A and Close

SMARTVISION Europe

Predictive Analytics for Smarter Business

Premier Business Partner IBM

- Premium, accredited partner to IBM specialising in the SPSS Advanced Analytics suite.
- Team each has 15 to 20 years of experience working in the analysis, statistics & predictive analytics sector - specifically as senior members of the heritage SPSS team

Maximise your investment in SPSS Statistics

Jarlath Quinn – Analytics Consultant

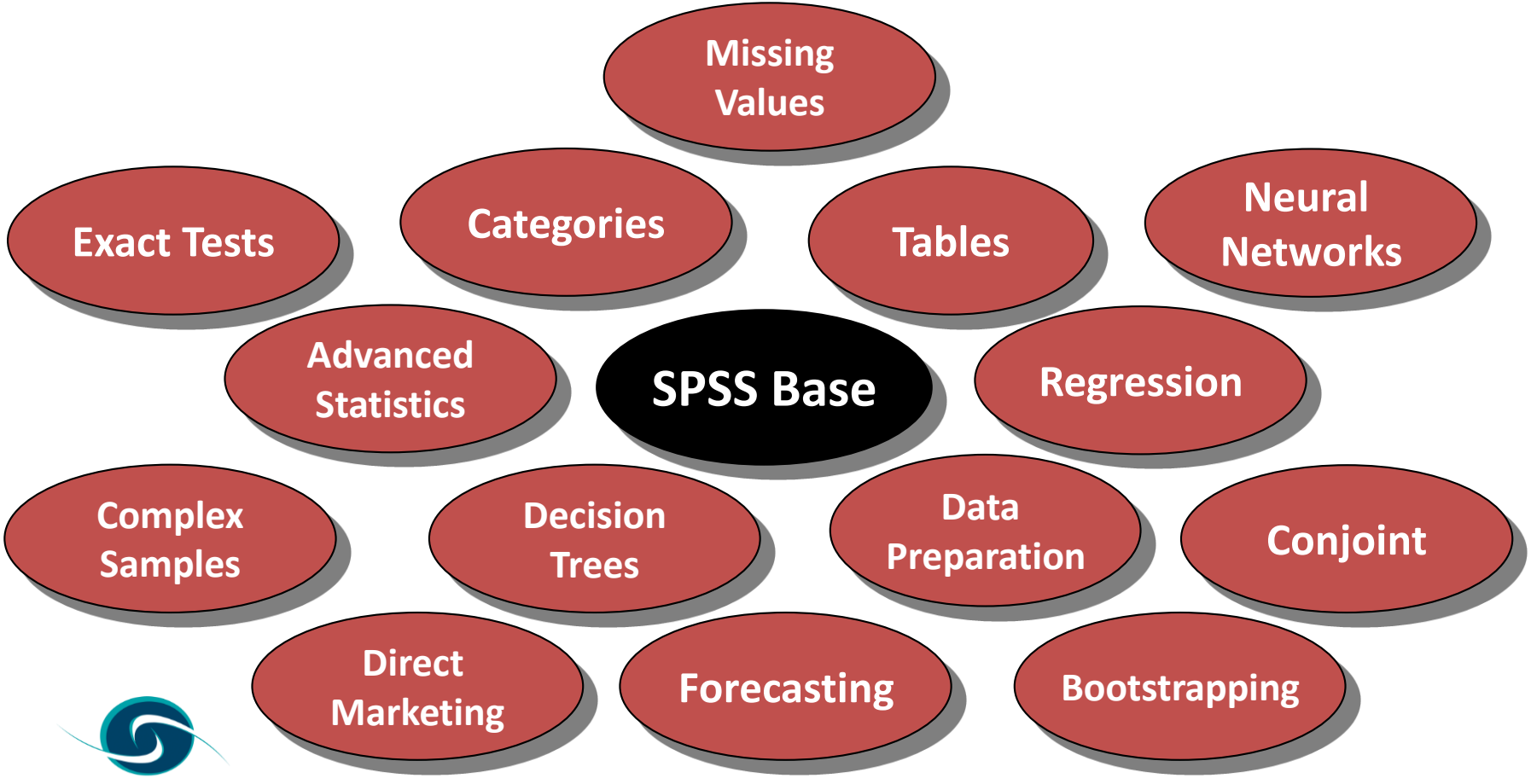
An Overview of IBM SPSS Statistics: Add-on Modules

SPSS Statistics

- Since 1968 one of the world's most popular data analysis and statistical interrogation platforms
- Used for everything from epidemiology studies , survey research and business reporting to direct marketing, credit risk, predictive modelling and asset management

- Statistics included in the base software:
  - Descriptive statistics: Cross tabulation, Frequencies, Descriptives statistics
  - Statistical Tests: T-test, ANOVA, Correlation
  - Prediction for numerical outcomes: Linear regression
  - Prediction for identifying groups: Factor analysis, Cluster analysis

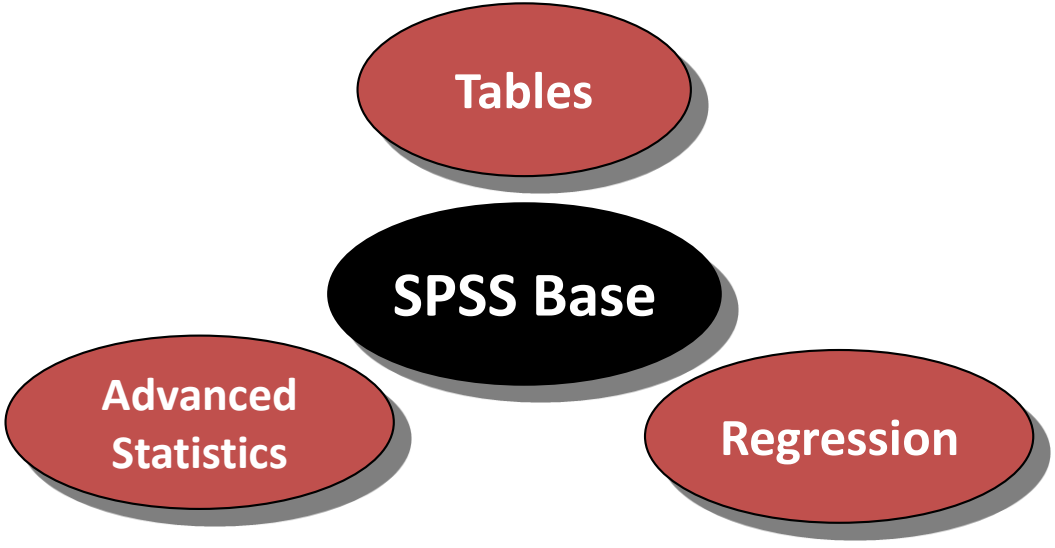
IBM SPSS Statistics Base & Associated Modules



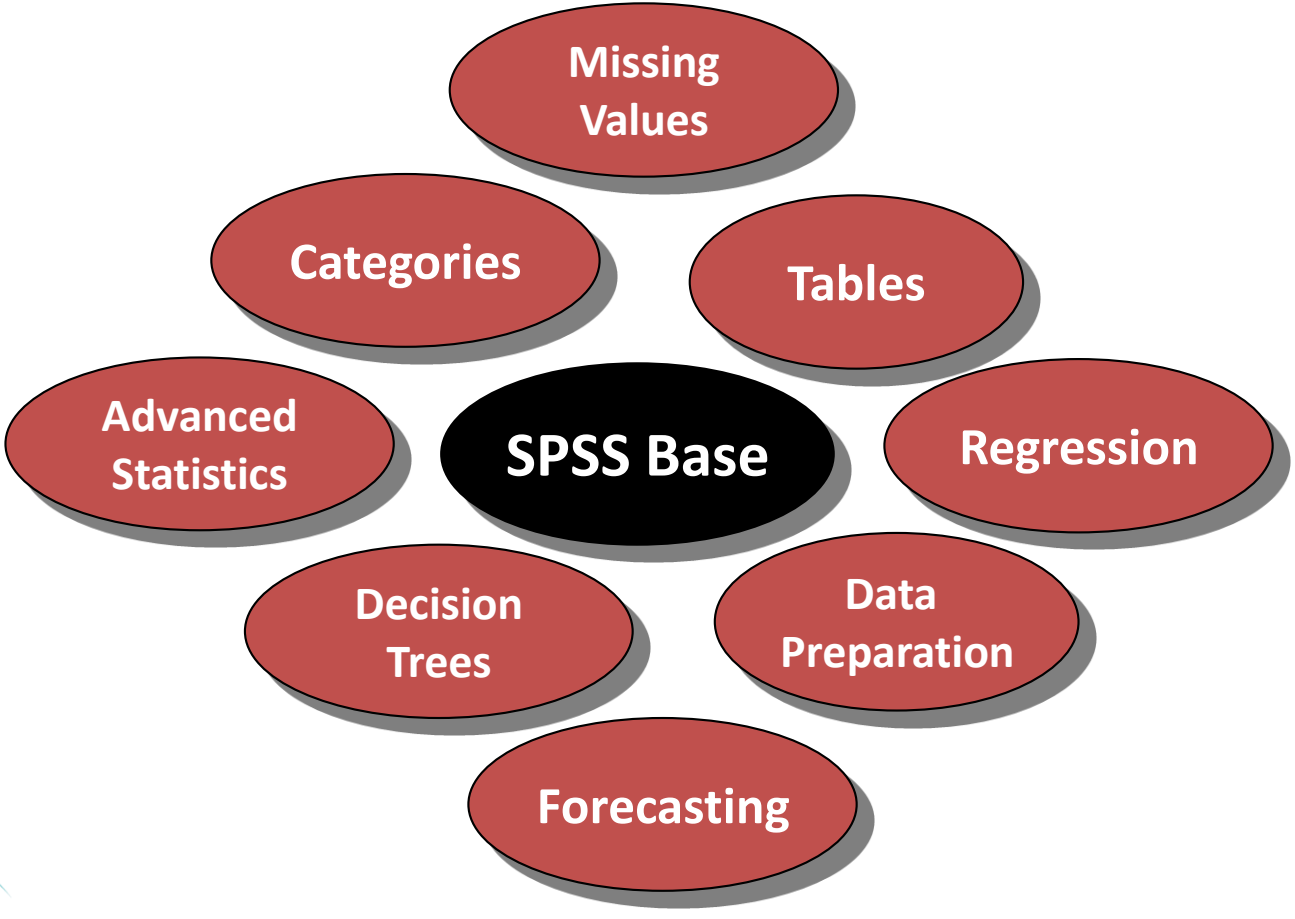
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

IBM SPSS Statistics Standard



IBM SPSS Statistics Professional

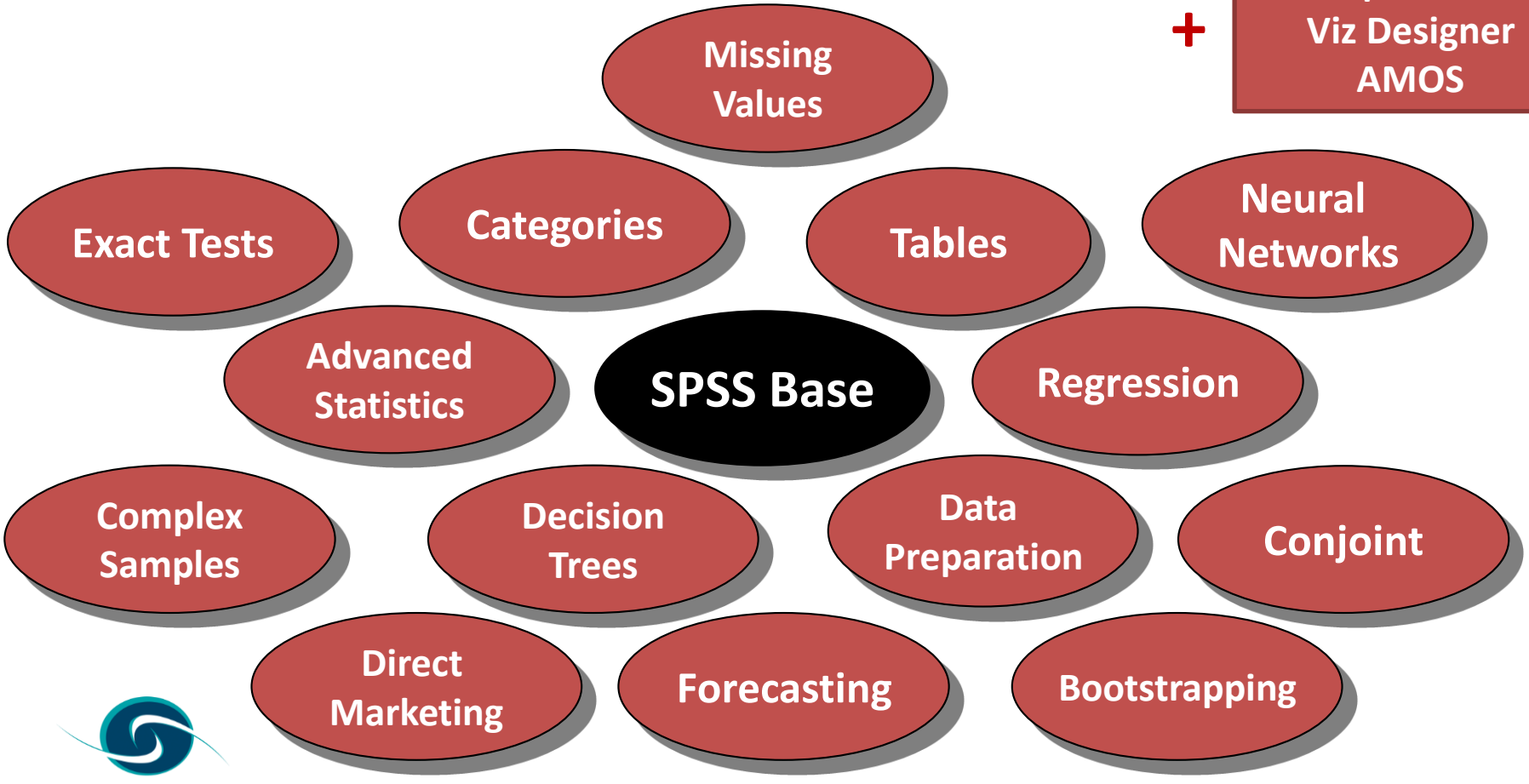


SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

IBM SPSS Statistics Premium

+ Sample Power
Viz Designer
AMOS



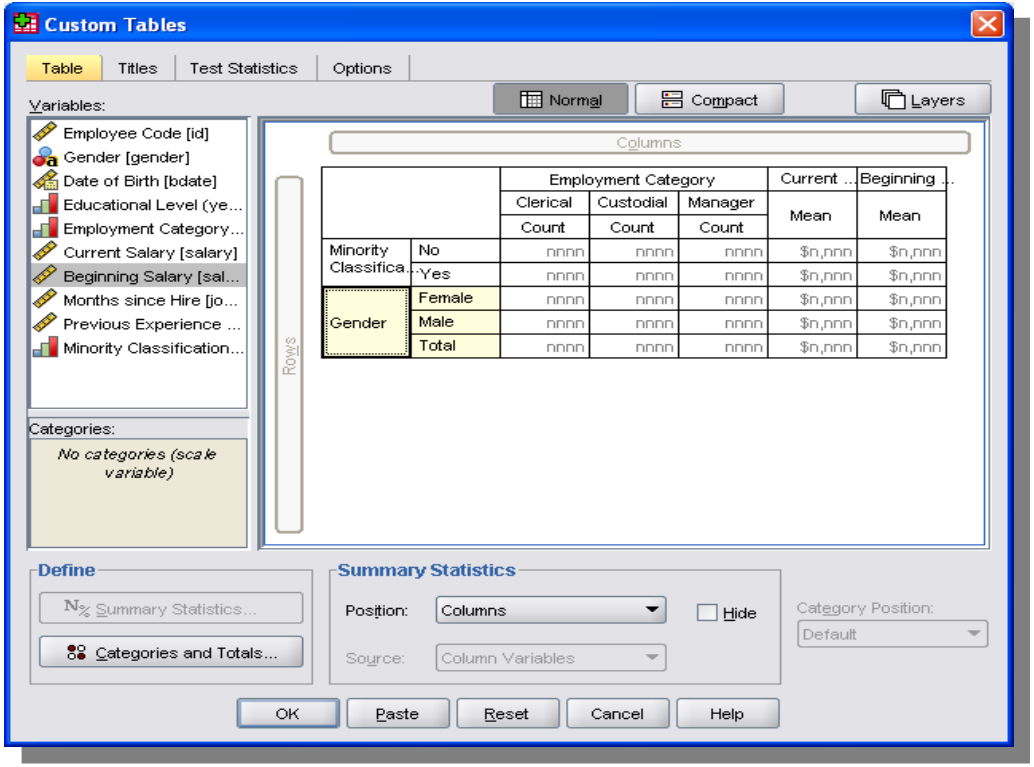
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

SMARTVISION
Europe

IBM SPSS Custom Tables

IBM SPSS Custom Tables



- A single dialog box for all tabulation and reports
- Preview the table as you drag and drop the fields into the rows and columns

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

IBM SPSS Custom Tables

Table 1

		Employment Category			Current Salary	Beginning Salary
		Clerical	Custodial	Manager		
		Row Percent	Row Percent	Row Percent	Mean	Mean
Minority Classification	No	75%	4%	22%	$36,023	$17,673
	Yes	84%	12%	4%	$28,714	$14,679
Gender	Female	95%	0%	5%	$26,032	$13,092

Table 1

		Employment Category			Current Salary	Beginning Salary
		Clerical	Custodial	Manager		
		Row Percent	Row Percent	Row Percent	Mean	Mean
Minority Classification	No	75%	4%	22%	$36,023	$17,673
	Yes	84%	12%	4%	$28,714	$14,679
Gender	Female	95%	0%	5%	$26,032	$13,092
	Male					
	Total					

Table 1

		Employment Category			Current Salary	Beginning Salary
		Clerical	Custodial	Manager		
		Row Percent	Row Percent	Row Percent	Mean	Mean
Minority Classification	No	75%	4%	22%	$36,023	$17,673
	Yes	84%	12%	4%	$28,714	$14,679
Gender	Female	95%	0%	5%	$26,032	$13,092
	Male	61%	10%	29%	$41,442	$20,301
	Total	77%	6%	18%	$34,420	$17,016

- **Multiple Table Looks**

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

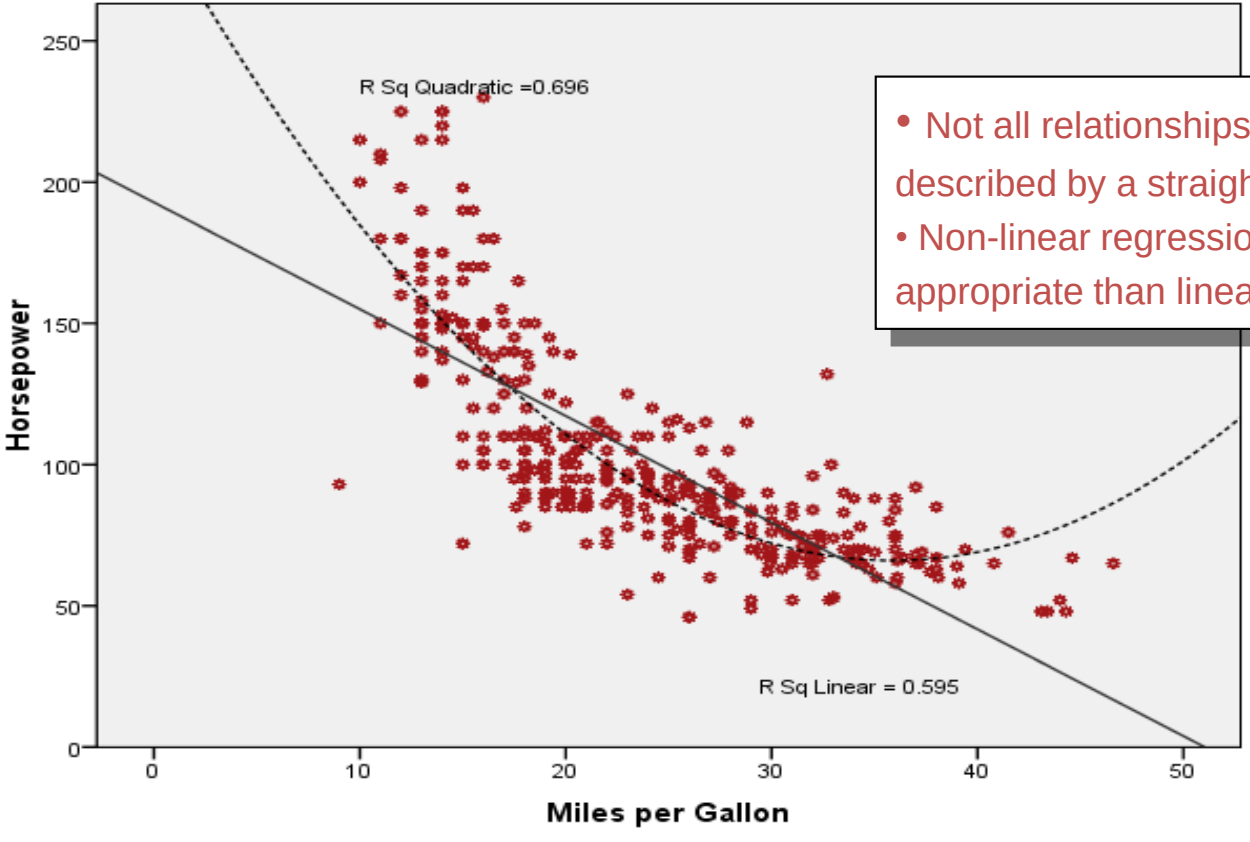
SMARTVISION
Europe

SPSS Regression Models

SPSS Regression Models

- The SPSS *Regression Models* module contains a wide range of nonlinear regression models that augment the linear regression functionality in SPSS Base.
- *Regression Models* is a family of classical predictive techniques all of which involve fitting (or *regressing*) a line or curve to a series of observations in order to model effects or predict outcomes.
- *Regression Models* is often used in situations where the Linear Regression functionality in SPSS base is either inappropriate or is too simplistic
- *Logistic Regression* is a very widely-used technique for predicting categorical outcomes. In *Regression Models* there are two forms of Logistic regression:
  - Binary Logistic – for predicting 2 category outcomes
  - Multinomial Logistic – for predicting more than 2 category outcomes
- Regression Models also contains:
  - Nonlinear regression and Constrained Nonlinear Regression
  - Probit, Weighted Least Squares and Two Stage Least Squares

SPSS Regression Models



- Not all relationships can be adequately described by a straight line
- Non-linear regression is sometimes more appropriate than linear regression

SMARTVISION
Europe

SPSS Advanced Models

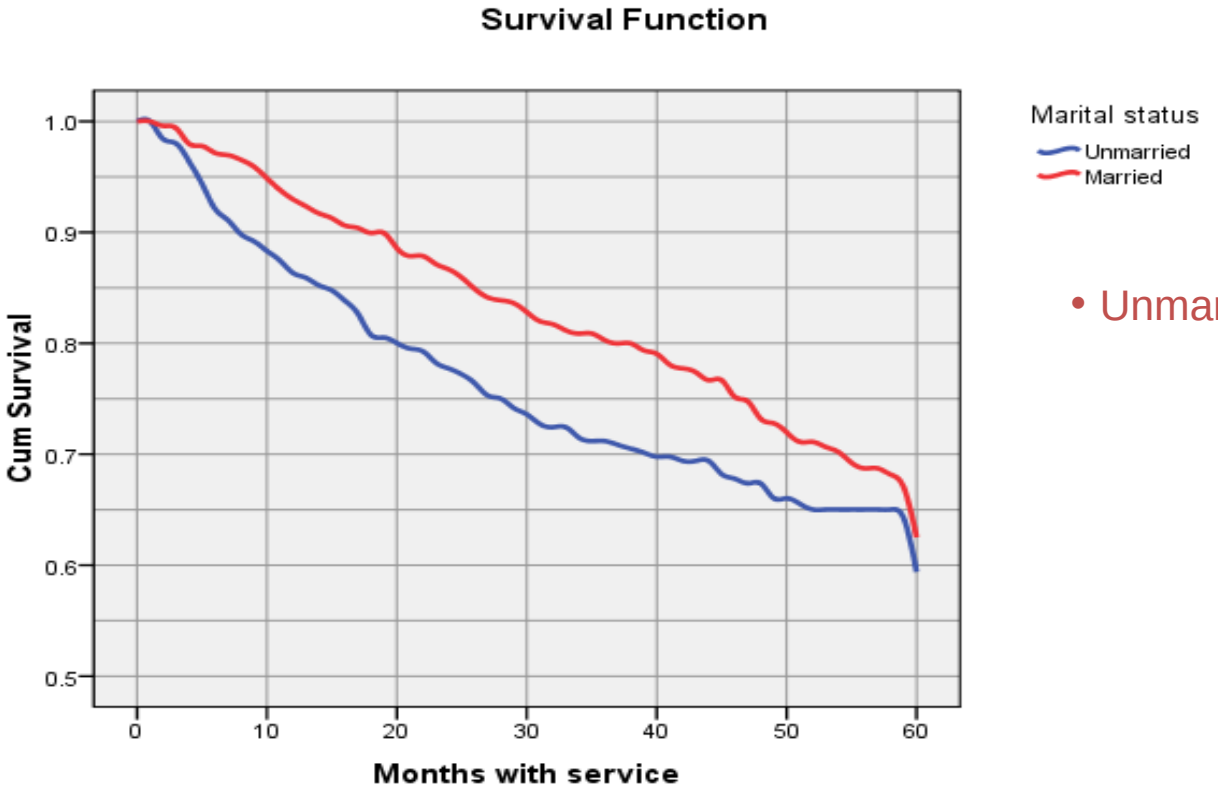
SPSS Advanced Models

- *Advanced Models* is the most 'analytically rich' SPSS module. *Advanced Models* includes a very wide range of multivariate procedures for investigating complex relationships in data.
- A number of the procedures in *Advanced Models* are relatively technical in a statistical sense. In particular, *Advanced Models* encompasses General and *Generalized* Linear Modelling capabilities.
- General Linear Models allow you to model relationships and interactions between many factors. The general linear model incorporates a number of different statistical models: ANOVA, MANOVA, ANCOVA, Repeated Measures etc.
- Generalized Linear Models are an extension of General Linear Models in that they are able to work with a greater range of data distributions. In particular, the model allows for the dependent variable to have a non-normal distribution.
- The Generalized Estimating Equations (GEE) procedure extends the generalized linear model to allow for analysis of repeated measurements or other correlated observations, such as clustered data.

SPSS Advanced Models

- *Advanced Models* also includes *Linear Mixed Models*. If you work with data that display correlation and non-constant variability, such as nested data that represent students within faculties or employees within departments, you can use the linear mixed models procedure to model means, variances, and covariances in your data.
- *Advanced Models* includes General Loglinear and LOGIT Loglinear analysis.
- *Advanced Models* also includes a number *Survival Analysis* algorithms. In recent times, Survival Analysis has also been used in application such as insurance claims and customer churn.
- *Advanced Models* offers 4 distinct Survival Analysis procedures:
  - **Life Tables**
  - **Kaplan-Meier**
  - **Cox Regression**
  - **Cox Regression with time-dependent covariate**

SPSS Advanced Models



- Unmarried customers churn sooner

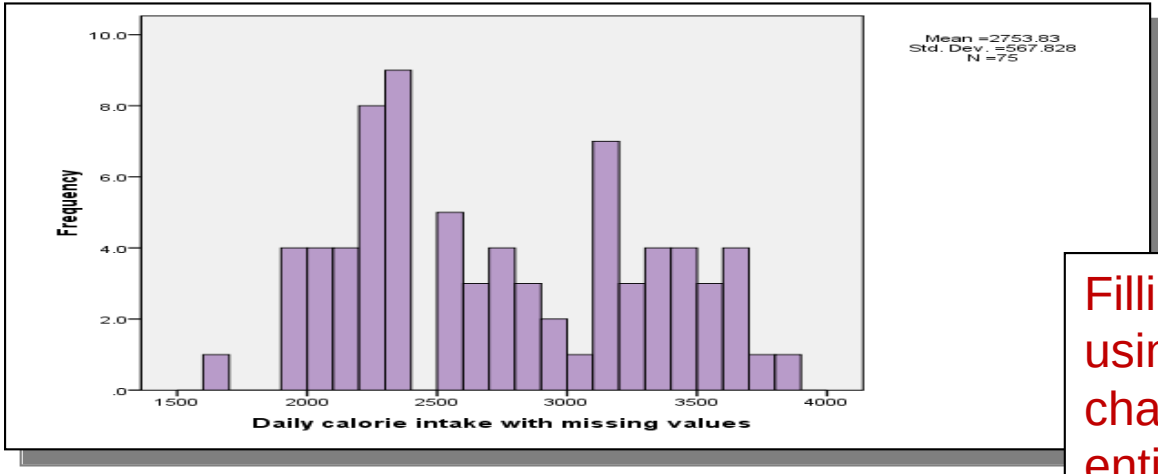
SMARTVISION Europe

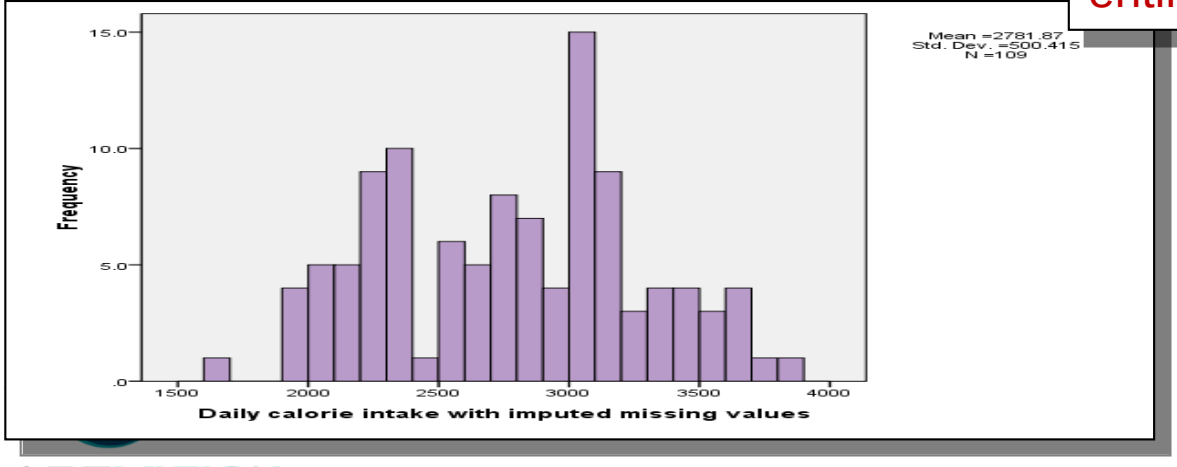
SPSS Missing Values

SPSS Missing Values

- The *Missing Values* module procedure provides three main functions:
  - It describes the pattern of missing data. Where are the missing values located? How extensive are they? Are values missing randomly?
  - Provides estimates of statistics like means, standard deviations and correlations for data series that contain missing values.
  - Fills in (imputes) missing data with estimated values using special methods like regression or EM (expectation-maximization).
- The *Missing Values* module helps address several concerns caused by incomplete data. By investigating patterns of missing data it can address questions such as 'Why are the data missing?'. The means estimation procedures address questions such as 'How does the missing data affect summary statistics?'

SPSS Missing Values





Filling in the missing data using imputation can change the shape of an entire distribution

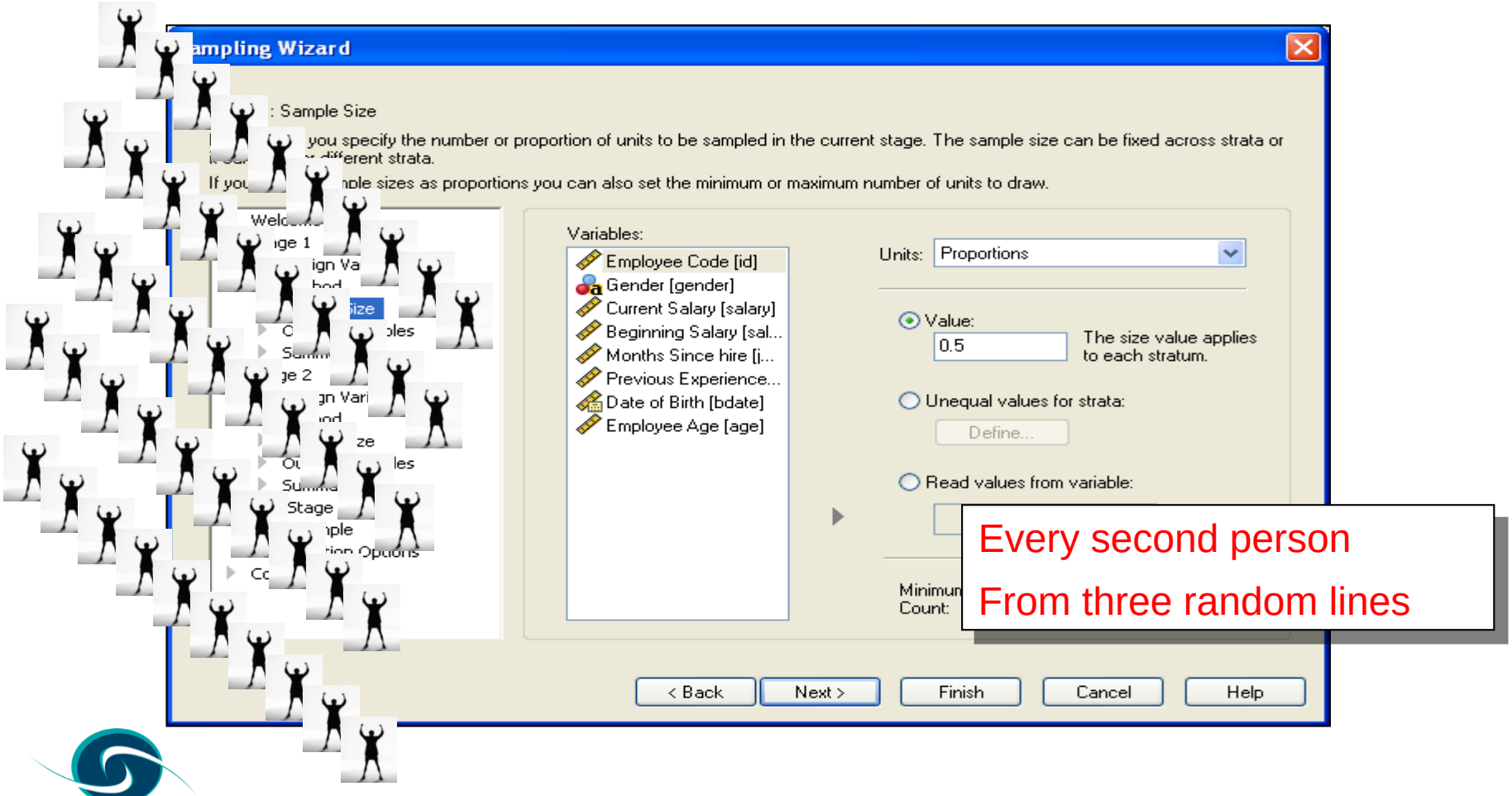
SMARTVISION
Europe

SPSS Complex Samples

SPSS Complex Samples

- An inherent assumption of many classical statistical procedures is that the data represents a *simple random sample* drawn from the population of interest.
- The SPSS *Complex Samples* module allows users to draw samples that are more complicated than simple random schemes.
- *Complex Samples* also allows statistical analyses to be carried out that *take account* of the complex sampling scheme used in collecting the data.
- An example of this would be carrying out a chi square test to see if larger households are more likely to recycle glass than smaller households. Using *Complex Samples*, the researchers could calculate a more appropriate test statistic based on a sample where every 3rd house was sampled from a random selection of 50 streets in a town.

SPSS Complex Samples



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

SMARTVISION Europe

SPSS Exact Tests

SPSS Exact Tests

- The *SPSS Exact Tests* module provides additional methods for calculating the significance levels for the statistical tests available through the *Crosstabs* and the *Nonparametric Tests* menus.
- Using the standard tests in SPSS Base (known as asymptotic tests) can lead to misleading or inaccurate results when working with small datasets or sparse groups in the sample data. *Exact Tests* enables users to obtain an accurate significance level without relying on assumptions that might not be met by the data.

- *Exact Tests* offers two extra methods of calculating probabilities on top of the normal asymptotic methods in SPSS Base.

  - Monte Carlo Estimate: An unbiased *estimate* of the *exact* significance level. This method is most useful when the data set is too large to compute exact significance but the data do not meet the assumptions of the asymptotic method.

  - Exact: The probability of the observed outcome or an outcome more extreme is calculated exactly.

SPSS Exact Tests

Employment Category * Minority Classification Crosstabulation

Count

		Minority Classification		
		No	Yes	Total
Employment Category	Clerical	17	12	29
	Custodial	7	6	13
	Manager	8	2	10
Total		32	20	52

Relatively small group sizes

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	1.869[a]	2	.393	.440		
Likelihood Ratio	2.004	2	.367	.411		
Fisher's Exact Test	1.819			.440		
Linear-by-Linear Association	.937[b]	1	.333	.374	.217	.092
N of Valid Cases	52					

a. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 3.85.

b. The standardized statistic is -.968.

Chi Square showing exact probabilities highlighted in red

SMARTVISION Europe

MPANY

SMARTVISION
Europe

SPSS Decision Trees

SPSS Decision Trees

- Decision trees are used *extensively and widely* within Predictive Analytics
- Decision trees can be used to
  - Build profiles of customers/employees/clients
  - Find key behavioural segments
  - Generate predictive models
- Decision Trees can be expressed as a series of hierarchical rules which means that they can be converted in languages like SQL for database scoring
- Decision Trees are especially popular because
  - they are fairly visual representations of models
  - relatively easy to understand

Understanding Decision Trees – a worked example

- What were the most important factors determining survival during the sinking of the RMS Titanic?

Survival on the RMS Titanic

		Count	Percent %
survive	Did not survive	1490	68%
	Survived	711	32%
	Total	2201	100%

Gender?

Age?

Class?

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		sex			
		female		male	
		Count	Column Percent %	Count	Column Percent %
survive	Did not survive	126	26.8%	1364	78.8%
	Survived	344	73.2%	367	21.2%
	Total	470	100.0%	1731	100.0%

Pearson Chi-Square Tests

		sex
survive	Chi-square	456.874
	df	1
	Sig.	.000*

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		age			
		adult		child	
		Count	Column Percent %	Count	Column Percent %
survive	Did not survive	1438	68.7%	52	47.7%
	Survived	654	31.3%	57	52.3%
	Total	2092	100.0%	109	100.0%

Pearson Chi-Square Tests

		age
survive	Chi-square	20.956
	df	1
	Sig.	.000*

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		class							
		1st		2nd		3rd		crew	
		Count	Column Percent %	Count	Column Percent %	Count	Column Percent %	Count	Column Percent %
survive	Did not survive	122	37.5%	167	58.6%	528	74.8%	673	76.0%
	Survived	203	62.5%	118	41.4%	178	25.2%	212	24.0%
	Total	325	100.0%	285	100.0%	706	100.0%	885	100.0%

Pearson Chi-Square Tests

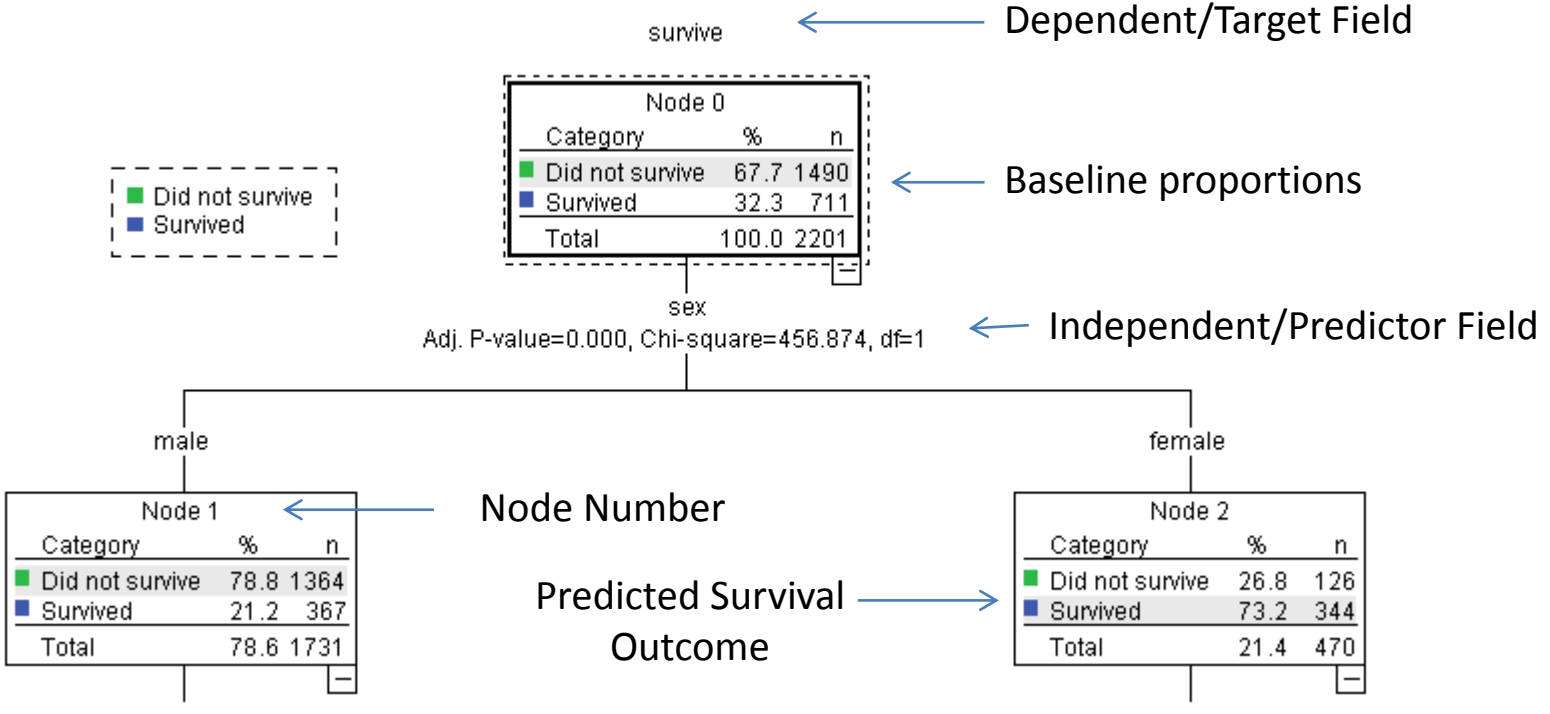
		class
survive	Chi-square	190.401
	df	3
	Sig.	.000*

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

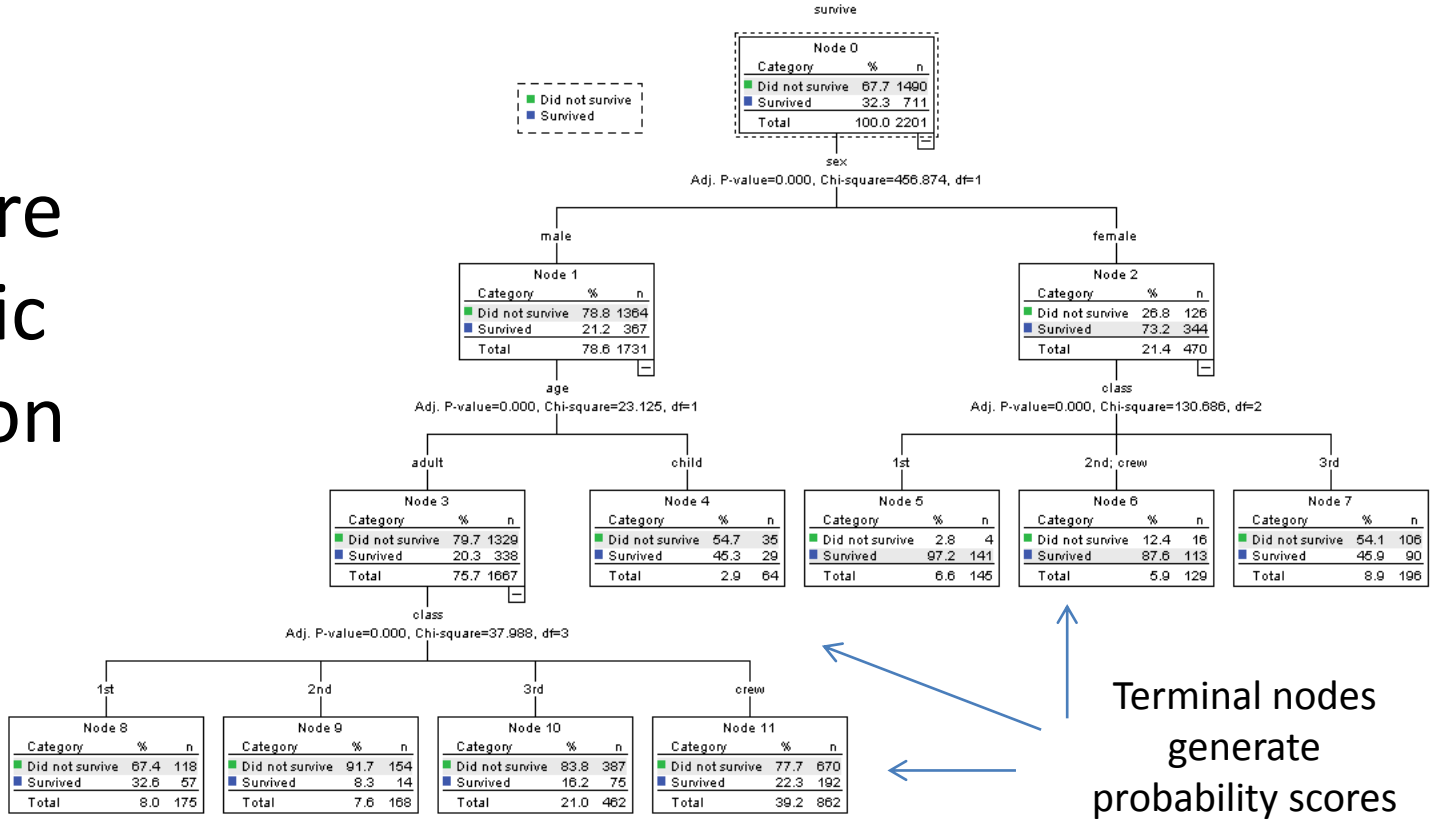
Gender is most important

…and a CHAID Decision tree will reflect this….



Full CHAID Decision Tree

C.H.A.I.D
Chi-Square
Automatic
Interaction
Detector



Terminal nodes generate probability scores

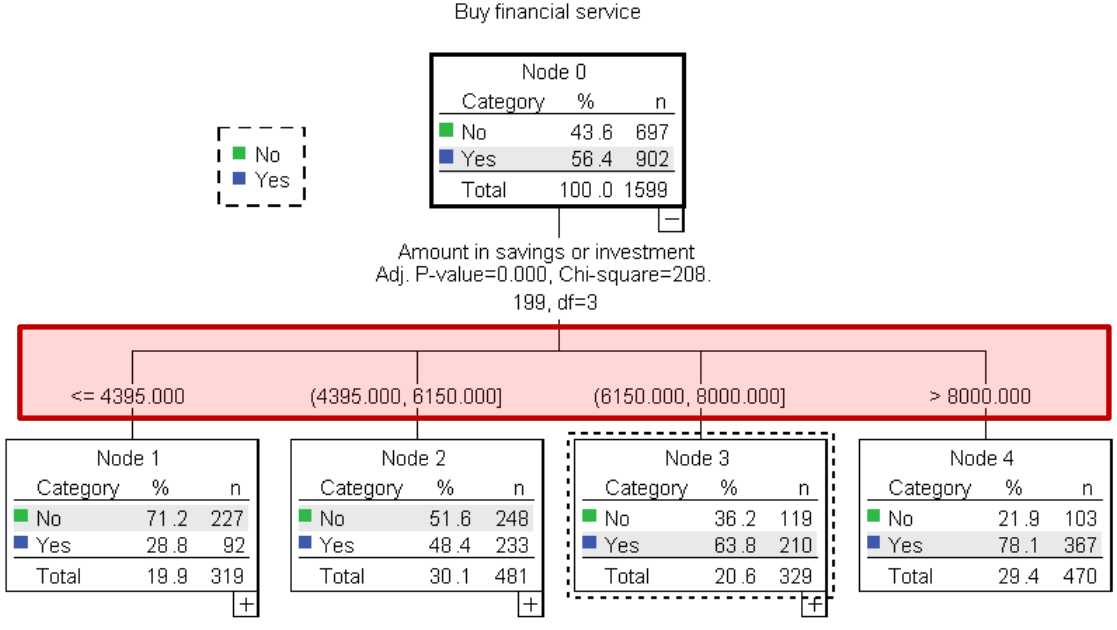
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Merging/Splitting in CHAID Trees

Decision Trees can merge values of numeric *and* categorical predictors together

This makes the tree more efficient and easier to read

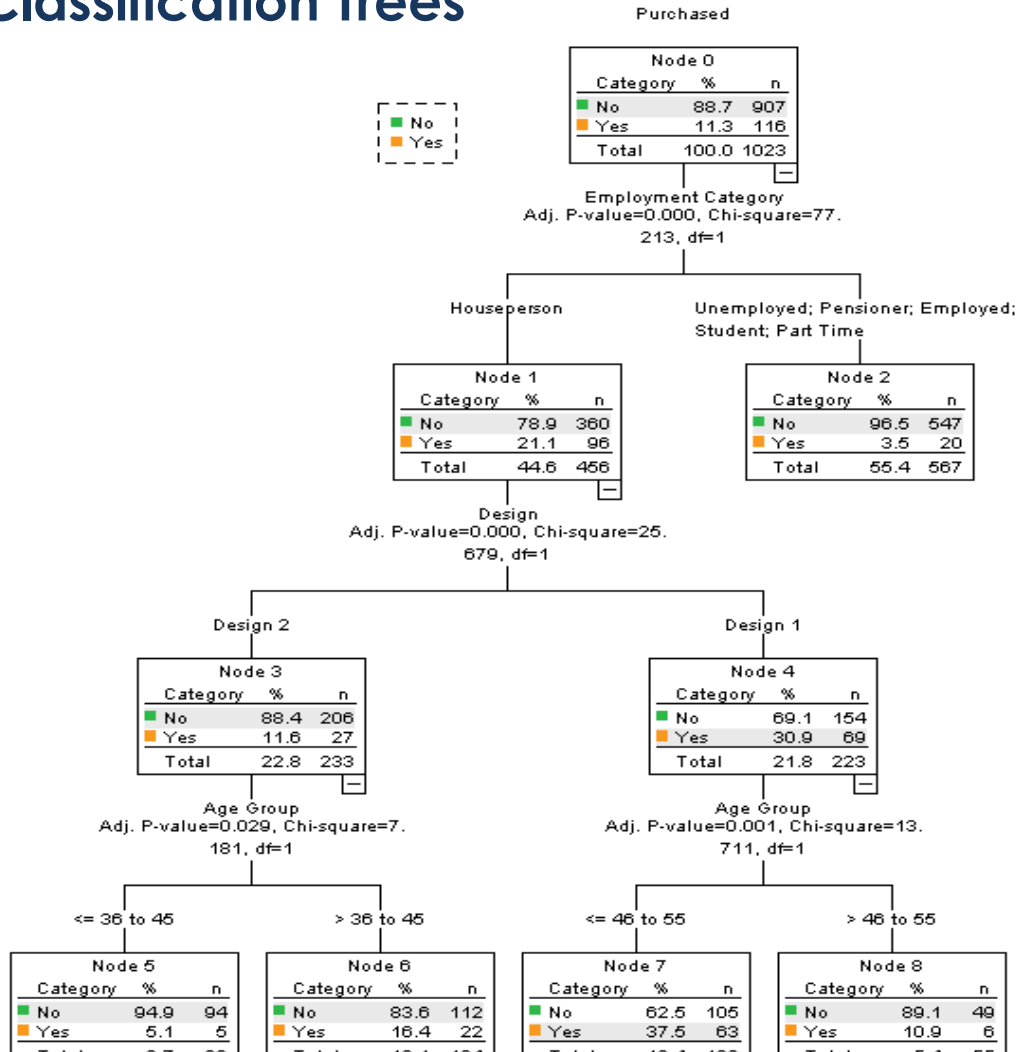


SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

SPSS Classification Trees

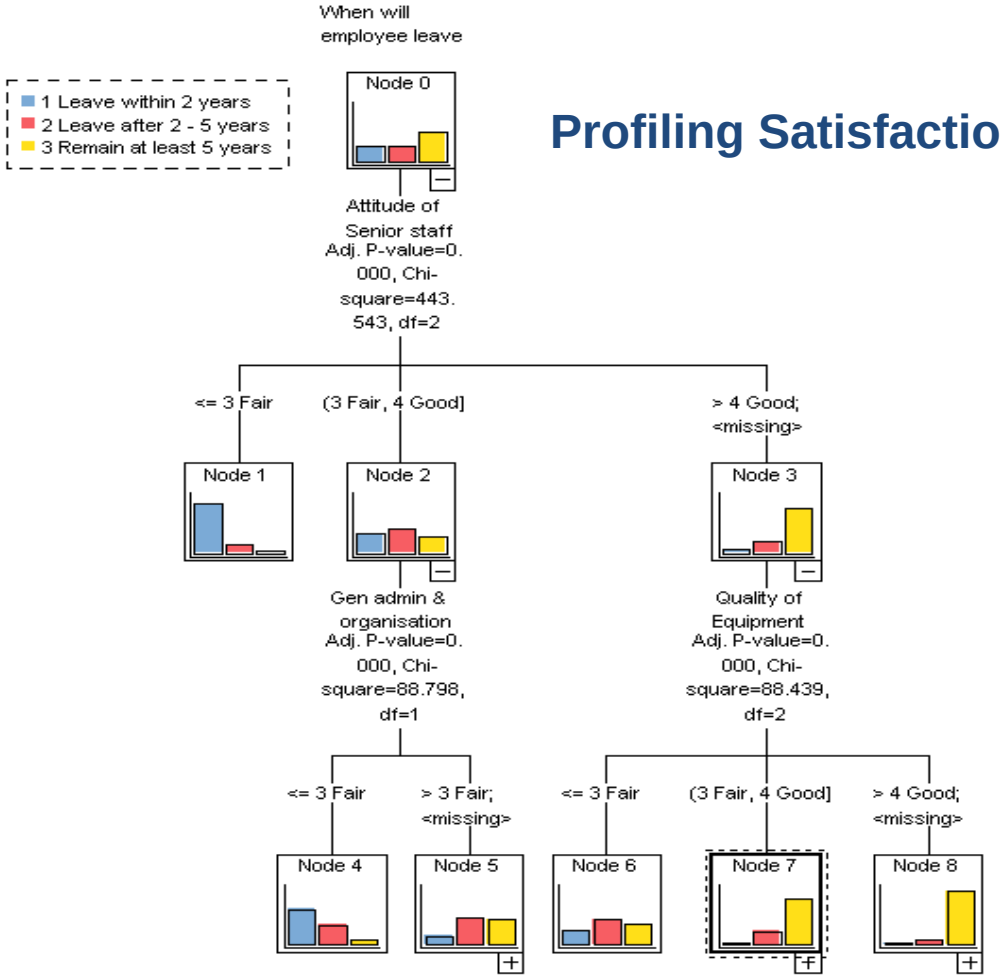
- Predicting
  - Customer Churn
  - Marketing Response
  - Fraud
  - Cross Sell
  - Asset Failure



SMARTVISION Europe

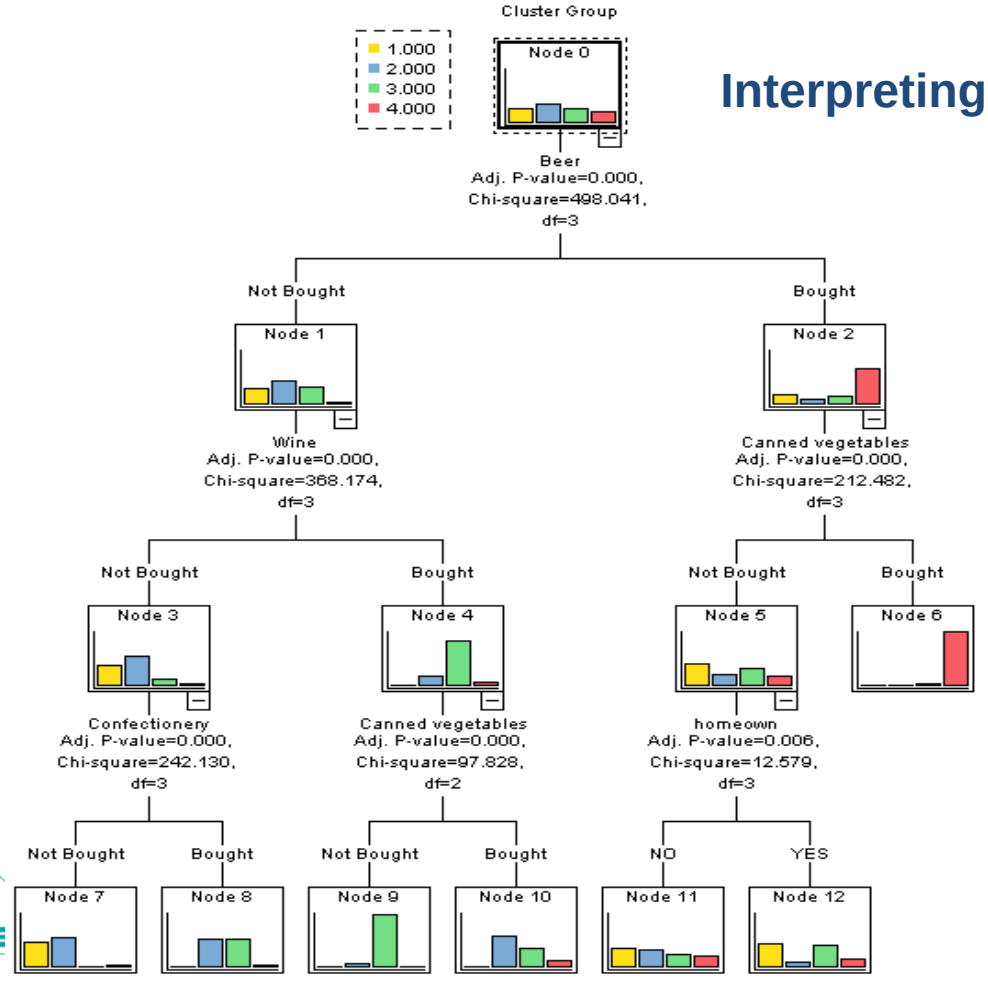
SPSS Classification Trees

Profiling Satisfaction



SPSS Classification Trees

Interpreting Cluster Membership



SMARTVIS Europe

A SELECT INTERNATIONAL COMPANY

SMARTVISION
Europe

SPSS Direct Marketing

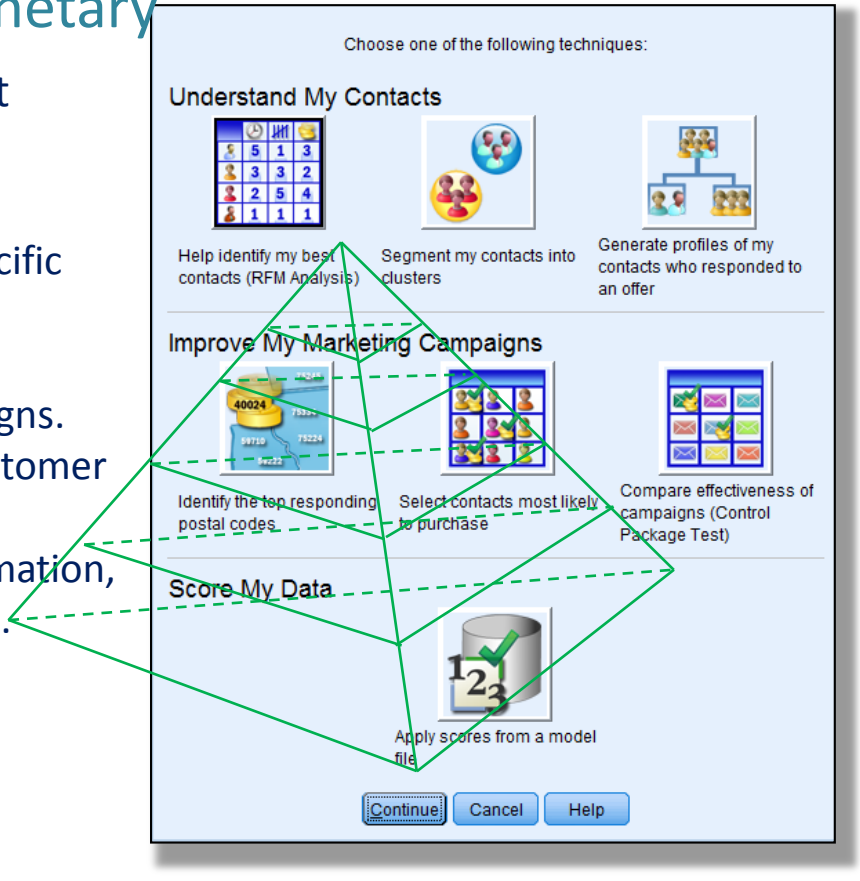
SPSS Direct Marketing

- RFM – Recency, Frequency, Monetary
- 3 dimensions of loyalty

SPSS Direct Marketing enables database and direct marketers to:

- specific
- ns.
- npaigns.
- customer
- formation,
- yses.

Recency	Frequency	Monetary
5	5	5
4	5	5
4	4	5
4	4	4
3	4	4
3	3	4
3	3	3
2	3	3
2	2	3
2	2	2
1	2	2
1	1	2
1	1	1



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

SMARTVISION Europe

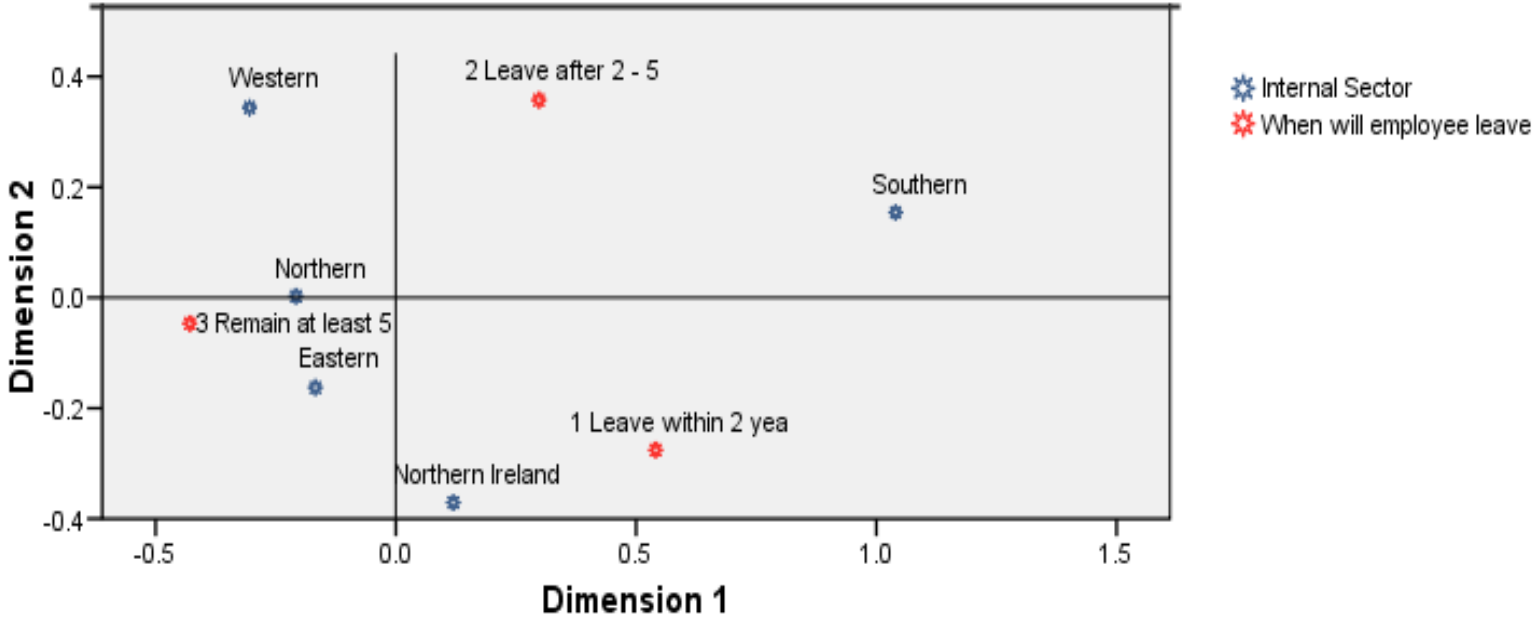
SPSS Categories

SPSS Categories

- The *SPSS Categories* module provides a number of algorithms based on a family of techniques called *optimal scaling*.
- Optimal scaling attempts to *quantify the* category groups of categorical fields i.e. assign numerical values to the categories *as if they existed on a scale.*
- By quantifying categories can be used as excellent exploratory tools when modelling multivariate categorical data.
- Examples of techniques include:
  - Correspondence Analysis
  - Categorical Regression
  - Categorical PCA

SPSS Categories

- Quantifying the two categorical fields '*Internal Sector*' and '*When will employee leave*' helps us to explore the relationship between the two variables as if they were continuous fields in a scatterplot



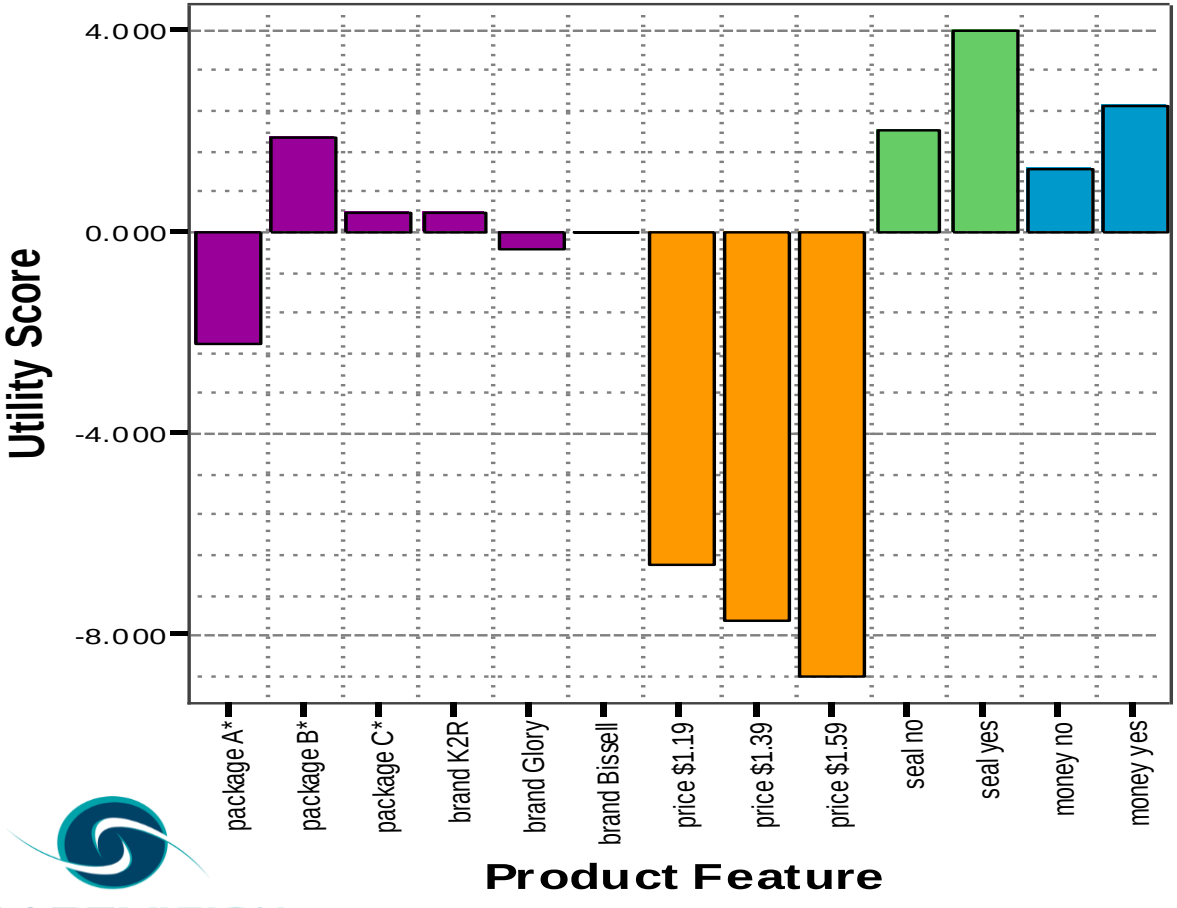
SMARTVISION
Europe

SPSS Conjoint

SPSS Conjoint

- *Conjoint* analysis is a technique pioneered by market research analysts to determine how people value the different features that make up an individual product or service.
- *Conjoint* analysis can be used to discover the optimal combination of product/service attributes in terms of the combination that is most influential on customer choice or decision making.
- *Conjoint* works by showing respondents a particular set of products (or services) and by analysing how they make preferences between these products.
- By mapping the different features or aspects of the products to the choices that the respondent makes, the *Conjoint* technique is able to infer the ideal set of characteristics for a product or service.

SPSS Conjoint



Results of a Conjoint analysis showing utility (preference) scores for different aspects of a cleaning product

Note: Conjoint analysis in SPSS is primarily run via SPSS syntax

SMARTVISION Europe

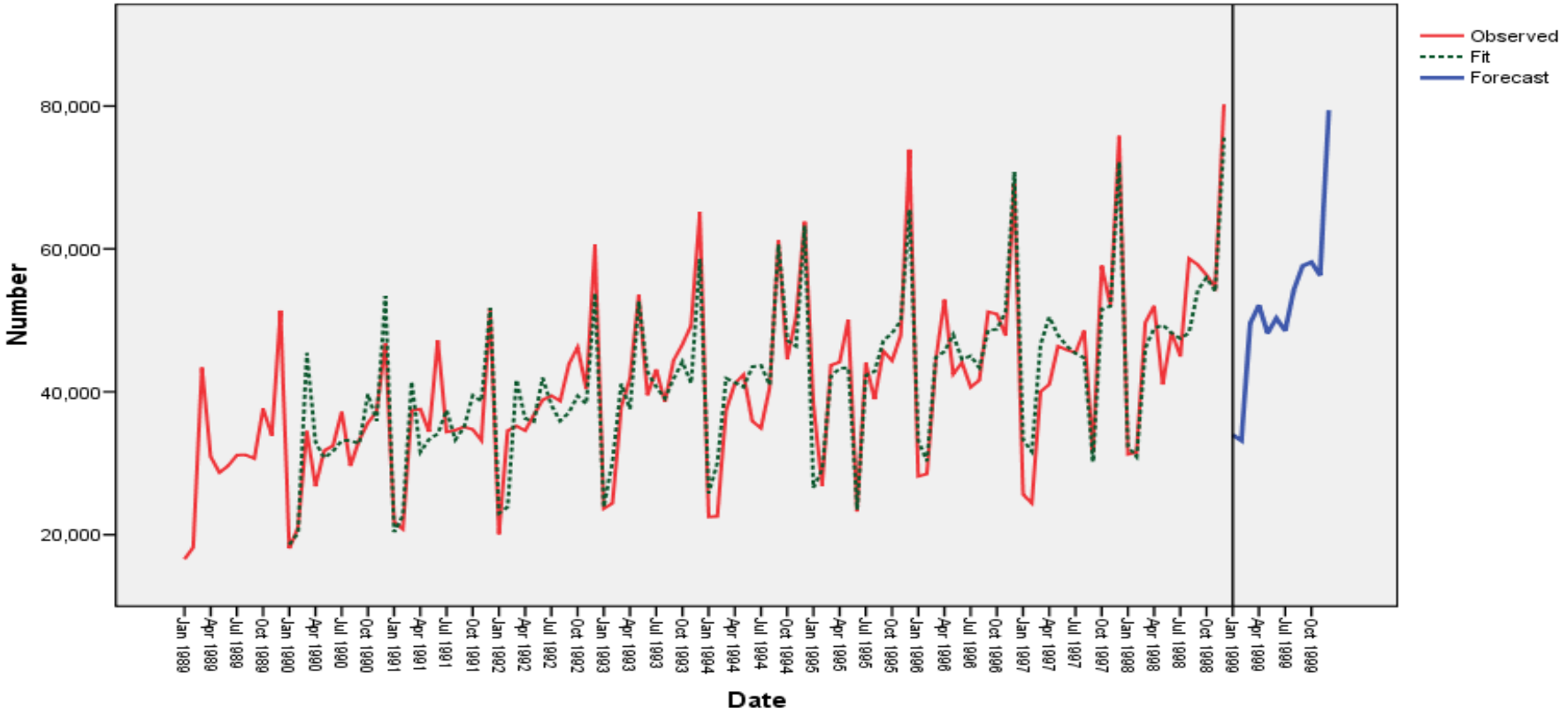
SPSS Forecasting

SPSS Forecasting

- *SPSS Forecasting* is the SPSS *time series* module. Time series forecasting is the use of a model to predict future events based on known past events.
- Examples of time series forecasting include:
  - Predicting the number of staff required on each day for a call centre
  - Forecasting the number of patients visiting the accident and emergency department
  - Predicting demand for a gas or electricity supplier
  - Estimating passenger numbers for a train company
- The time factor, is in itself, a predictor of the dependent variable. In other words, in time series, the *past provides a model for the future.*

- *SPSS Forecasting is particularly powerful as it can automatically select and fit a Time Series Model*

SPSS Forecasting

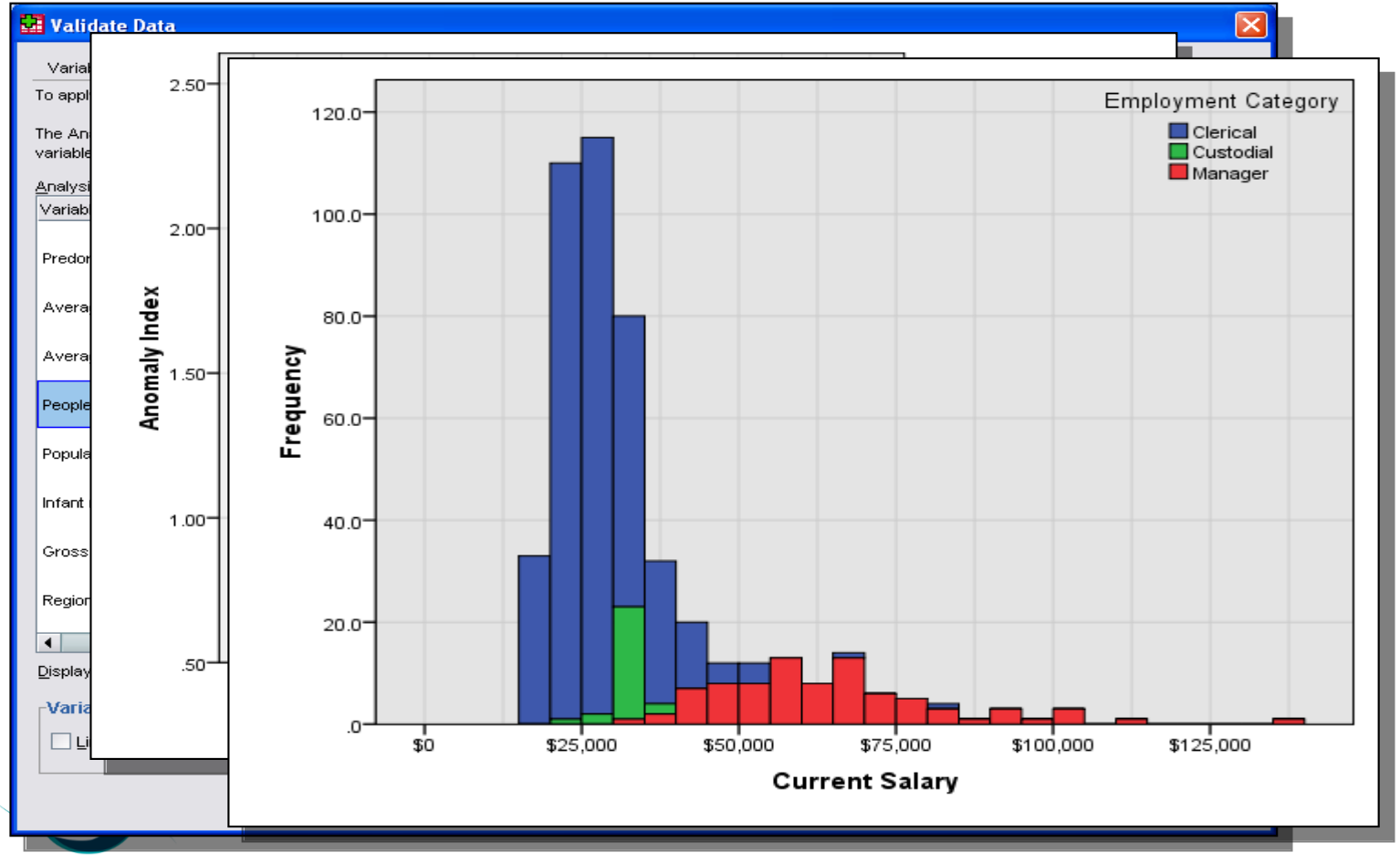


SPSS Data Preparation

SPSS Data Preparation

- The SPSS *Data Preparation* module allows users to identify data errors or unusual cases in their datasets. Using a combination of basic checks, validation rules or anomaly detection algorithms, the *Data Preparation* module will generate new variables or output reports that identify problematic cases or unusual records.
- It can be used to:
  - Identify records with a high percentage of missing values, a high degree of variability or conversely, too little variability as well as incomplete id fields or duplicate records.
  - Provide a graphical overview of each of the fields and the capability to create validation rules for individual fields. An example of this would be a rule that ensures a field is an integer (i.e. no decimal places) such as age.
  - Create rules that ensure that the values in combinations of variables do not contradict each other or imply errors in the data. An example would be a cross-variable rule that ensures that all car drivers are at least 17 years old..

SPSS Data Preparation



SMARTVISION
Europe

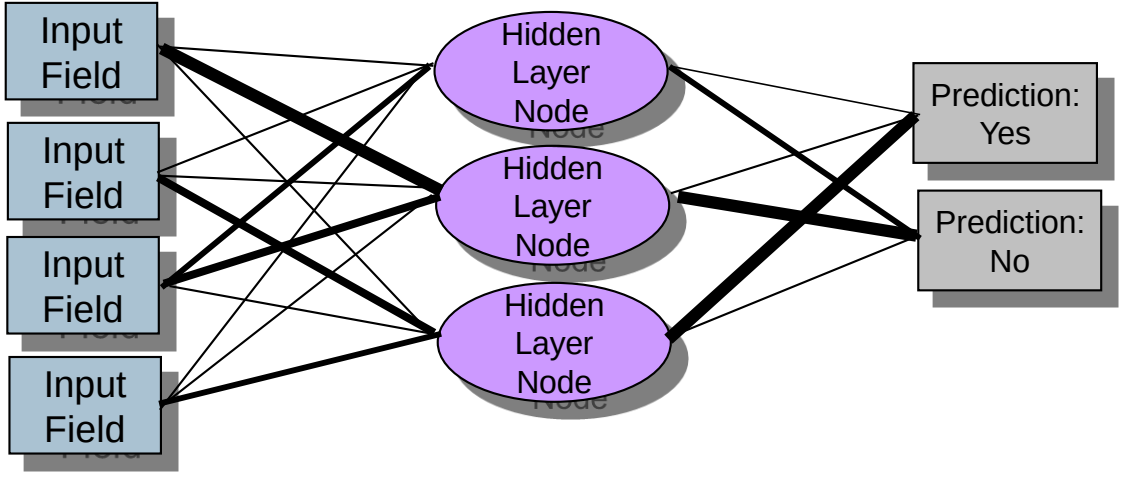
SPSS Neural Networks

SPSS Neural Networks

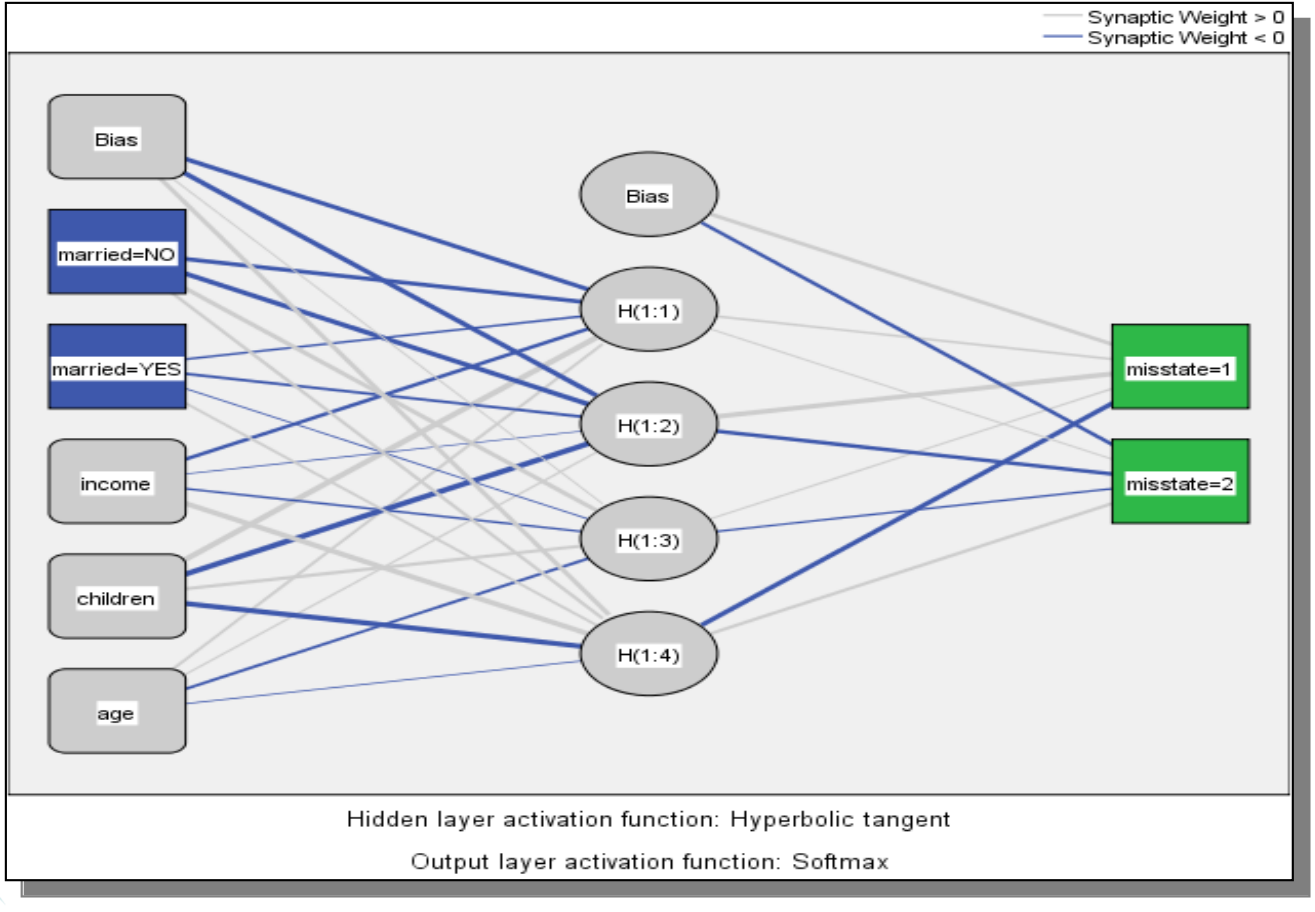
- SPSS *Neural Networks* provides an alternative predictive capability to approaches such as regression or classification trees. Predictive neural networks are particularly useful in applications where the data from the underlying phenomena is complex such as fraud detection, credit scoring and pattern recognition.
- Neural Networks attempt to 'learn' the outcomes of a target field by constantly updating the model with increasingly smaller changes until model accuracy can no longer be improved
- One of the primary advantages of neural networks when compared to classical statistical techniques, is their flexibility and lack of distributional assumptions.

SPSS Neural Networks

- A neural network works by taking the values of predictor or input fields and feeding them into the algorithm as an *input layer*.
- The *input layer* is used to create a *hidden layer* containing unseen nodes (or units) where each node is some function of the input fields (in fact some networks can create more than one hidden layer).
- The *output layer* contains the responses or predictions. The network is continually rebuilt or refined so that the *synaptic weights* in the nodes correctly predict the outcome.



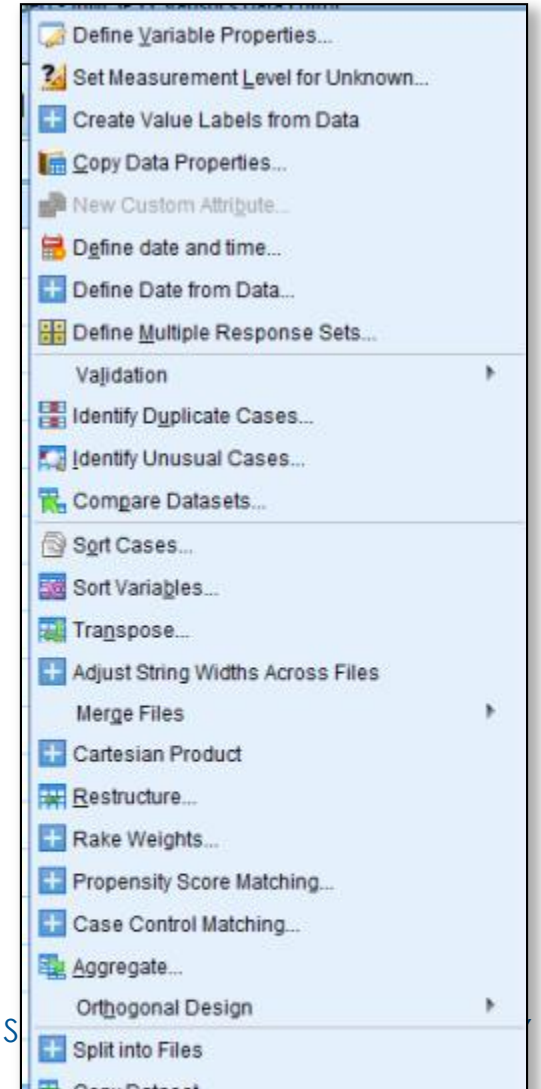
SPSS Neural Networks



What's new in IBM SPSS Statistics v23?

IBM SPSS Statistics v23

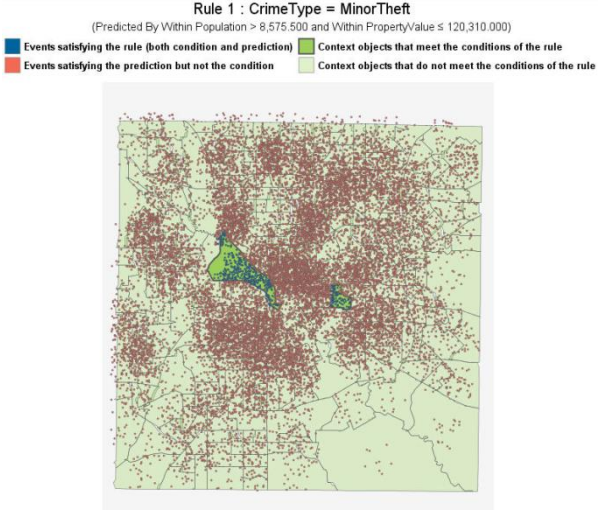
- Loads of enhancements enabled via the Python Essentials Pack (available at installation or via a separate download)
- Examples include –
  - Manage Datasets
  - Read Triple S Data
  - Connect to Internet Data
  - Weibull Plots
  - Anonymize Variables
  - Simulate Active Dataset



SMARTVISION Europe

IBM SPSS Statistics v23

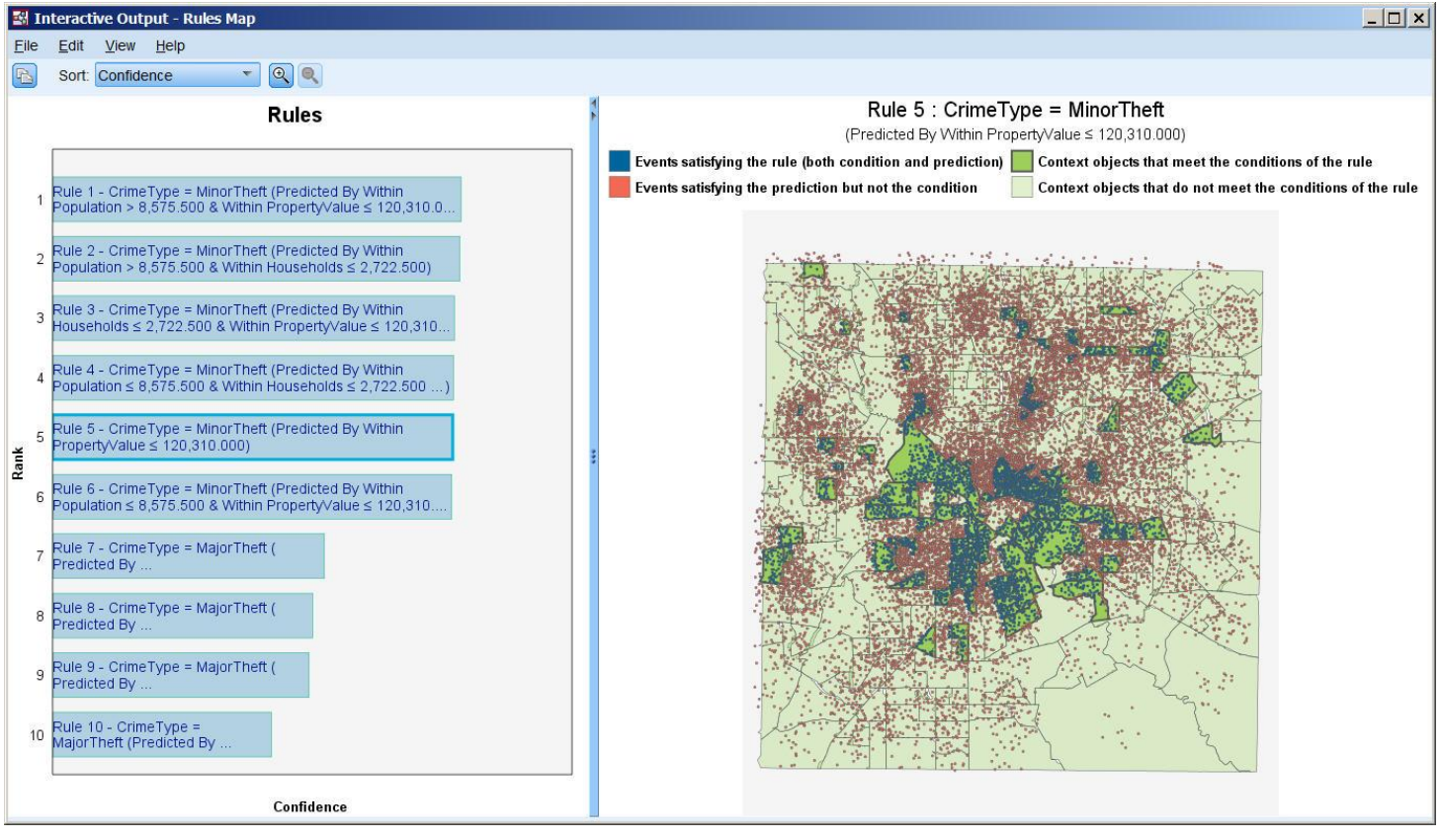
- **Geospatial Association Rules**
- Using geospatial association rules, you can find patterns in data based on both the spatial and non-spatial properties. For example, you might identify patterns in crime data by location and demographic attributes. From these patterns, you can build rules that predict where certain types of crimes are likely to occur.
- This procedure is available in the *Base Statistics* option.



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

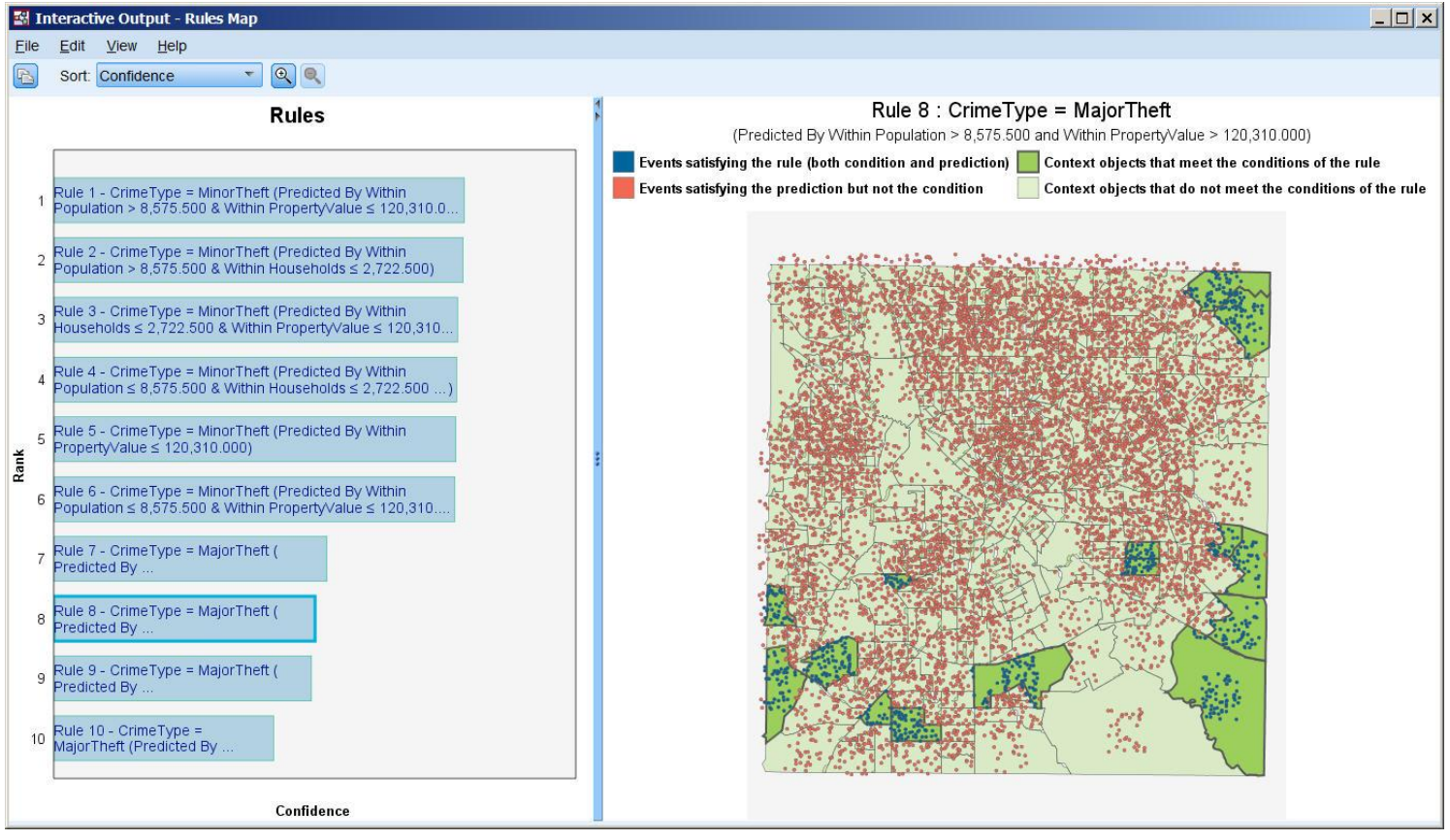
IBM SPSS Statistics v23: Geospatial Association Rules



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

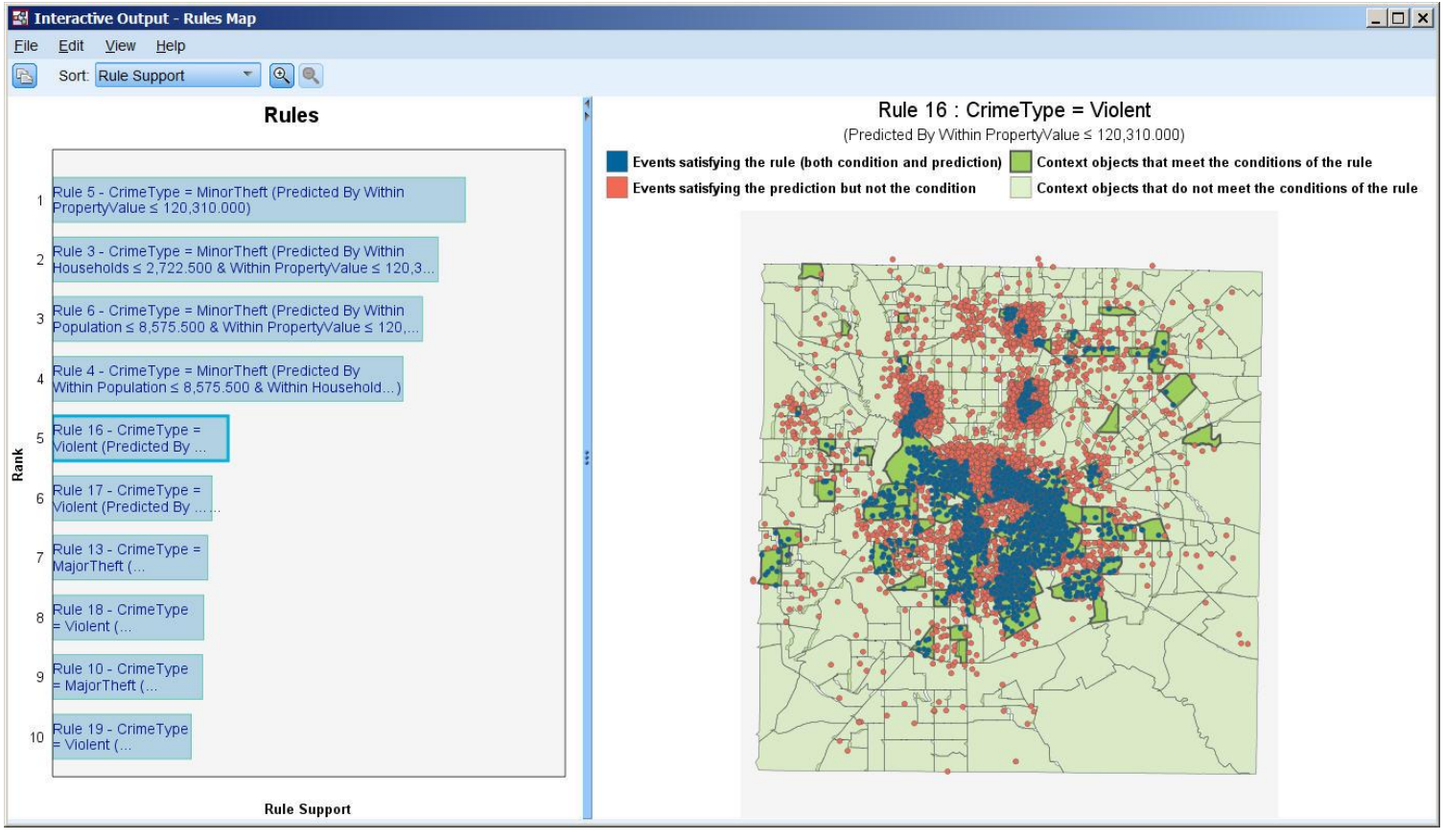
IBM SPSS Statistics v23: Geospatial Association Rules



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

IBM SPSS Statistics v23: Geospatial Association Rules

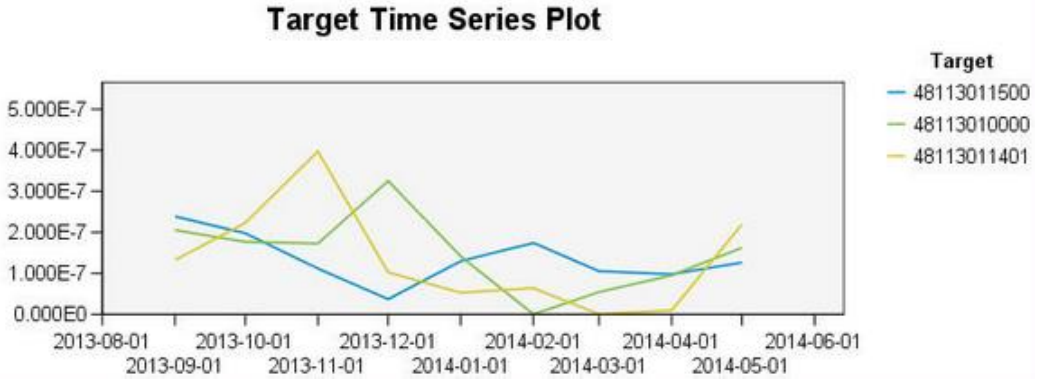


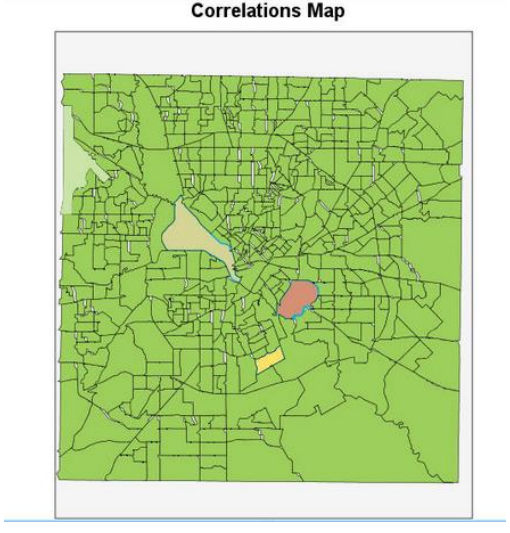
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

IBM SPSS Statistics v23

- **Spatial Temporal Prediction**
- Spatial temporal prediction uses data that contains location data, input fields for prediction (predictors), a time field, and a target field. Each location has numerous rows in the data that represents the values of each predictor at each time interval at each location.
- This procedure is available in the *Base Statistics* option.



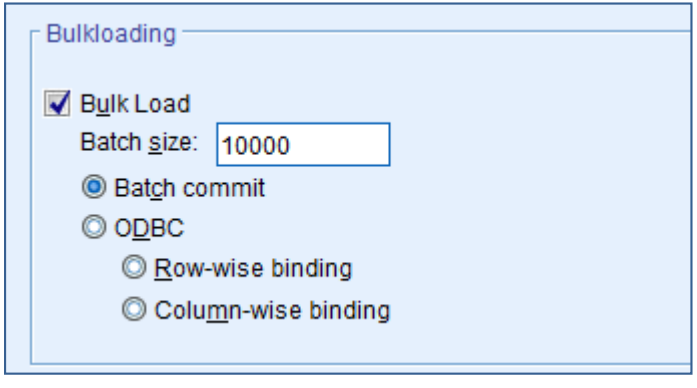


IBM SPSS Statistics v23

- **Temporal Causal Models**
- Temporal causal modelling attempts to discover key causal relationships in time series data. In temporal causal modelling, you specify a set of target series and a set of candidate inputs to those targets.
- The procedure then builds an autoregressive time series model for each target and *includes only those inputs that have a causal relationship with the target*. This approach differs from traditional time series modelling where you must explicitly specify the predictors for a target series.
- Temporal causal modelling procedures are available in the *Forecasting* option.

IBM SPSS Statistics v23

- **Bulk Loading to a database**
- When you export data to a database, bulk loading submits data to the database in batches instead of one record at a time. This action can make the operation much faster, particularly for large data files.



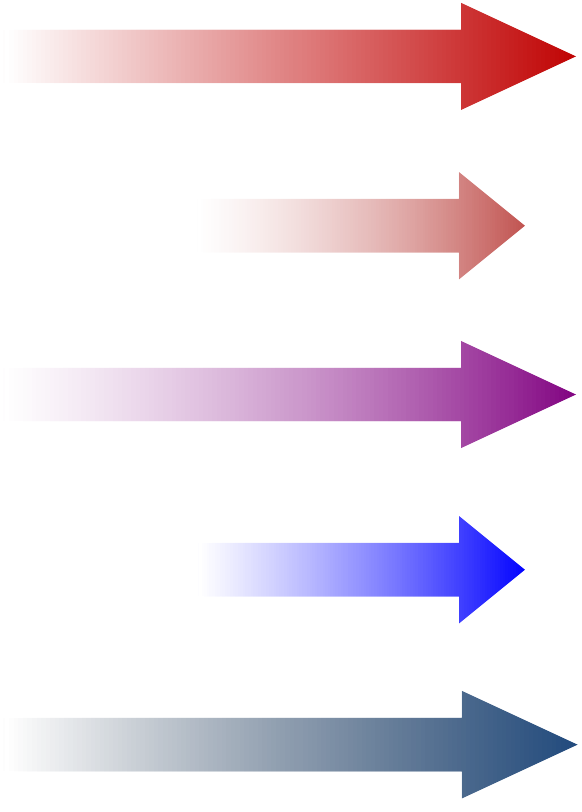
Getting more from SPSS
- Automating and Extending

John McConnell – Services

Contents

- Background
- Levels of automation with syntax
- Automating beyond syntax
- Extensions
- Automating SPSS from the outside
- Support from Smart Vision

Some reasons to automate

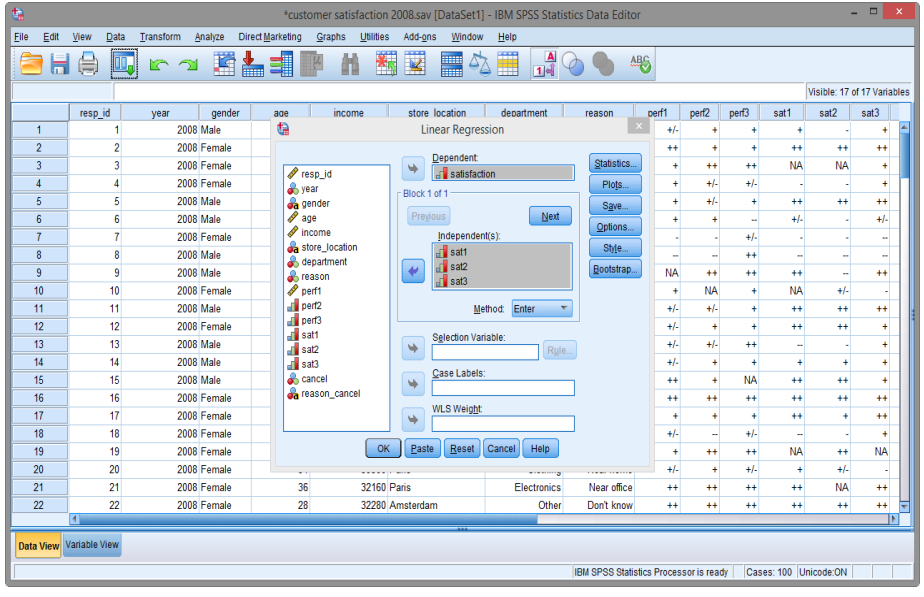


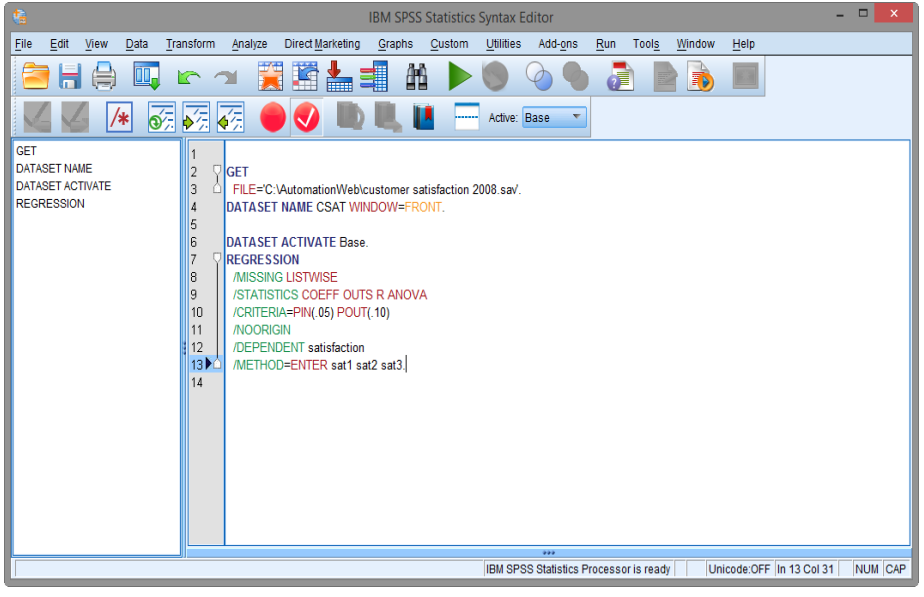
- Productivity
- Repeatability
- Governance
- Communication
- Delegation

Contents

- Background
- Levels of automation with syntax and streams
- Automating beyond syntax and streams
- Automating SPSS from the outside

Automation – Level 1





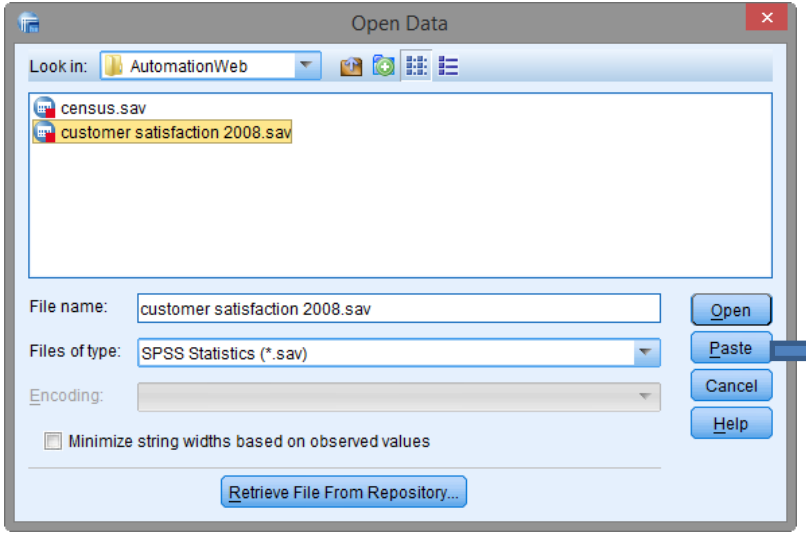
From the GUI to …

Syntax

SMARTVISION Europe

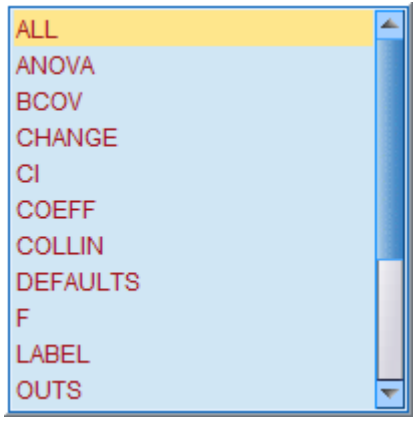
A SELECT INTERNATIONAL COMPANY

Defining and pasting



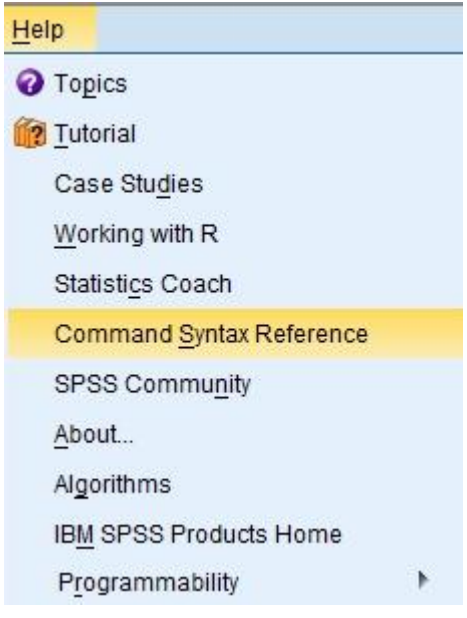

```
GET
  FILE='C:\AutomationWeb\customer satisfaction 2008.sav'.
DATASET NAME CSAT WINDOW=FRONT.

DATASET ACTIVATE CSAT.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT satisfaction
  /METHOD=ENTER sat1 sat2 sat3.
```

Getting help



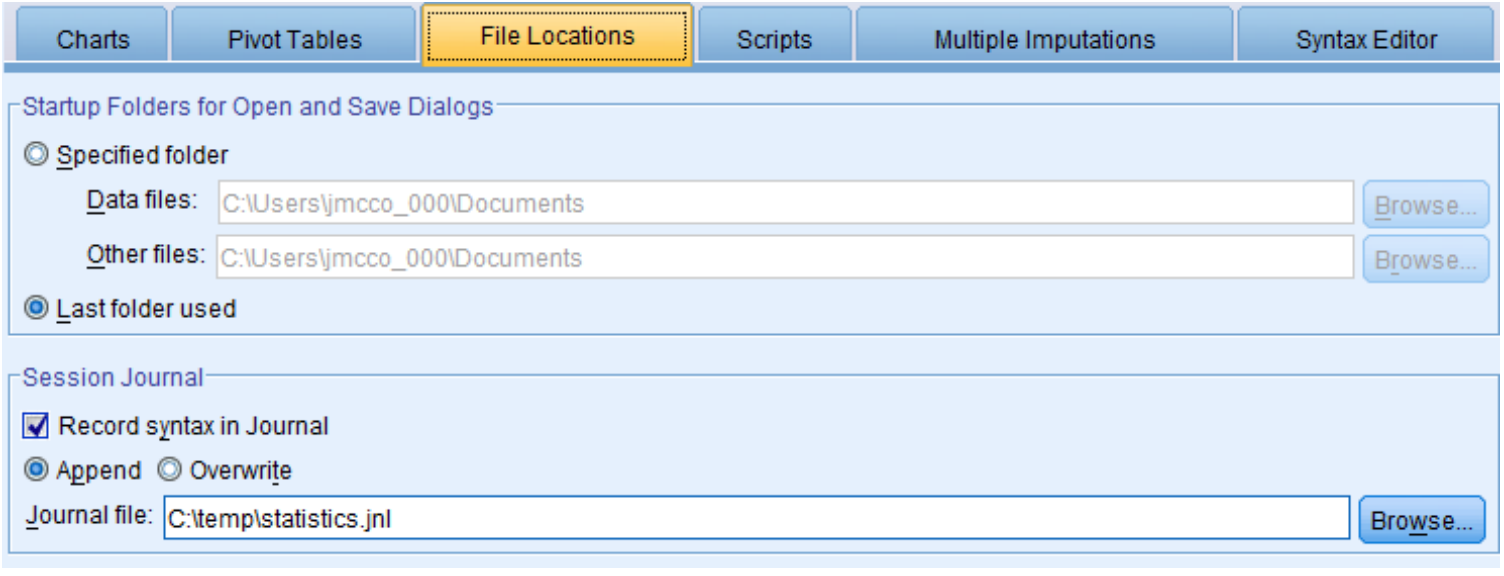
Auto or <ctrl>+<space>
Pops up relevant options

Tool to show us the syntax options for the selected command



The PDF of all commands and options

Forgot to Paste?

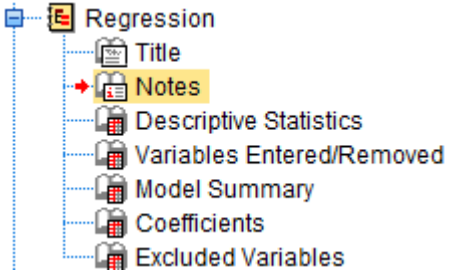


The **Journal File** is set (in **Edit > Options**) to record syntax automatically – until overwritten or deleted

Forgot to Paste?

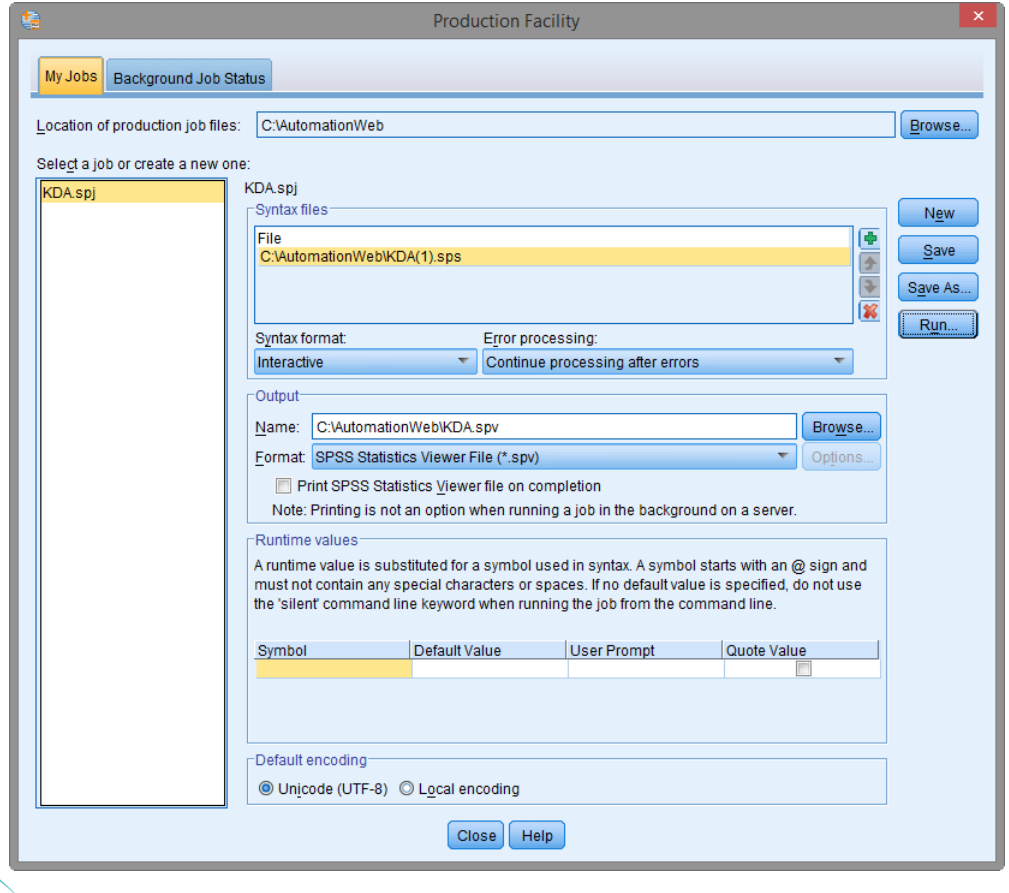
Notes

Output Created		03-DEC-2014 07:12:18
Comments		
Input	Data	C:\AutomationWeb\customer satisfaction 2008.sav
	Active Dataset	Base
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	140
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any variable used.
Syntax		REGRESSION /DESCRIPTIVES MEAN /MISSING LISTWISE /STATISTICS R COEFF OUTS /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT satisfaction /METHOD= STEPWISE sat1 sat2 sat3.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.02
	Memory Required	5088 bytes
	Additional Memory Required for Residual Plots	0 bytes



The (usually hidden) **Notes** table in output contains the syntax for each output

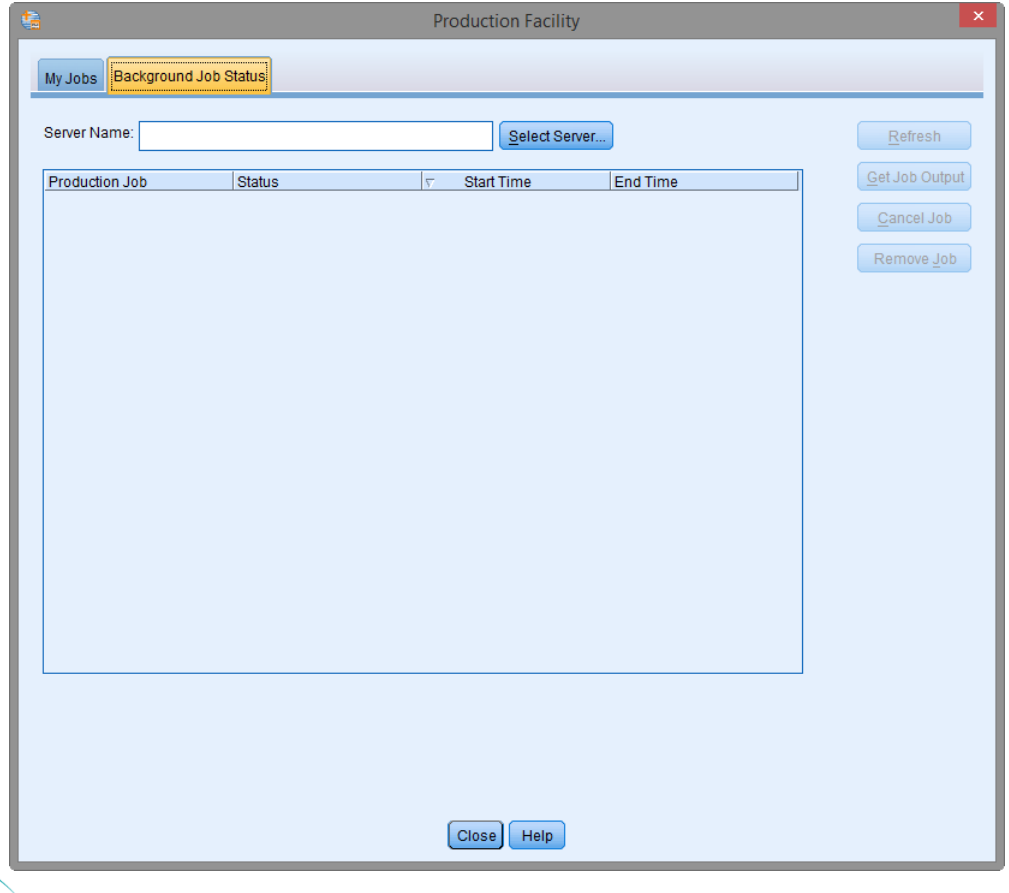
Batch running Syntax – The Production Facility



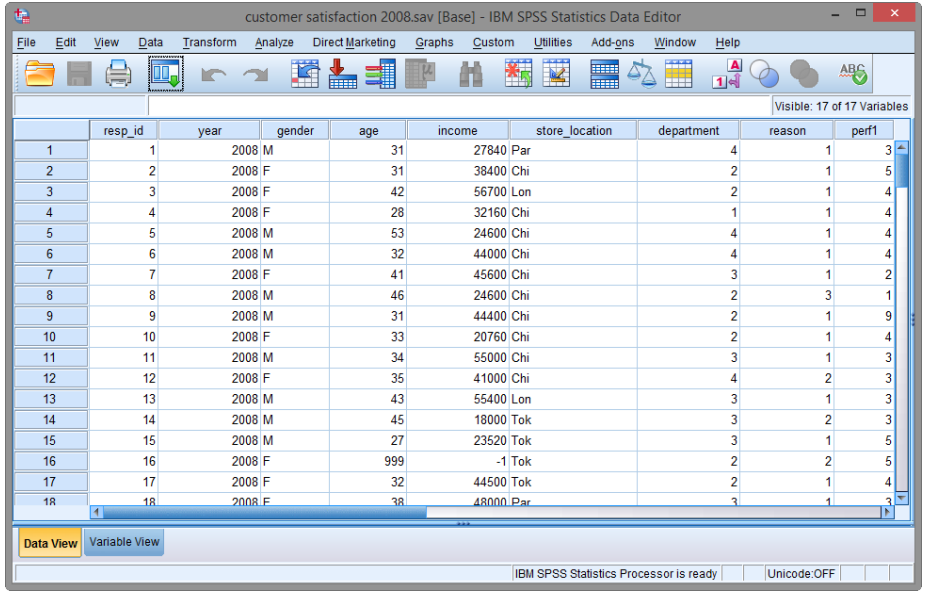
Menu path:

Utilities > Production Facility

Background mode runs production jobs on a server



The server side batch engine

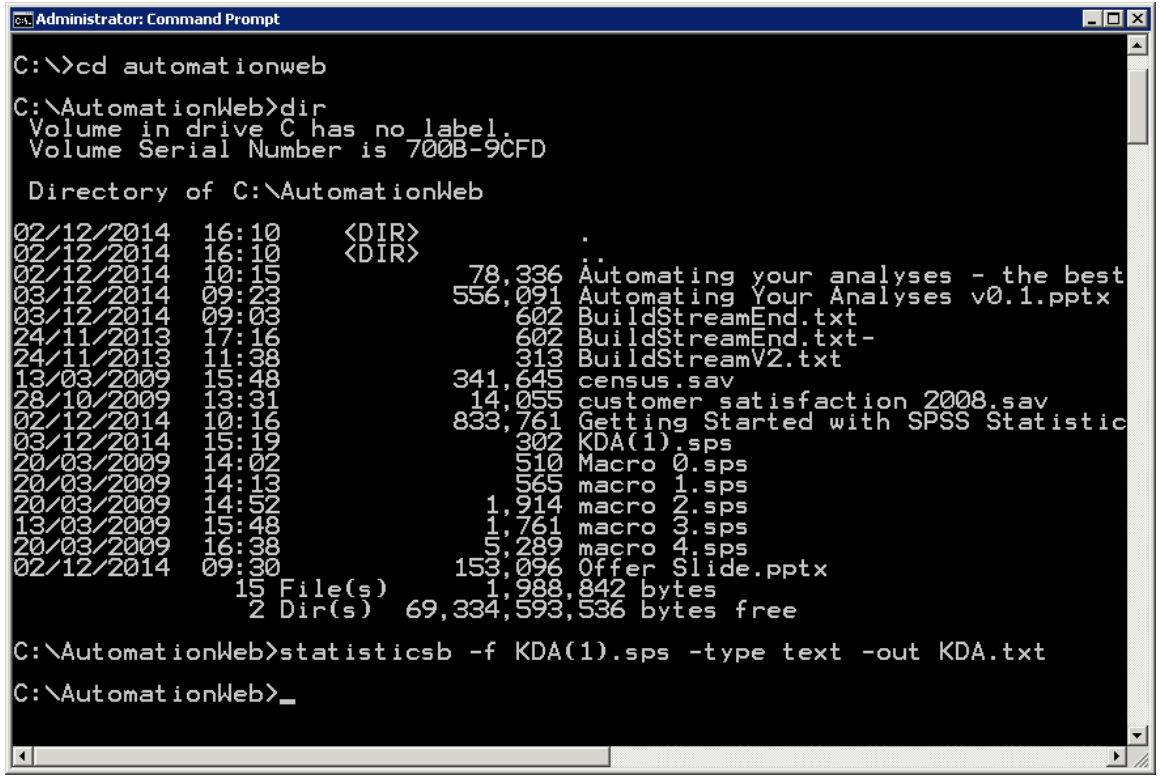


statisticsb

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

On the server side

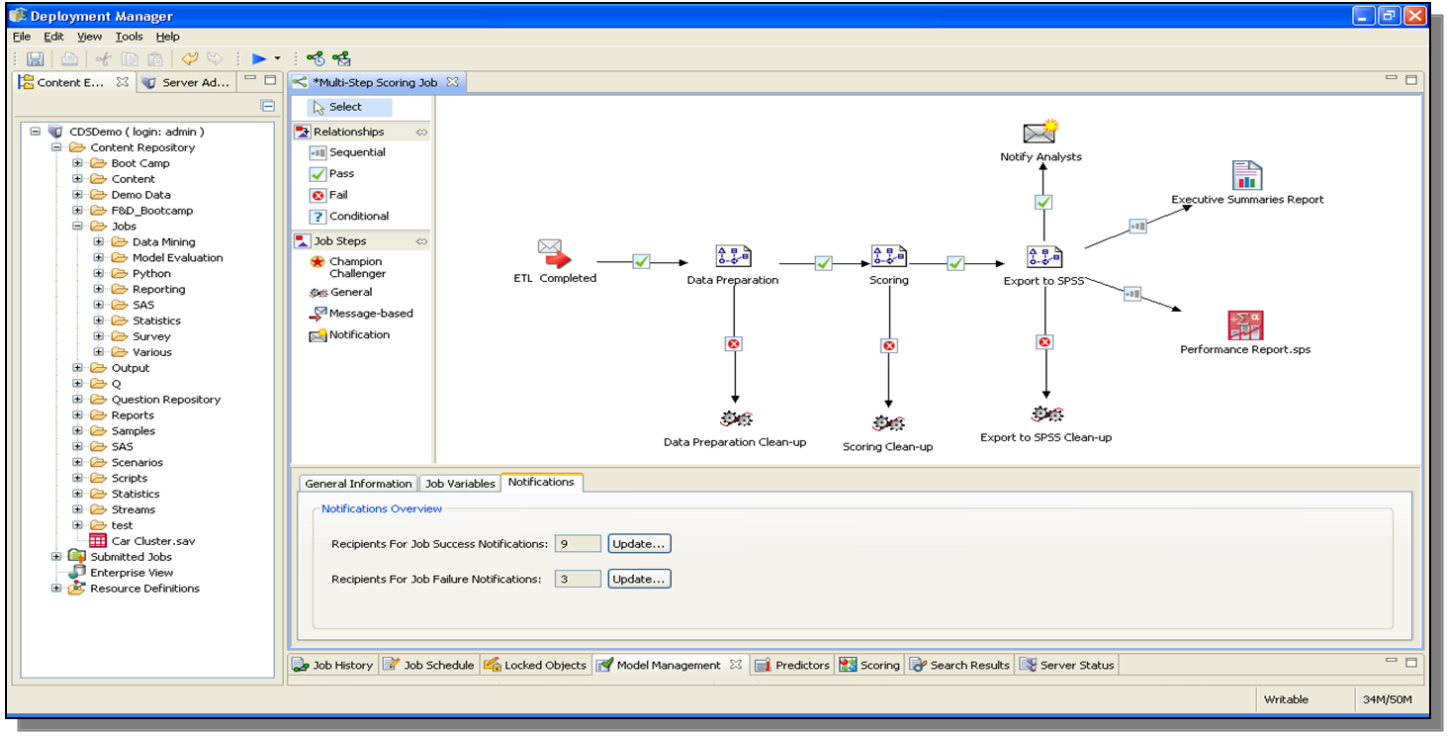


Batch jobs can be scheduled to run using the Windows Task Scheduler

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

IBM/SPSS C&DS is the next level of automation



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Contents

- Background
- Levels of automation with syntax and streams
- Automating beyond syntax and streams
- Automating SPSS from the outside

Automating beyond standard syntax - Statistics

- Macros
- Visual Basic
- Python
- Java
- R

More programming power
This includes:

- Creating re-usable blocks of code
- Creating our own User Interfaces
- Automating processes beyond SPS
  - e.g. controlling Excel, PowerPoint etc.

Automating beyond standard syntax - Macros

Pros:

- An extension of the SPSS syntax language
- Run inside the same files(s)

Cons:

- They have their own syntactic rules
- Functionally limited
  - Don’t support some more advanced programming constructs
  - Can’t control other tools

Example Macros

A simple to define a re-usable variable

```
DEFINE !MYFOLDER ()
"C:\TRAIN\SYNTAX_II\"
!ENDDEFINE.
```

Using the macro variable in syntax

```
GET FILE = !MYFOLDER + 'census.sav'.
DATASET NAME census WINDOW=FRONT.
```

A macro to create a new “command”

```
DEFINE !CLOSEALL (DATASETS     = !CHAREND ("/")
                  /VIEWERDOCS  = !CMDEND       )


!IF (!DATASETS = YES) !THEN
NEW FILE.
DATASET CLOSE ALL.
!IFEND

!IF (!VIEWERDOCS = YES) !THEN
OUTPUT CLOSE ALL.
!IFEND

!ENDDEFINE.
```

Calling that macro

```
!CLOSEALL DATASETS    = YES
        /VIEWERDOCS  = YES.
```

Automating beyond standard syntax – VB, Python, Java, R

Pros:

- More powerful / widely used languages
- Allow us to add **extended** functionality
- Go beyond automating SPSS

Cons:

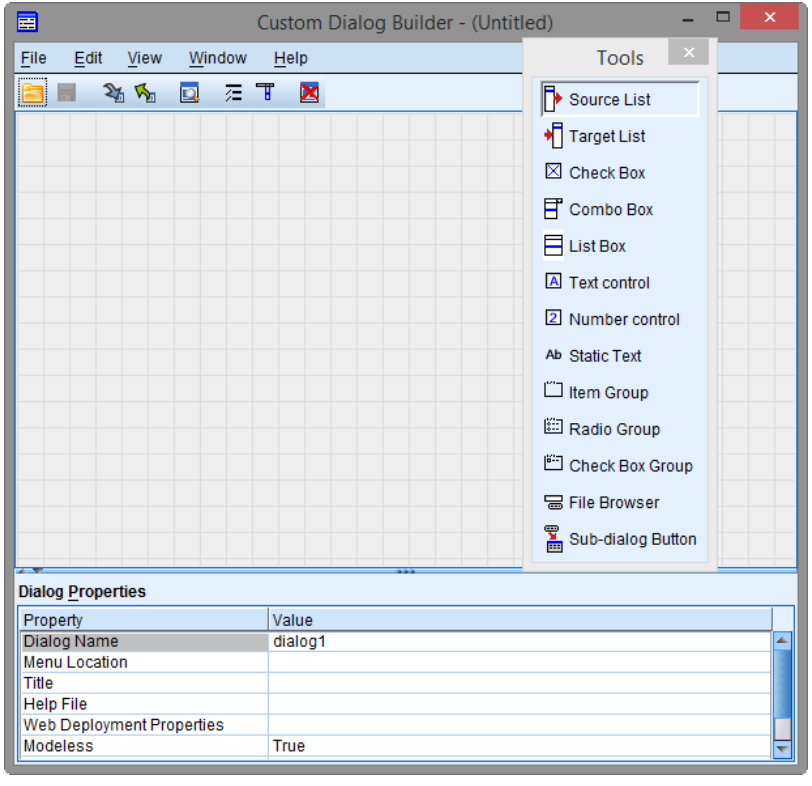
- They run separately so we need to integrate syntax into them
- Need to learn / have access to programming expertise

An example VB script

```
'Begin Description
'This file removes upper diagonal of correlation matrix and highlights
'correlations significant at the .01 level.
'End Description
Sub Main
    Dim objPivotTable As PivotTable
    Dim objItem As ISpssItem
    Dim bolFoundOutputDoc As Boolean
    Dim bolPivotSelected As Boolean
    Dim lngIndex As Long
    Dim objOutputDoc As ISpssOutputDoc
    Call GetFirstSelectedPivot(objPivotTable, objItem, bolFoundOutputDoc, bolPivotSelected)

    Call Correlations_Table_Correlations_Create(objPivotTable, objOutputDoc, lngIndex)
    'Deactivate the correlation pivot table
    objItem.Deactivate

End Sub
```

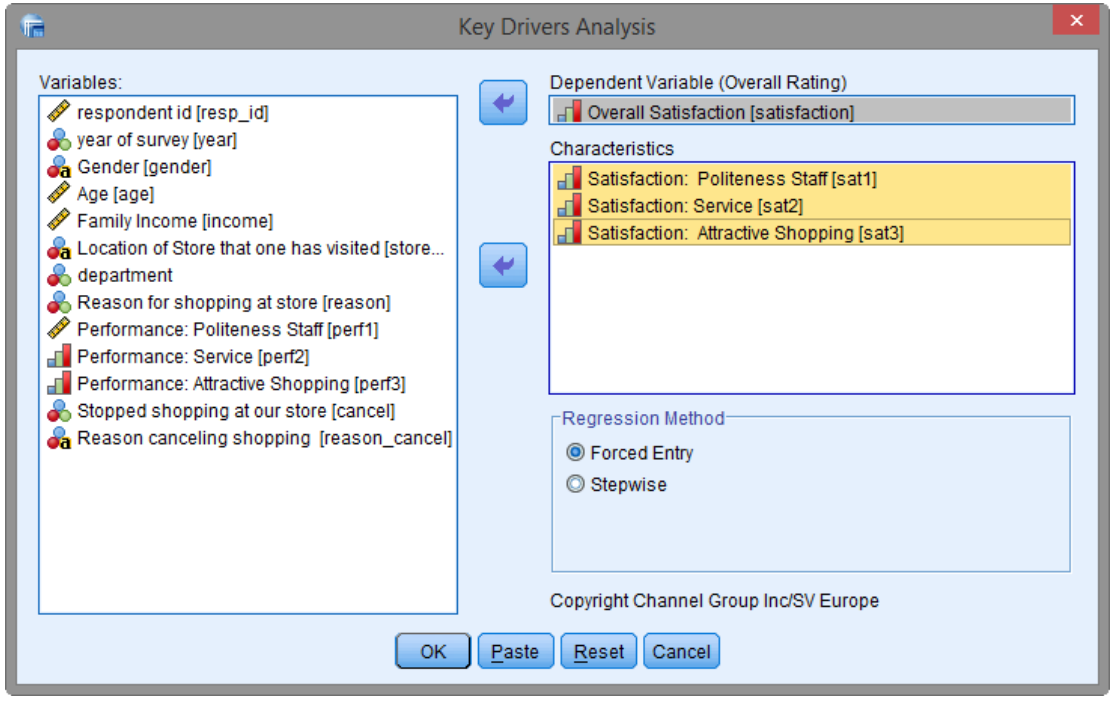
This script looks inside a correlation table
Identifies statistically significant correlations

Extensibility



We can use the Custom Dialog builder in SPSS to create our own UIs and automate behind them With Syntax, Python, R, etc.

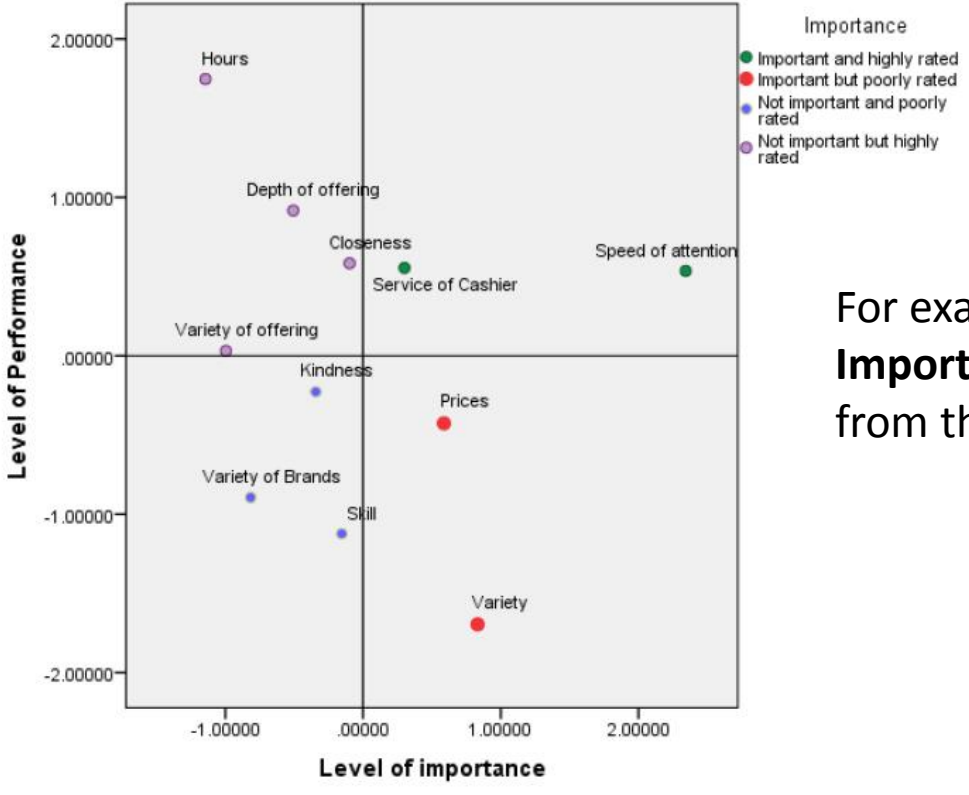
A KDA extension



This example (available for download from our web site shortly) was developed by Channel Group in the US

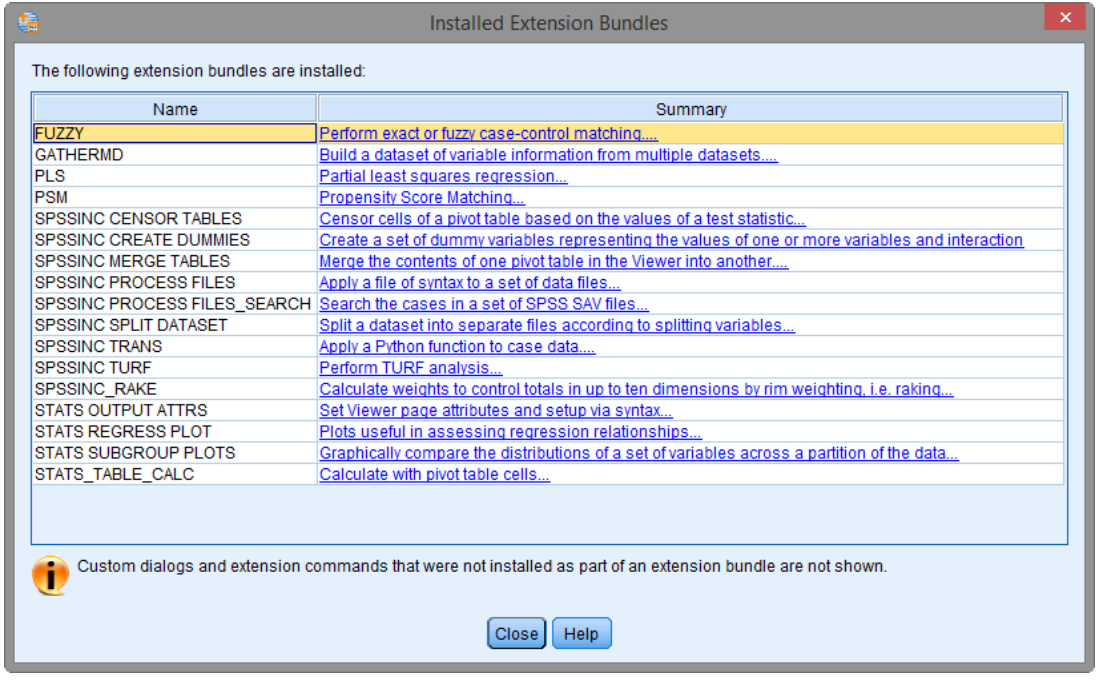
It simplifies several steps beyond the KDA syntax that we ran earlier

A KDA extension



For example it automatically produces the **Importance v Performance** quadrant chart from the SPSS regression output

Extension Bundles



Typically written in Python (or R)

Check out the SPSS Developer Central for more resources

[www.ibm.com/**spss**/dev**central**](www.ibm.com/spss/devcentral)

R

[Home]

Download

CRAN

R Project

About R
Contributors
What's New?
Mailing Lists
Bug Tracking
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **R version 3.2.0** (Full of Ingredients) has been released on 2015-04-16.
- **R version 3.1.3** (Smooth Sidewalk) has been released on 2015-03-09.
- **The R Journal Volume 6/2** is available.
- **useR! 2015**, will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.
- **useR! 2014**, took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.

Contributed Packages

Available Packages

Currently, the CRAN package repository features 6646 available packages.

Table of available packages, sorted by date of publication

Table of available packages, sorted by name

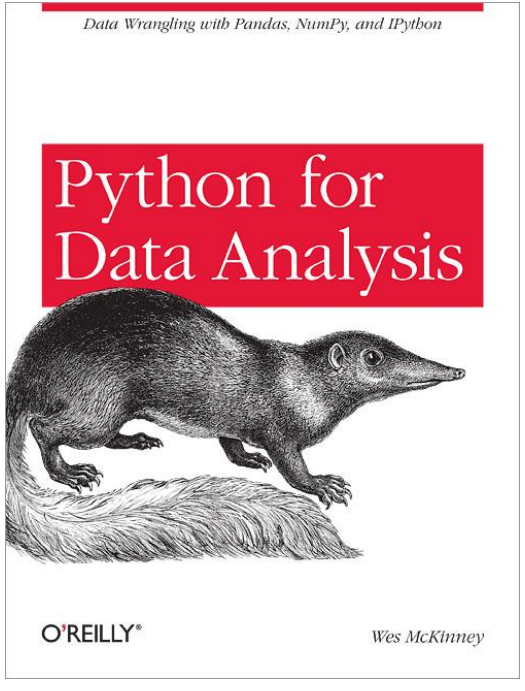
http://www.r-project.org/

An Alternative KDA in R

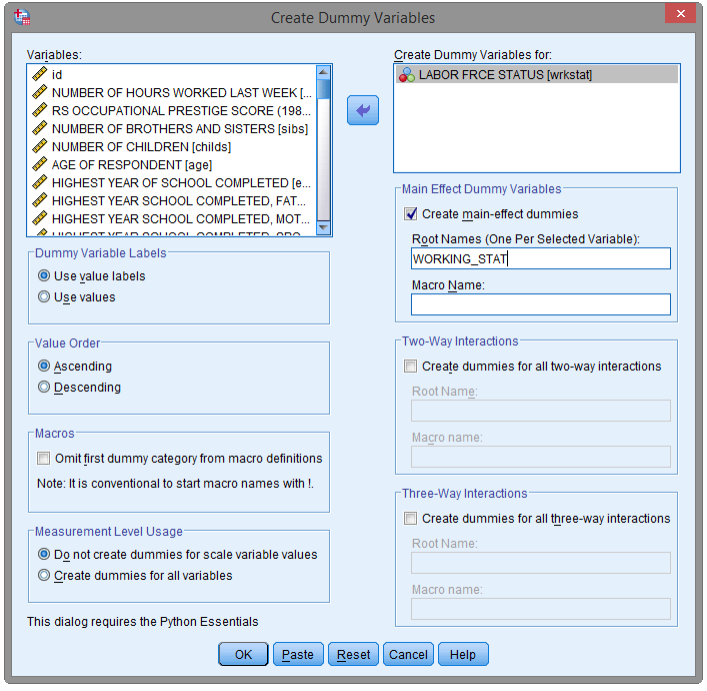
```
* Install additional packages

BEGIN PROGRAM R.

install.packages("kappalab")
install.packages("relaimpo")
require(relaimpo)

testdata =  spssdata.GetDataFromSPSS()

END PROGRAM.

* now run the shapley regression.

BEGIN PROGRAM R.

regdata =  spssdata.GetDataFromSPSS()

reg1 <- lm( overallrating ~ characteristic_1 +characteristic_2 +characteristic_3 +characteristic_4 +characteristic_5
   +characteristic_6 +characteristic_7 +characteristic_8 +characteristic_9 +characteristic_10
   +characteristic_11 , data=regdata)

summary(reg1)

shap <- calc.relimp(reg1, rela=TRUE)
shap

END PROGRAM.
```

Python

- In SPSS Python is – in the first instance - a more powerful scripting language
- It can also be used for Data Analysis



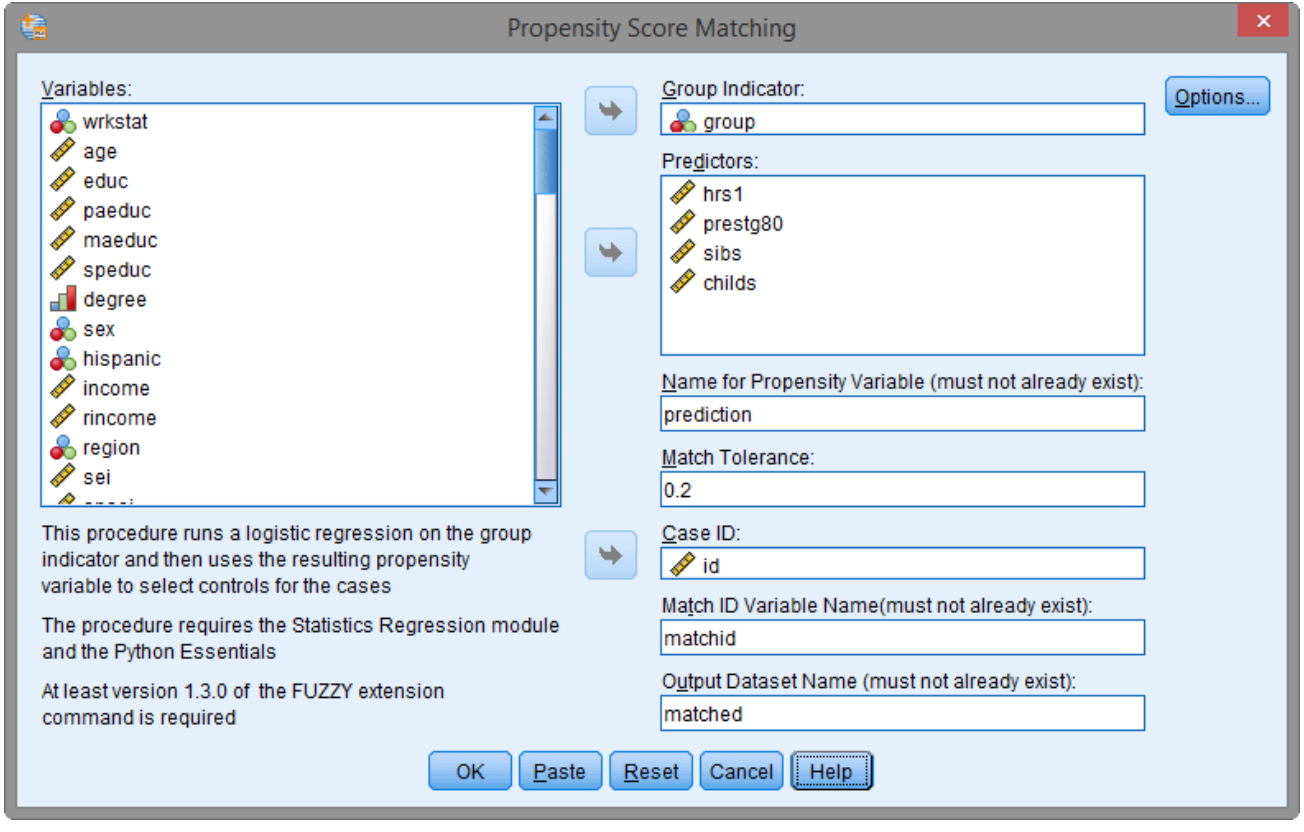
The Create Dummy Variables extension



Variable Creation

	Label
WORKING_STAT_1	wrkstat=WORKING FULLTIME
WORKING_STAT_2	wrkstat=WORKING PARTTIME
WORKING_STAT_3	wrkstat=TEMP NOT WORKING
WORKING_STAT_4	wrkstat=UNEMPL, LAID OFF
WORKING_STAT_5	wrkstat=RETIRED
WORKING_STAT_6	wrkstat=SCHOOL
WORKING_STAT_7	wrkstat=KEEPING HOUSE
WORKING_STAT_8	wrkstat=OTHER

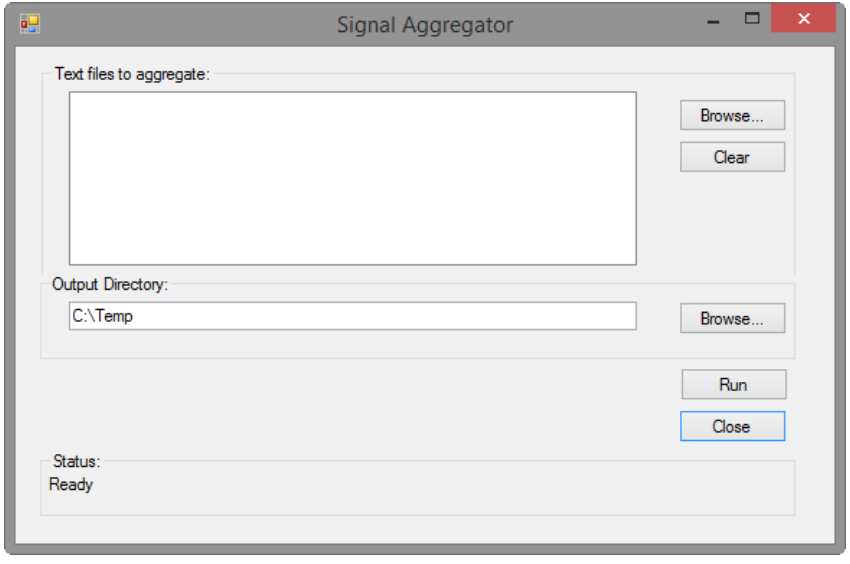
The Propensity Matching extension

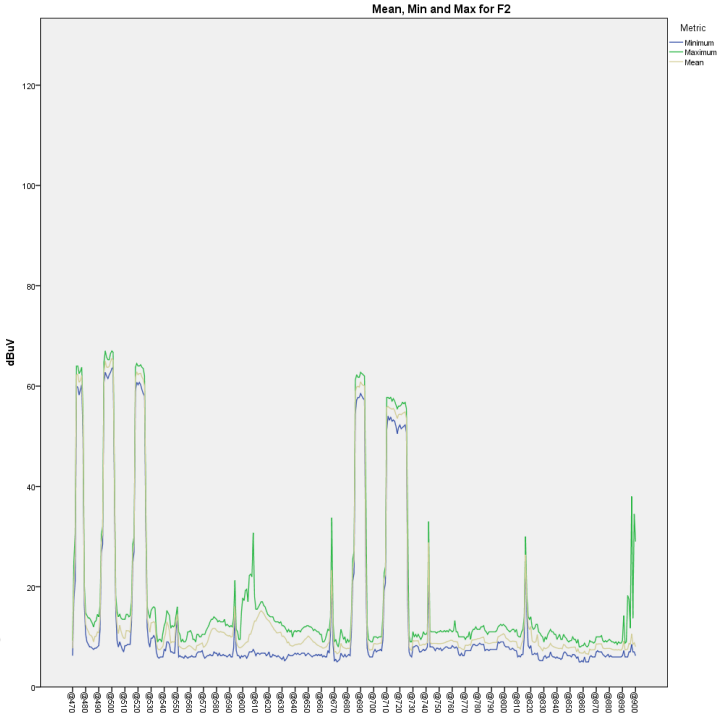


Contents

- Background
- Levels of automation with syntax and streams
- Automating beyond syntax and streams
- Automating SPSS from the outside

Automating from the outside



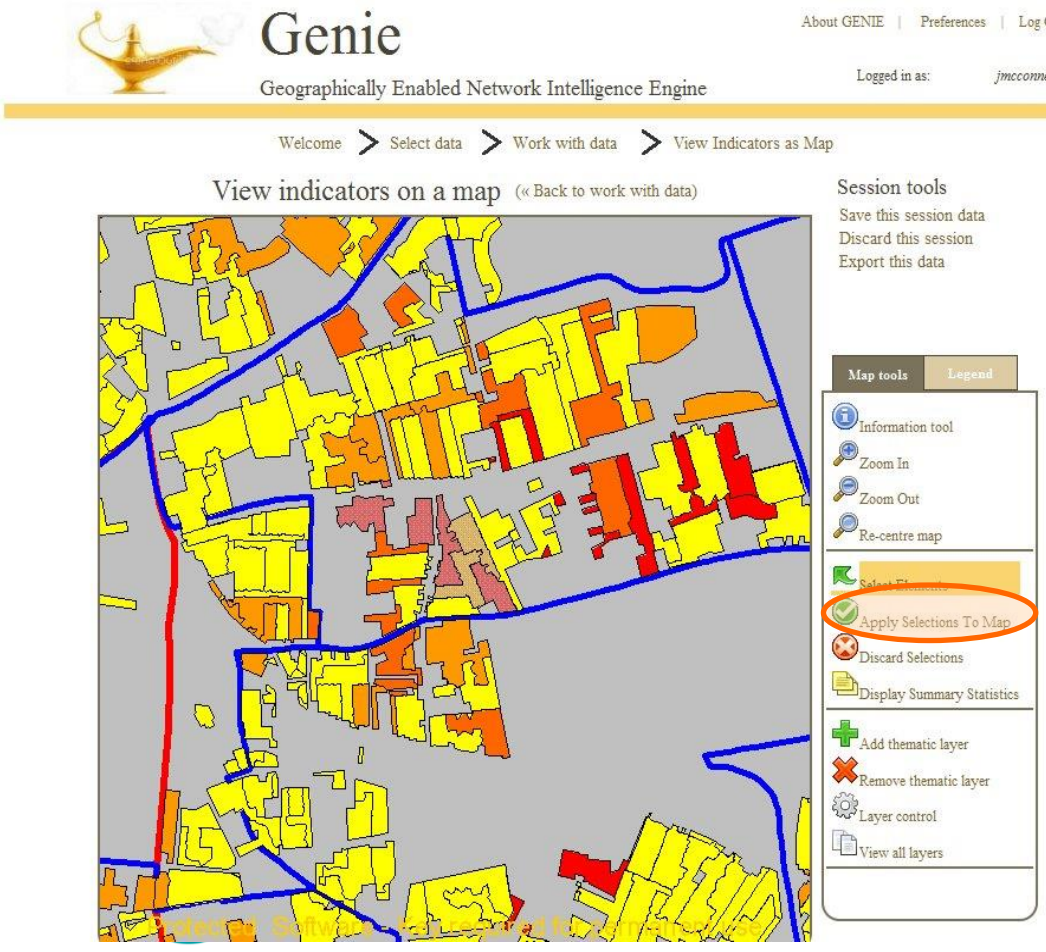


This UI runs a standalone app

a) Reads and cleans data coming from sensors
b) Produces summary graphs as jpegs for integration into reports

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

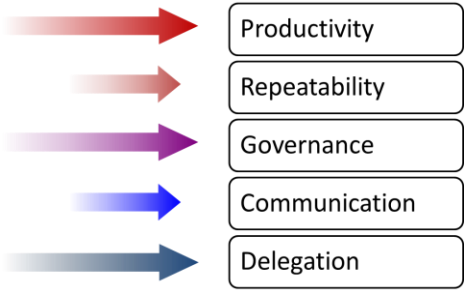


This on-line GIS app is designed for local planners

- It **runs factor analysis** models based on selected criteria to **create indices of sustainability**

In Summary

- The interface to R allows us to mix and match R analysis with SPSS Analysis inside the SPSS UI
- It is possible to automate just about anything in and around SPSS
- This can lead to significant time saving, increased productivity, higher quality and better governance
- As usual the key question is whether the build (development) time is worth investing
  - Does it save time, money etc. in the long run?



Training and support options with Smart Vision Europe Ltd

- As experts in SPSS / Staistics / Analytics / Predictive Analytics we
  - Deliver classroom training courses
    - Public and private
    - Optionally create custom classroom courses on your data
  - Offer side by side training support
  - Offer “skills transfer” consulting
  - Run booster and refresher sessions to get more from your SPSS licences
  - Give no strings attached advice
- We are a support providing partner so if you already have SPSS you can source your technical support directly from us (identical costs to IBM)
  - We offer telephone support with real people as well as web tickets / email queries
  - We offer “how to” support to help you get moving on your project quickly

SMARTVISION Europe

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on Linked In
Sign up for our Newsletter

Thank you

How can we help?

- Sell and support the full SPSS suite of tools
- Run an accredited UK SPSS Support Desk
- Deliver the SPSS public training schedule on behalf of IBM
- Deliver a complete range of professional services
  - Training
  - Custom training
  - Guided consulting
  - Project management & analytical consulting
  - Technical integration (data integration, application development, BI etc.)

SMARTVISION
Europe

Summary, Next Steps & Close

Working with Smart Vision Europe Ltd

- As a premier partner we sell the IBM SPSS suite of software to you directly
  - We're agile, responsive and generally easier to deal with
- As experts in SPSS / Analytics / Predictive Analytics we will
  - deliver classroom training courses
  - offer side by side training support
  - offer "skills transfer" consulting
  - run booster and refresher sessions to get more from your SPSS licences
  - Give no strings attached advice
- We are a support providing partner so if you already have SPSS you can source your technical support directly from us (identical costs to IBM)
  - We offer telephone support with real people as well as web tickets / email queries
  - We offer "how to" support to help you get moving on your project quickly

SMARTVISION Europe

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on Linked In
Sign up for our Newsletter

Thank you