



Automating with IBM SPSS

John McConnell – Services

Rachel Clinton – Business Development

www.sv-europe.com

A SELECT INTERNATIONAL COMPANY

Contents

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

Today's Objectives

- To introduce types and levels of automation in SPSS
- To sow some seeds of what is possible
- To provide some pointers of how to start automating more
 - Skills needed
 - Resources available

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

IBM/SPSS Statistics & IBM/SPSS Modeler

customer satisfaction 2008.sav [Base] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Custom Utilities Add-ons Window Help

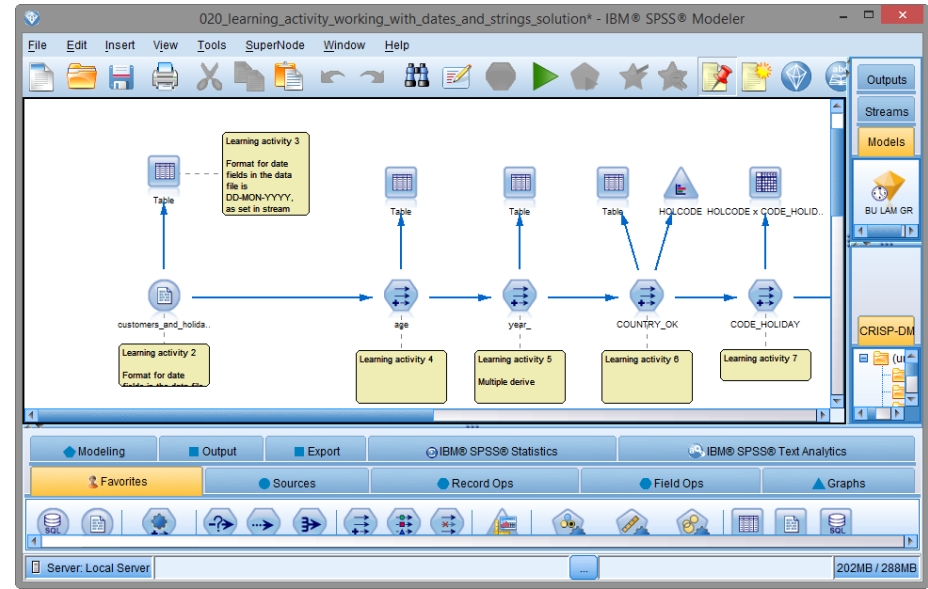
Visible: 17 of 17 Variables

	resp_id	year	gender	age	income	store_location	department	reason	perf1
1	1	2008	M	31	27840	Par	4	1	3
2	2	2008	F	31	38400	Chi	2	1	5
3	3	2008	F	42	56700	Lon	2	1	4
4	4	2008	F	28	32160	Chi	1	1	4
5	5	2008	M	53	24600	Chi	4	1	4
6	6	2008	M	32	44000	Chi	4	1	4
7	7	2008	F	41	45600	Chi	3	1	2
8	8	2008	M	46	24600	Chi	2	3	1
9	9	2008	M	31	44400	Chi	2	1	9
10	10	2008	F	33	20760	Chi	2	1	4
11	11	2008	M	34	55000	Chi	3	1	3
12	12	2008	F	35	41000	Chi	4	2	3
13	13	2008	M	43	55400	Lon	3	1	3
14	14	2008	M	45	18000	Tok	3	2	3
15	15	2008	M	27	23520	Tok	3	1	5
16	16	2008	F	999	-1	Tok	2	2	5
17	17	2008	F	32	44500	Tok	2	1	4
18	18	2008	F	38	48000	Par	3	1	3

Data View Variable View

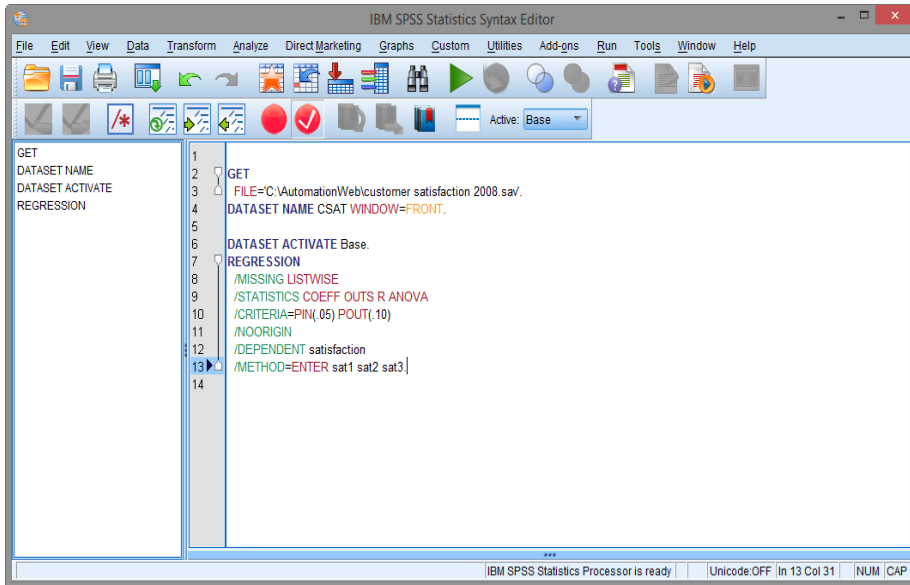
IBM SPSS Statistics Processor is ready Unicode:OFF

Statistics

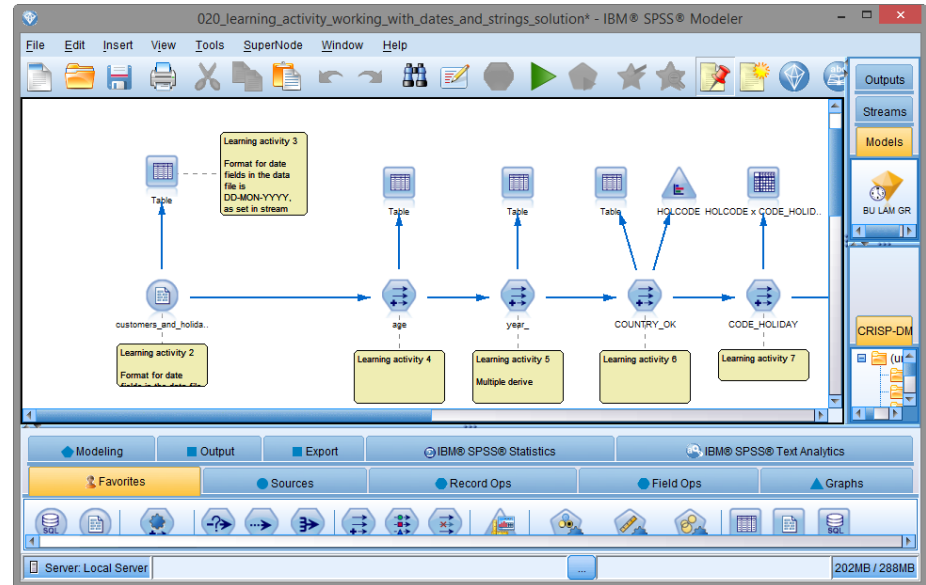


Modeler

Automation – Level 1

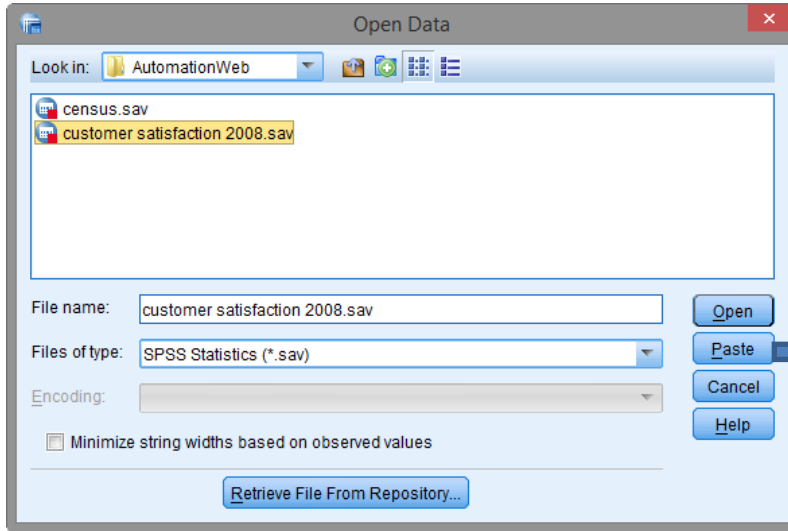


Syntax



Streams

Defining and pasting



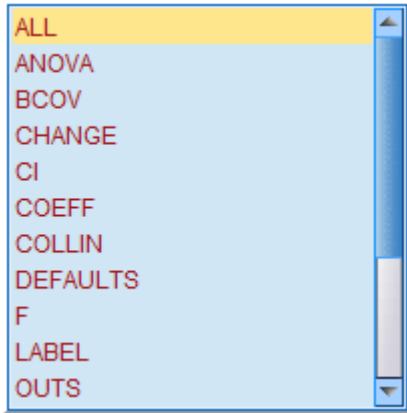
GET

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

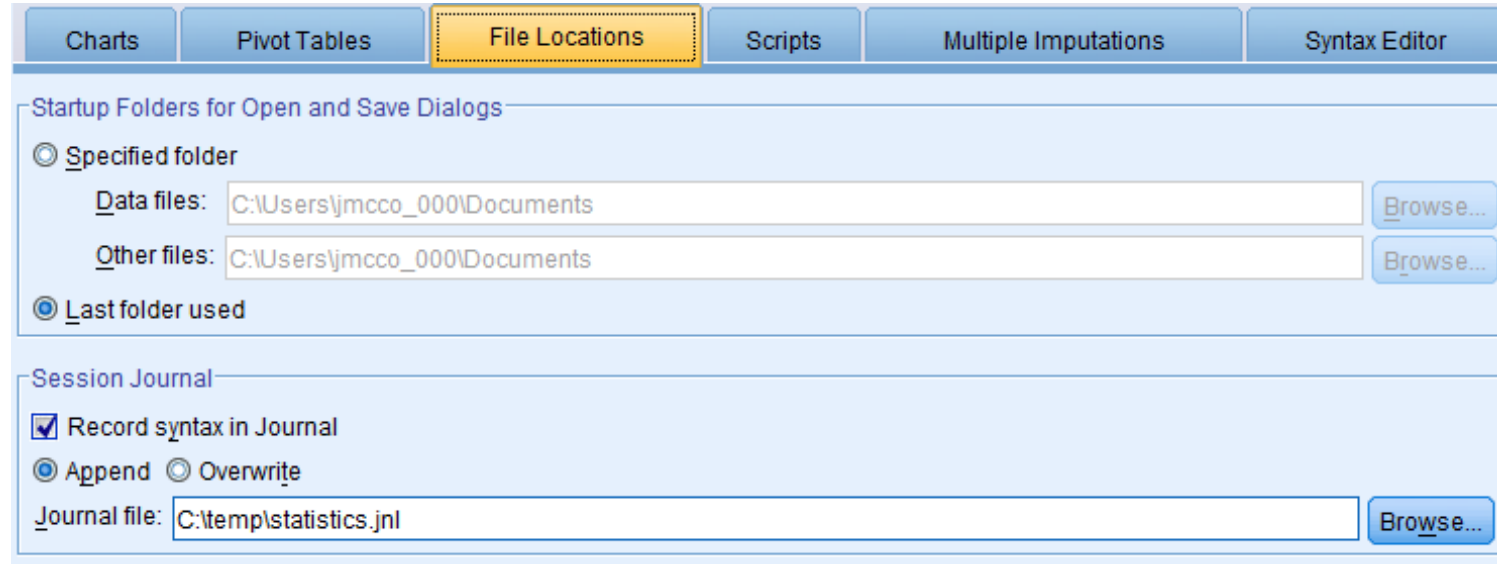


Tool to show us the
syntax options for
the selected
command



The PDF of all
commands and
options

Forgot to Paste?



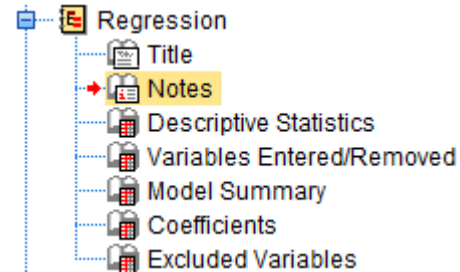
The screenshot shows a software interface with a top navigation bar containing six tabs: 'Charts', 'Pivot Tables', 'File Locations' (which is highlighted with a yellow border), 'Scripts', 'Multiple Imputations', and 'Syntax Editor'. Below the tabs, the 'File Locations' section is divided into two main areas. The first area, titled 'Startup Folders for Open and Save Dialogs', contains two radio button options. The first option, 'Specified folder', is selected and has two text input fields below it: 'Data files:' and 'Other files:', both containing the path 'C:\Users\jmcoco_000\Documents'. Each field has a 'Browse...' button to its right. The second option, 'Last folder used', is also selected. The second area, titled 'Session Journal', contains a checked checkbox 'Record syntax in Journal', followed by two radio button options 'Append' (selected) and 'Overwrite'. Below these is a 'Journal file:' label and a text input field containing 'C:\temp\statistics.jnl', with a 'Browse...' button to its right.

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

The screenshot shows the 'Production Facility' dialog box in SPSS. It has two tabs: 'My Jobs' and 'Background Job Status'. The 'Location of production job files' is set to 'C:\AutomationWeb'. A list on the left shows 'KDA.spj' selected. The 'Syntax files' section contains a list with 'C:\AutomationWeb\KDA(1).sps'. The 'Syntax format' is set to 'Interactive' and 'Error processing' is set to 'Continue processing after errors'. The 'Output' section shows the name 'C:\AutomationWeb\KDA.spv' and format 'SPSS Statistics Viewer File (*.spv)'. There is a checkbox for 'Print SPSS Statistics Viewer file on completion' which is unchecked. The 'Runtime values' section contains a table with columns 'Symbol', 'Default Value', 'User Prompt', and 'Quote Value'. The 'Default encoding' section has radio buttons for 'Unicode (UTF-8)' (selected) and 'Local encoding'. At the bottom are 'Close' and 'Help' buttons.

Production Facility

My Jobs Background Job Status

Location of production job files: C:\AutomationWeb Browse...

Select a job or create a new one:

KDA.spj

Syntax files

File

C:\AutomationWeb\KDA(1).sps

Syntax format: Interactive Error processing: Continue processing after errors

Output

Name: C:\AutomationWeb\KDA.spv Browse...

Format: SPSS Statistics Viewer File (*.spv) Options...

☐ Print SPSS Statistics Viewer file on completion

Note: Printing is not an option when running a job in the background on a server.

Runtime values

A runtime value is substituted for a symbol used in syntax. A symbol starts with an @ sign and must not contain any special characters or spaces. If no default value is specified, do not use the 'silent' command line keyword when running the job from the command line.

Symbol	Default Value	User Prompt	Quote Value

Default encoding

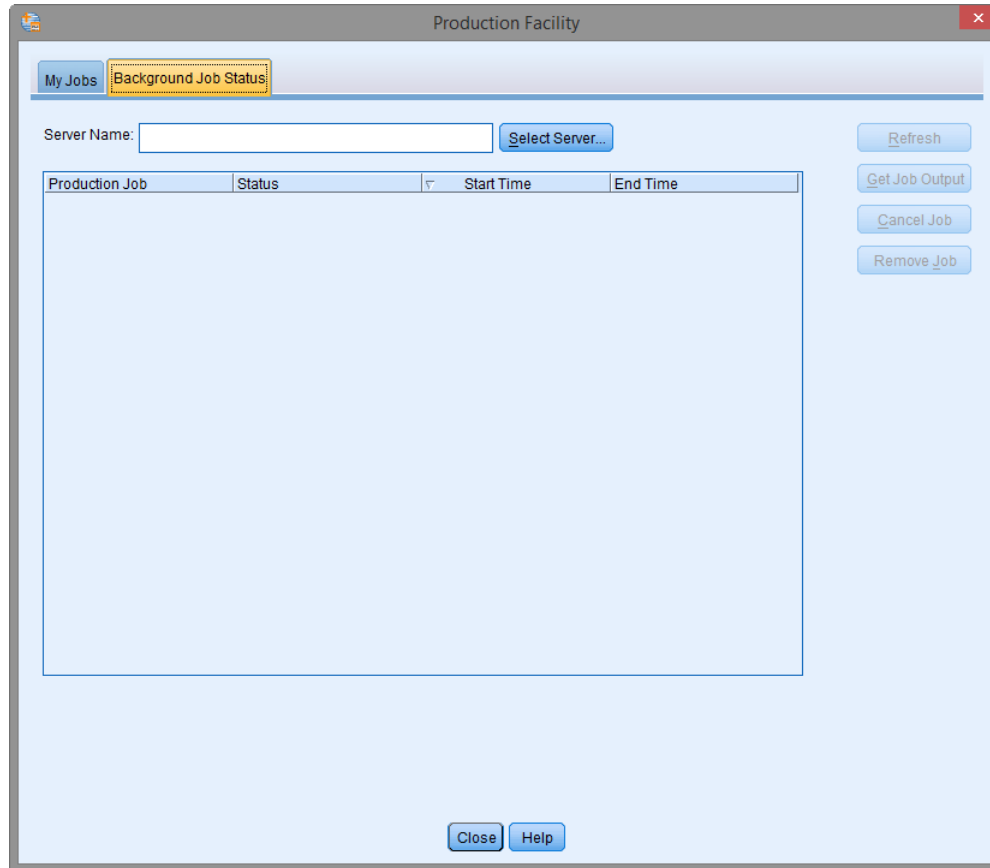
☒ Unicode (UTF-8) ☐ Local encoding

Close Help

Menu path:

Utilities > Production Facility

Background mode runs production jobs on a server



Server side batch engines

customer satisfaction 2008.sav [Base] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Custom Utilities Add-ons Window Help

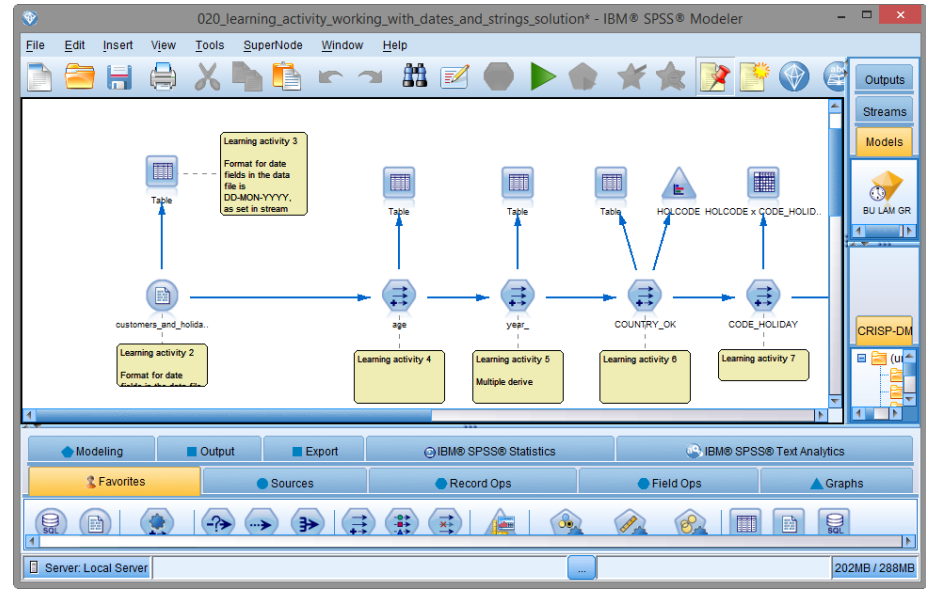
Visible: 17 of 17 Variables

	resp_id	year	gender	age	income	store_location	department	reason	perf1
1	1	2008	M	31	27840	Par	4	1	3
2	2	2008	F	31	38400	Chi	2	1	5
3	3	2008	F	42	56700	Lon	2	1	4
4	4	2008	F	28	32160	Chi	1	1	4
5	5	2008	M	53	24600	Chi	4	1	4
6	6	2008	M	32	44000	Chi	4	1	4
7	7	2008	F	41	45600	Chi	3	1	2
8	8	2008	M	46	24600	Chi	2	3	1
9	9	2008	M	31	44400	Chi	2	1	9
10	10	2008	F	33	20760	Chi	2	1	4
11	11	2008	M	34	55000	Chi	3	1	3
12	12	2008	F	35	41000	Chi	4	2	3
13	13	2008	M	43	55400	Lon	3	1	3
14	14	2008	M	45	18000	Tok	3	2	3
15	15	2008	M	27	23520	Tok	3	1	5
16	16	2008	F	999	-1	Tok	2	2	5
17	17	2008	F	32	44500	Tok	2	1	4
18	18	2008	F	38	48000	Par	3	1	3

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:OFF

statisticsb



clemb

On the server side

```
Administrator: Command Prompt
C:\>cd automationweb
C:\AutomationWeb>dir
Volume in drive C has no label.
Volume Serial Number is 700B-9CFD

Directory of C:\AutomationWeb

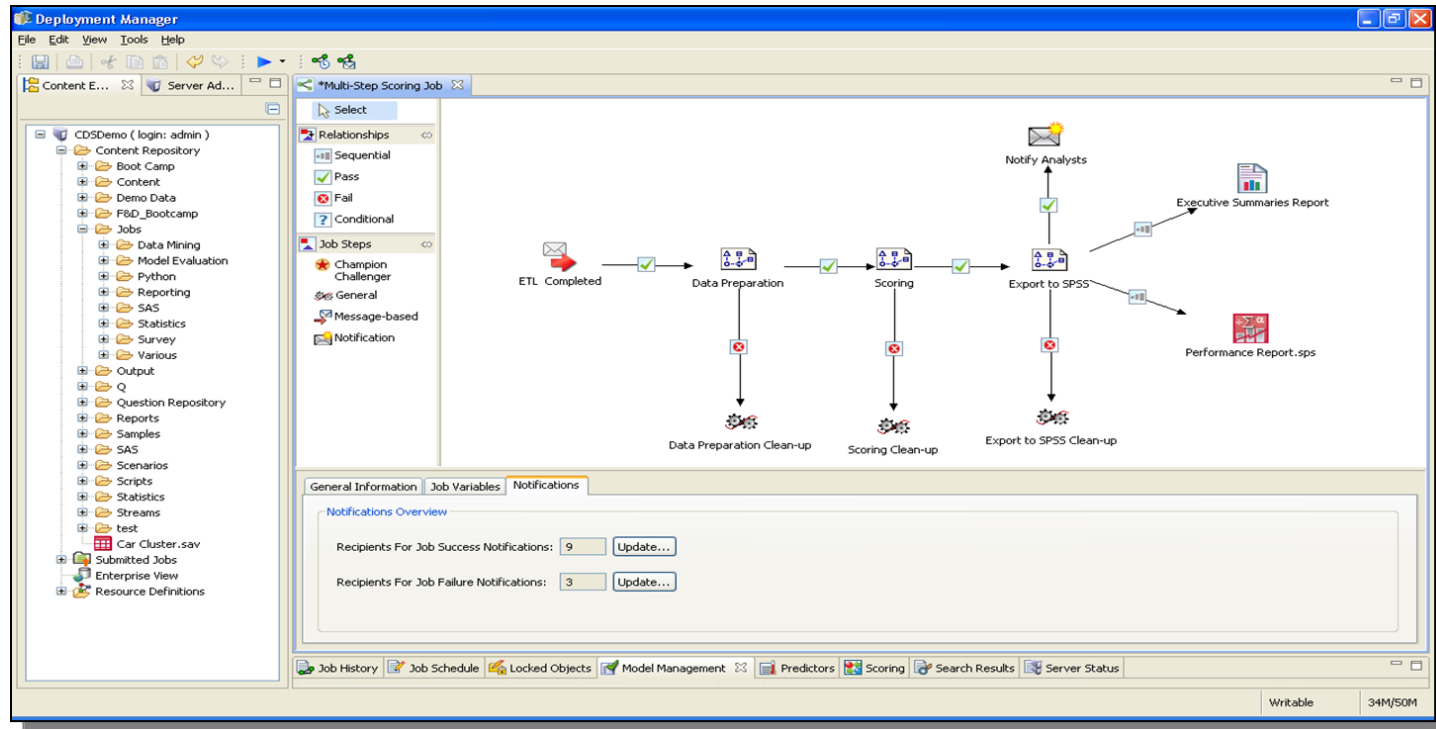
02/12/2014  16:10    <DIR>          .
02/12/2014  16:10    <DIR>          ..
02/12/2014  10:15             78,336 Automating your analyses - the best
03/12/2014  09:23          556,091 Automating Your Analyses v0.1.pptx
03/12/2014  09:03             602 BuildStreamEnd.txt
24/11/2013  17:16             602 BuildStreamEnd.txt-
24/11/2013  11:38             313 BuildStreamV2.txt
13/03/2009  15:48          341,645 census.sav
28/10/2009  13:31          14,055 customer satisfaction 2008.sav
02/12/2014  10:16          833,761 Getting Started with SPSS Statistic
03/12/2014  15:19             302 KDA(1).sps
20/03/2009  14:02             510 Macro 0.sps
20/03/2009  14:13             565 macro 1.sps
20/03/2009  14:52             1,914 macro 2.sps
13/03/2009  15:48             1,761 macro 3.sps
20/03/2009  16:38             5,289 macro 4.sps
02/12/2014  09:30          153,096 Offer Slide.pptx
15 File(s)          1,988,842 bytes
2 Dir(s)  69,334,593,536 bytes free

C:\AutomationWeb>statisticsb -f KDA(1).sps -type text -out KDA.txt
C:\AutomationWeb>_
```

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



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



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\  
!ENDDDEFINE.
```

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  
  
!ENDDDEFINE.
```

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 (statistics only)
- 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

Automating beyond streams - Modeler

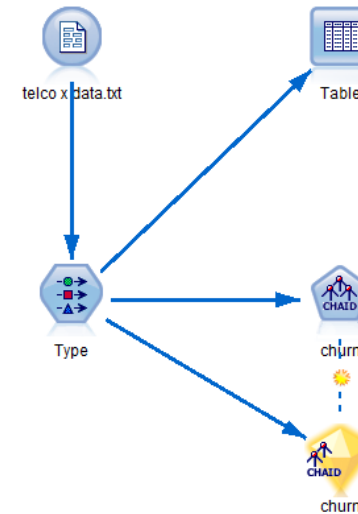
- Scripting
 - Legacy
 - Python
- R Nodes

There is a general direction in the IBM/SPSS products to integrate and apply Python more

An example script in Modeler

```
Standalone Script
Script Python Legacy

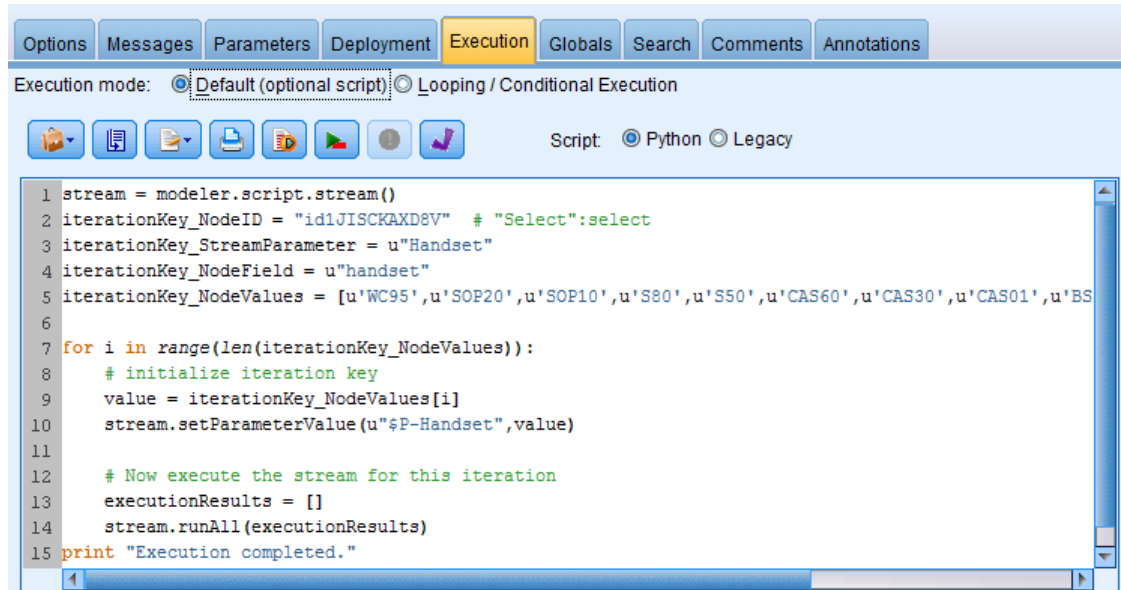
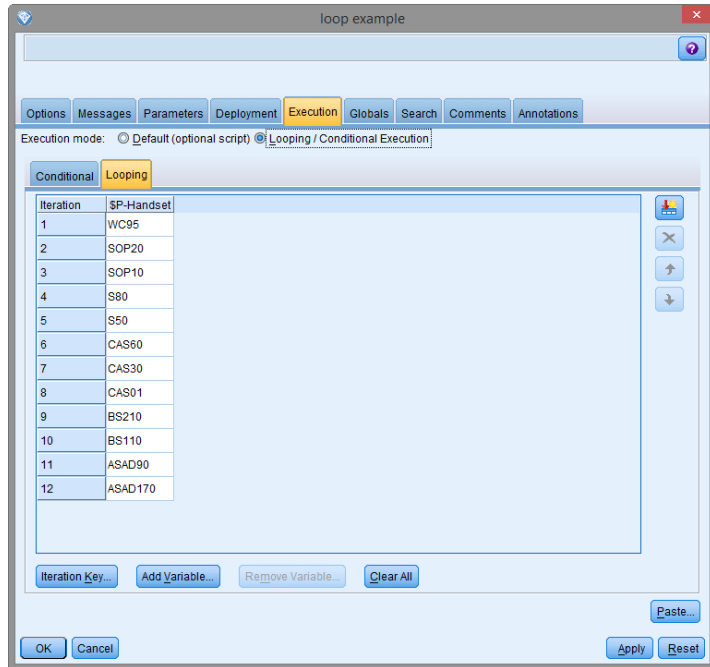
1 # Var file
2
3 create variablefilenode at 200 200
4 set :variablefilenode.full_filename = "C:/Train/Modeler_Scripting/telco x data.txt"
5 set :variablefilenode.read_field_names = True
6
7 #type
8
9 create typenode at 200 400
10 set :typenode.direction.'customer_id' = None
11 set :typenode.direction.'retention' = None
12 set :typenode.direction.'churn' = Target
13 connect :variablefilenode to :typenode
14
15 #table
16 create tablenode at 400 200
17 connect :typenode to :tablenode
18
19 #this will cause the type node to instantiate
20 execute :tablenode
21
22 #chaid
23 create chaidnode at 400 400
24 connect :typenode to :chaidnode
25 execute :chaidnode
26
```



This script:

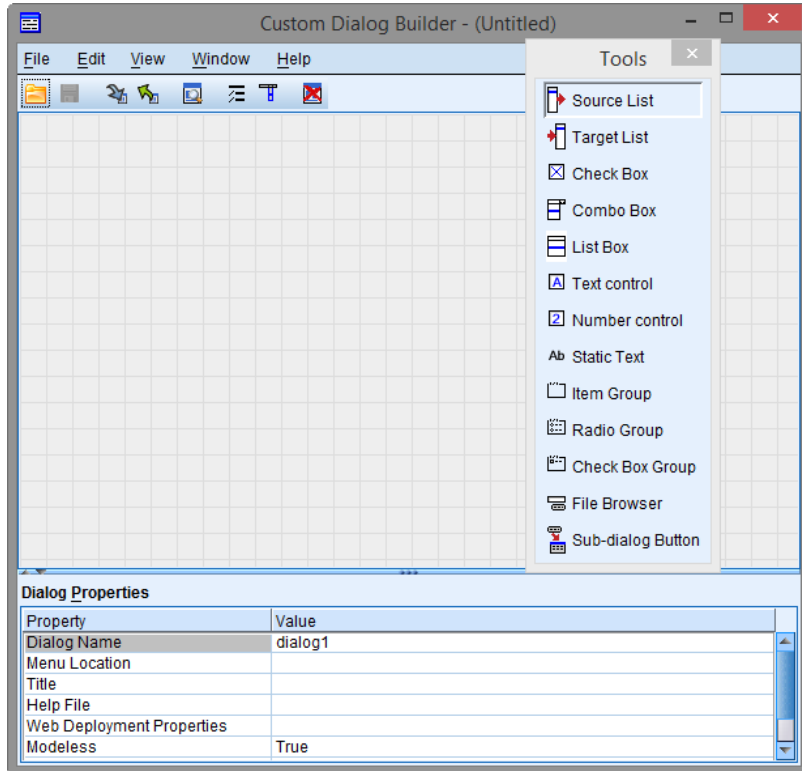
- a) creates a simple stream
- b) Read and displays data in a Table
- c) Builds a CHAID model

An example python script - looping



In the **looping and conditional execution** option in Modeler we can **Paste** Python syntax to execute the operation

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

Key Drivers Analysis

Variables:

- respondent id [resp_id]
- year of survey [year]
- Gender [gender]
- Age [age]
- Family Income [income]
- Location of Store that one has visited [store...]
- department
- Reason for shopping at store [reason]
- Performance: Politeness Staff [perf1]
- Performance: Service [perf2]
- Performance: Attractive Shopping [perf3]
- Stopped shopping at our store [cancel]
- Reason canceling shopping [reason_cancel]

Dependent Variable (Overall Rating)

Overall Satisfaction [satisfaction]

Characteristics

- Satisfaction: Politeness Staff [sat1]
- Satisfaction: Service [sat2]
- Satisfaction: Attractive Shopping [sat3]

Regression Method

☒ Forced Entry

☐ Stepwise

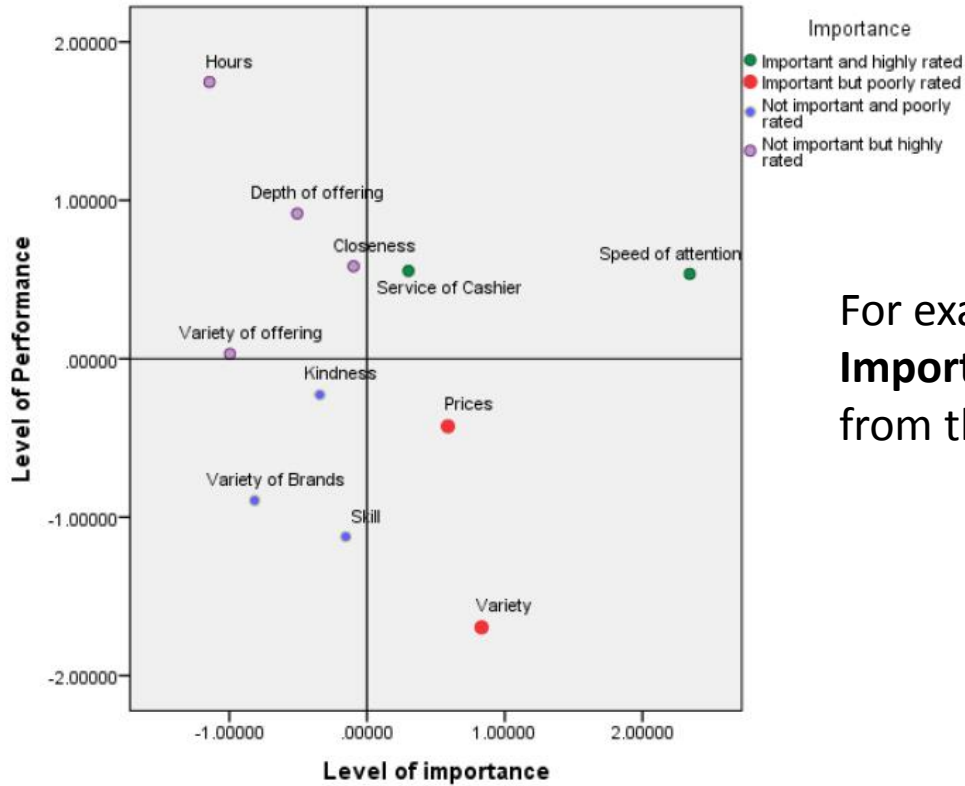
Copyright Channel Group Inc/SV Europe

OK Paste Reset Cancel

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

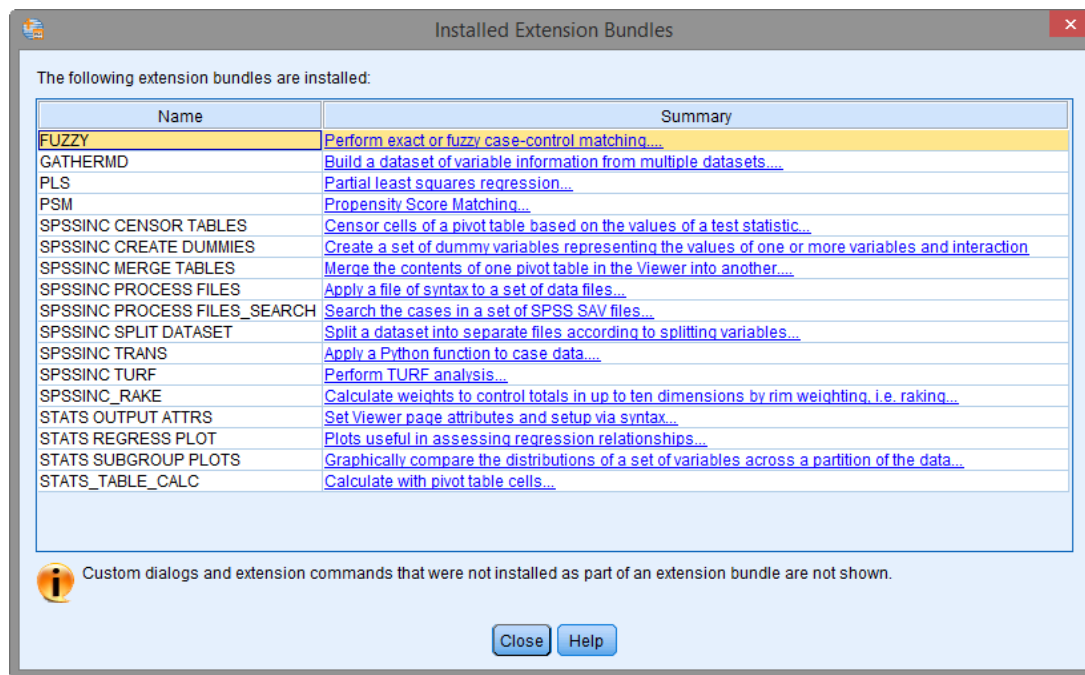
It simplifies several steps beyond the KDA syntax that we ran earlier

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

Check out the SPSS Developer Central for more resources

www.ibm.com/spss/devcentral

The Create Dummy Variables extension

Create Dummy Variables

Variables:

- id
- NUMBER OF HOURS WORKED LAST WEEK [...]
- RS OCCUPATIONAL PRESTIGE SCORE (198...
- NUMBER OF BROTHERS AND SISTERS [sibs]
- NUMBER OF CHILDREN [chlds]
- AGE OF RESPONDENT [age]
- HIGHEST YEAR OF SCHOOL COMPLETED [e...]
- HIGHEST YEAR SCHOOL COMPLETED, FAT...
- HIGHEST YEAR SCHOOL COMPLETED, MOT...
- HIGHEST YEAR SCHOOL COMPLETED, APP...

Create Dummy Variables for:

LABOR FRCE STATUS [wrkstat]

Main Effect Dummy Variables

☒ Create main-effect dummies

Root Names (One Per Selected Variable):

WORKING_STAT

Macro Name:

Two-Way Interactions

☐ Create dummies for all two-way interactions

Root Name:

Macro name:

Three-Way Interactions

☐ Create dummies for all three-way interactions

Root Name:

Macro name:

Dummy Variable Labels

☒ Use value labels

☐ Use values

Value Order

☒ Ascending

☐ Descending

Macros

☐ Omit first dummy category from macro definitions

Note: It is conventional to start macro names with 1.

Measurement Level Usage

☒ Do not create dummies for scale variable values

☐ Create dummies for all variables

This dialog requires the Python Essentials

OK Paste Reset Cancel Help

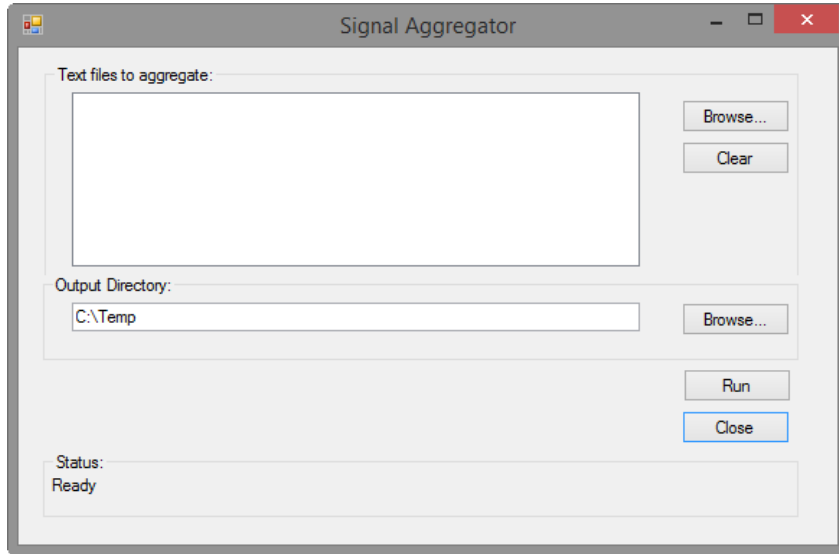
Variable Creation

	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

Contents

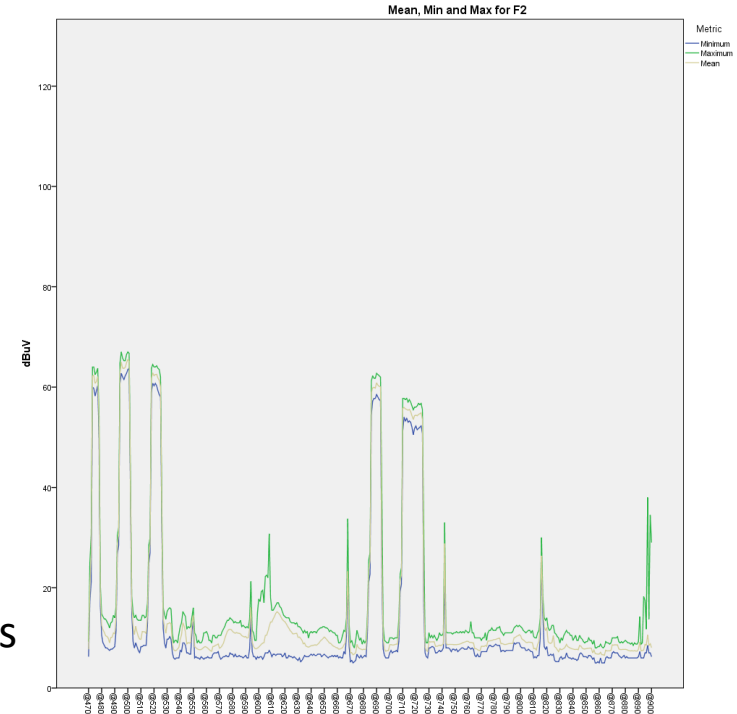
- Background
- Levels of automation with syntax and streams
- Automating beyond syntax and streams
- Automating SPSS from the outside

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





Genie

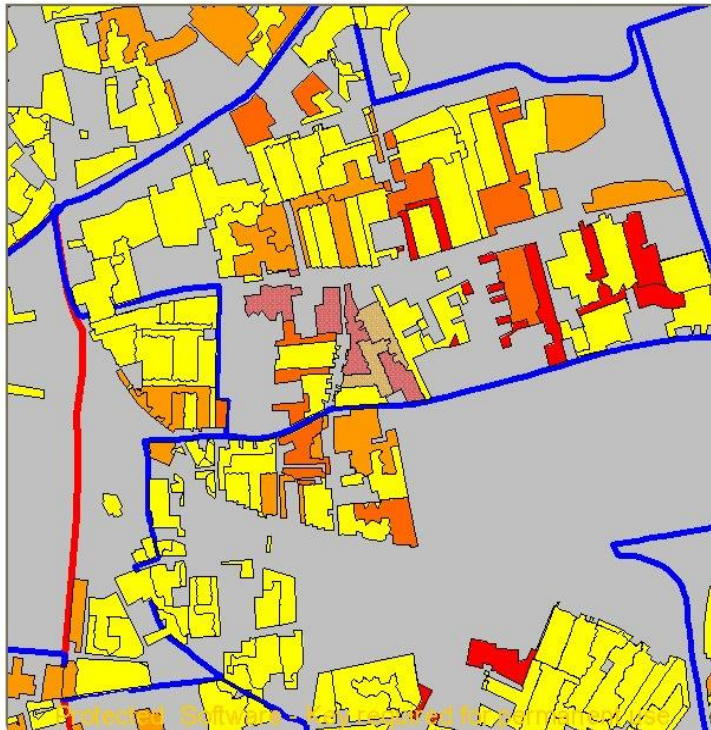
Geographically Enabled Network Intelligence Engine

[About GENIE](#) | [Preferences](#) | [Log Out](#)

Logged in as: *jmcconnell*

[Welcome](#) > [Select data](#) > [Work with data](#) > [View Indicators as Map](#)

View indicators on a map ([« Back to work with data](#))



Session tools

[Save this session data](#)
[Discard this session](#)
[Export this data](#)

Map tools

Legend

Information tool

Zoom In

Zoom Out

Re-centre map

Select Elements

Apply Selections To Map

Discard Selections

Display Summary Statistics

Add thematic layer

Remove thematic layer

Layer control

View all layers

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

- 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?

Exclusive Smart Vision Bundle Offers



SPSS Automation Services

- If you already have SPSS Base & are looking for help
 - 10% discount on services
 - £1080 per day
 - Expect a project to be 2 -5 days
 - Offer code **SVAUTO001**

SPSS Automation Starter Pack

- If you do not have SPSS licence already
 - 1 user perpetual licence of SPSS Base & Tables
 - 1st year support + 2 days personalised training
 - £4228 (+VAT)
 - Saving £1000 on IBM list price
 - Offer Code **SVAUTOSP002**



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