Automating with IBM SPSS

John McConnell – Services

Rachel Clinton – Business Development

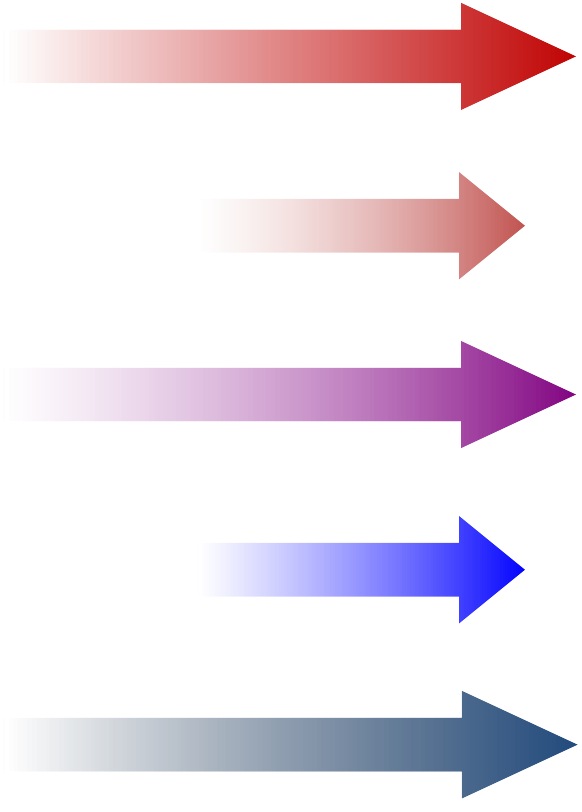
Contents

- 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

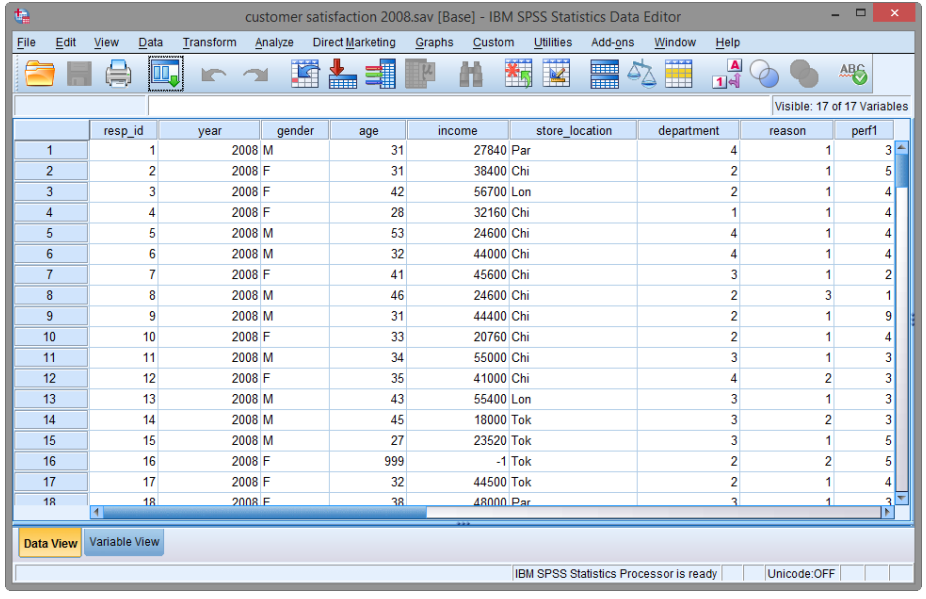


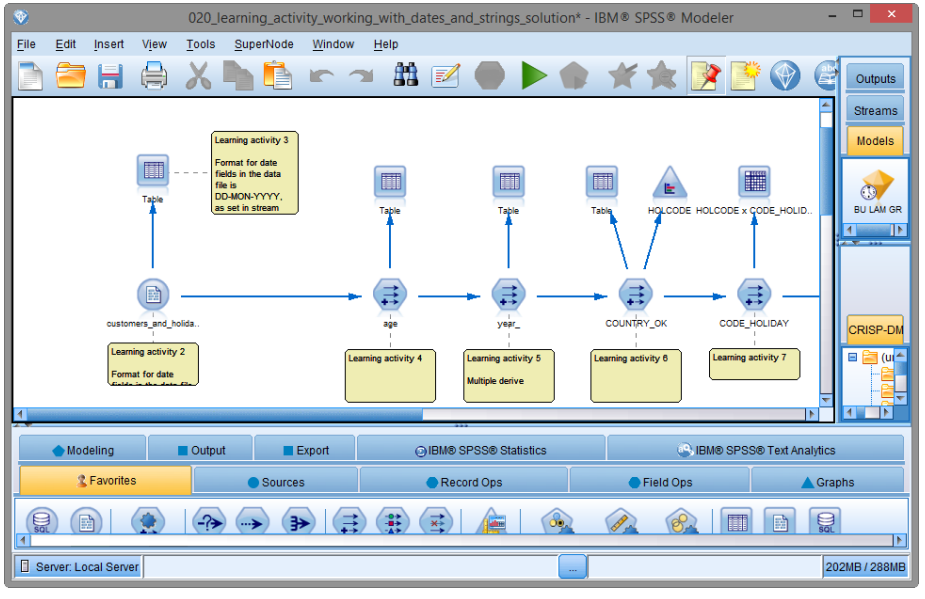
- Productivity
- Repeatability
- Governance
- Communication
- Delegation

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



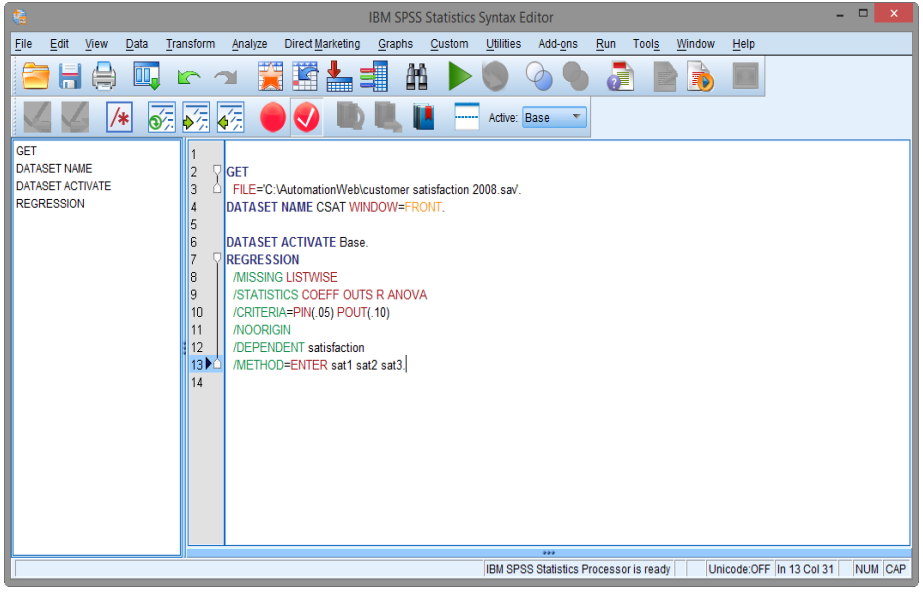


Statistics **Modeler**

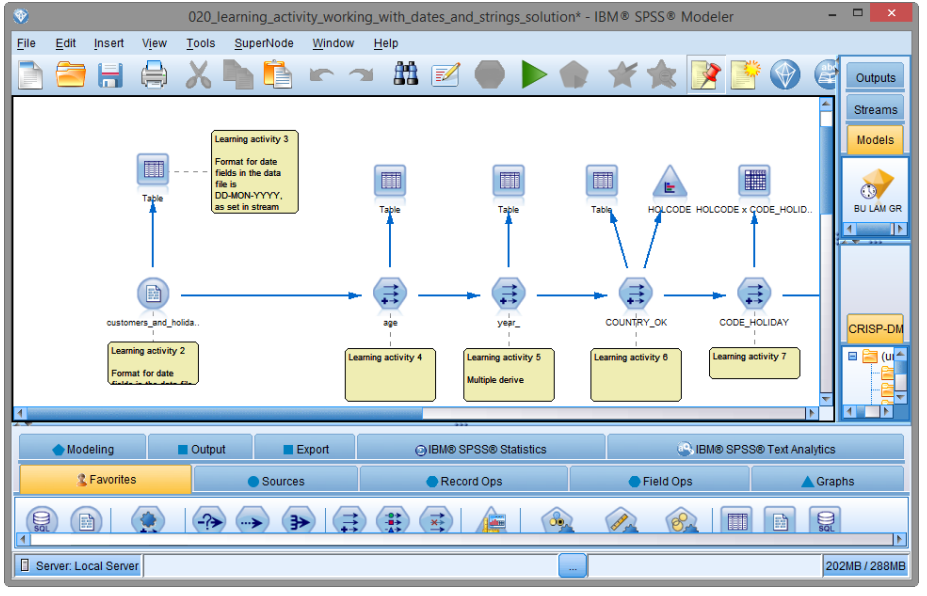
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Automation – Level 1



Syntax

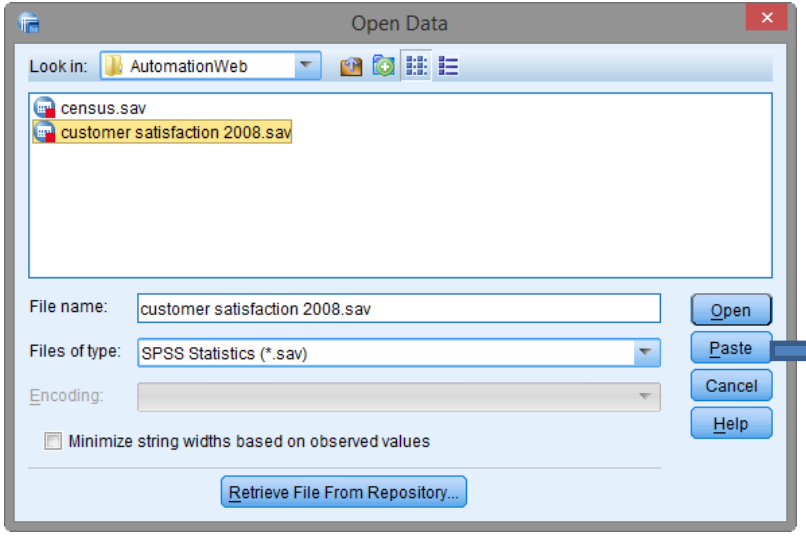


Streams

SMARTVISION Europe

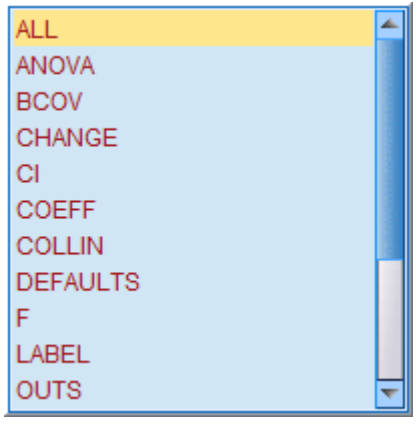
A SELECT INTERNATIONAL COMPANY

Defining and pasting



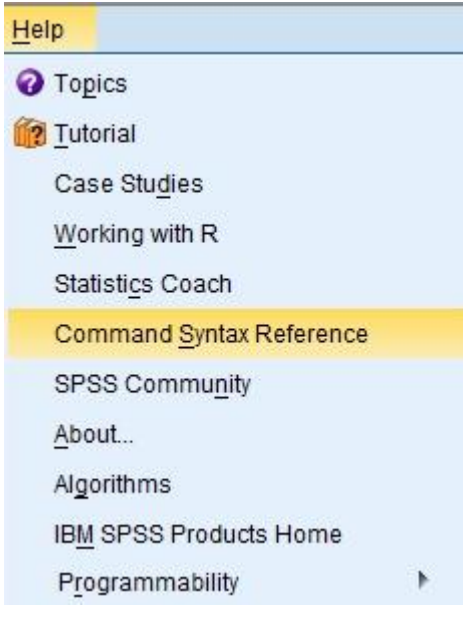

```
GET
  FILE='C:\AutomationWeb\customer satisfaction 2008.sav'.
DATASET NAME CSAT WINDOW=FRONT.

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



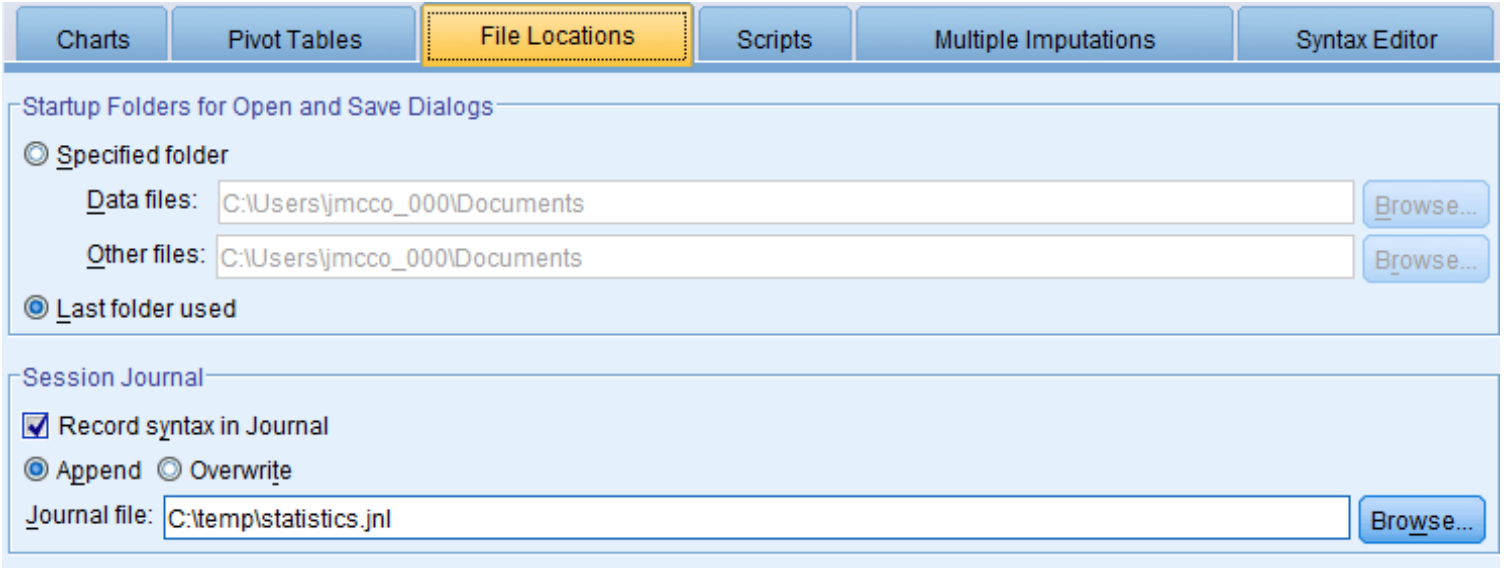
Auto or <ctrl>+<space>
Pops up relevant options

Tool to show us the syntax options for the selected command



The PDF of all commands and options

Forgot to Paste?

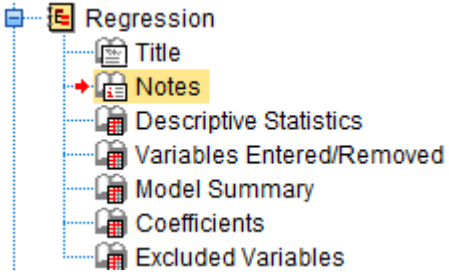


The **Journal File** is set (in **Edit > Options**) to record syntax automatically – until overwritten or deleted

Forgot to Paste?

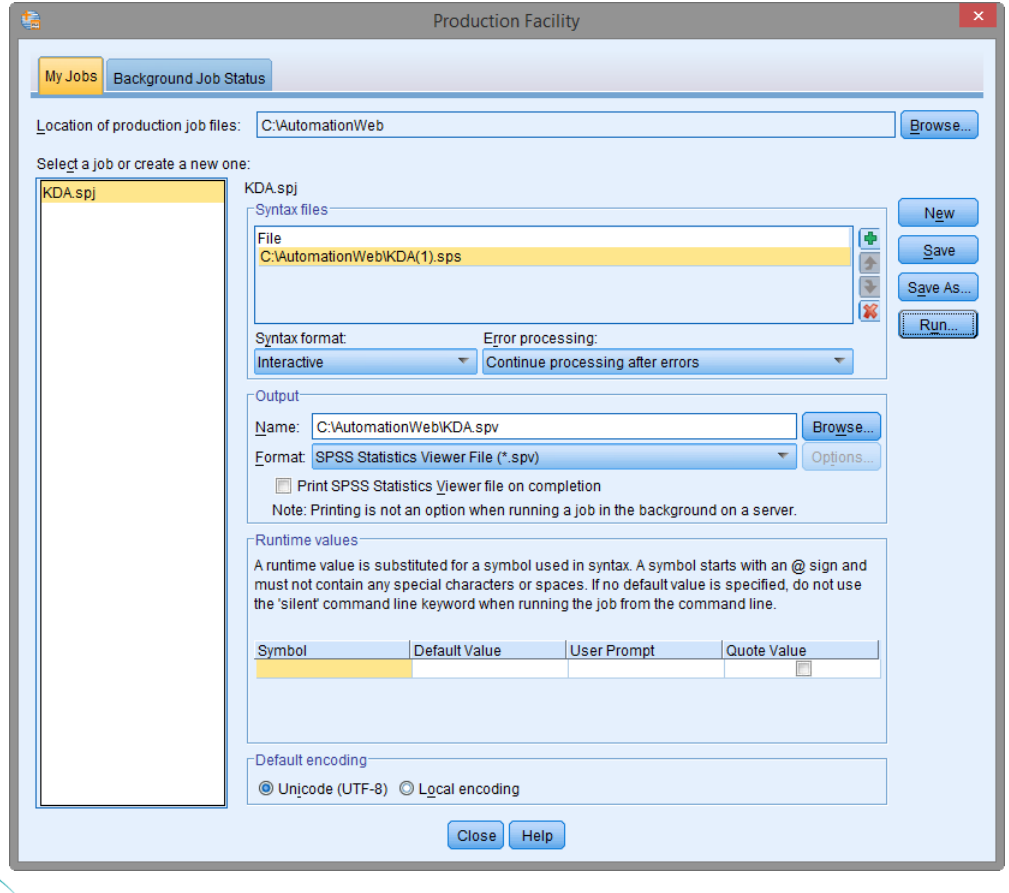
Notes

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

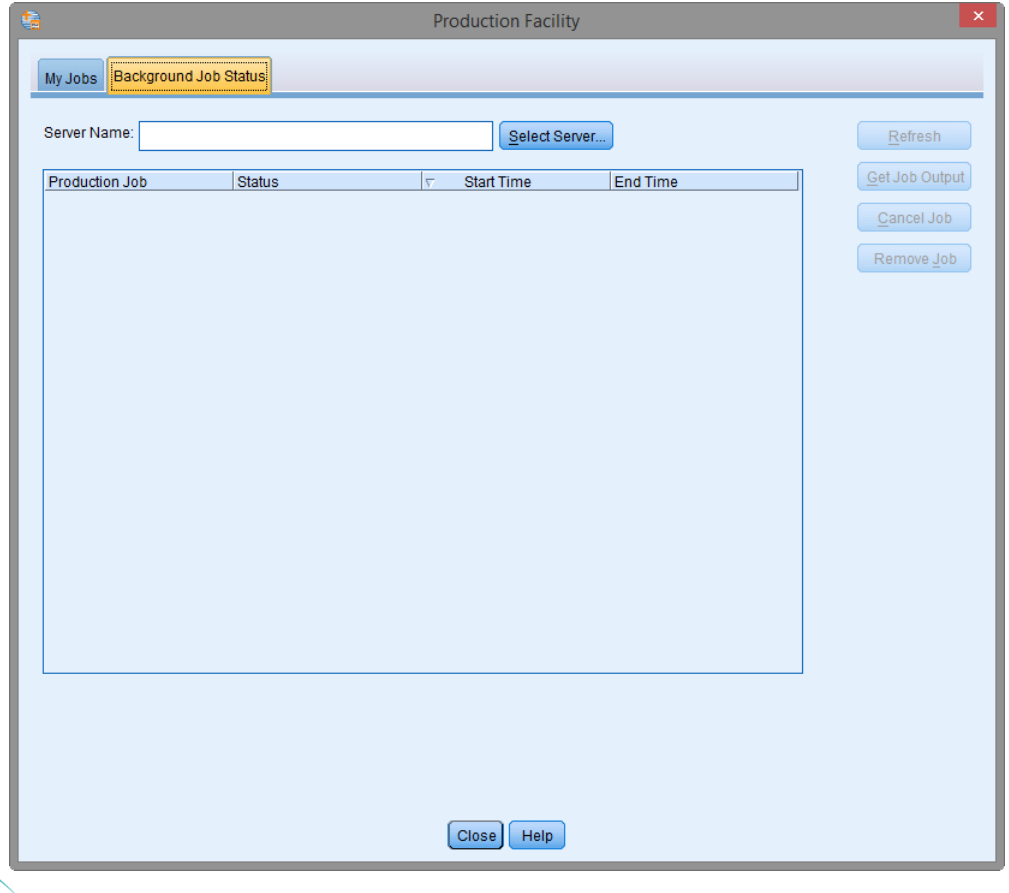
Batch running Syntax – The Production Facility



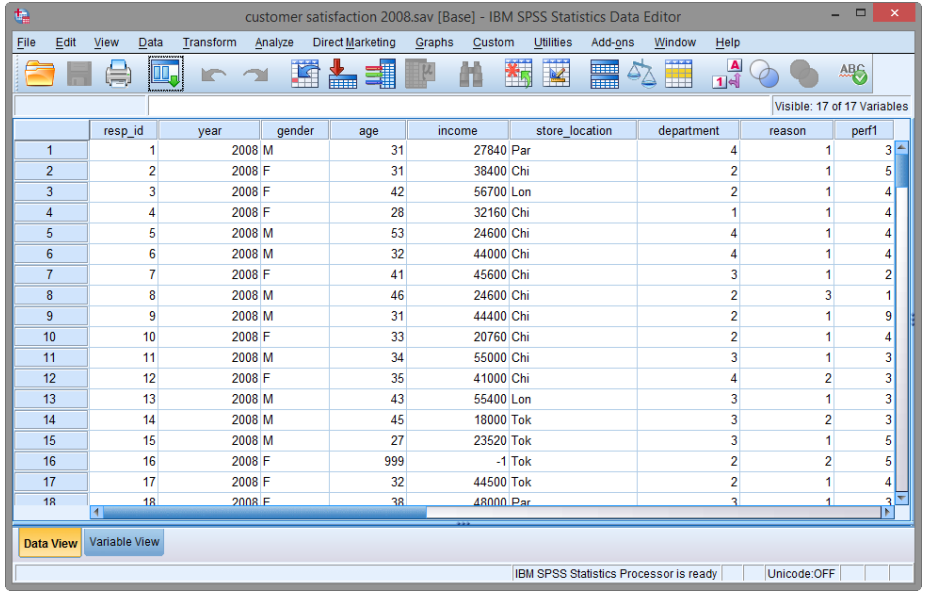
Menu path:

Utilities > Production Facility

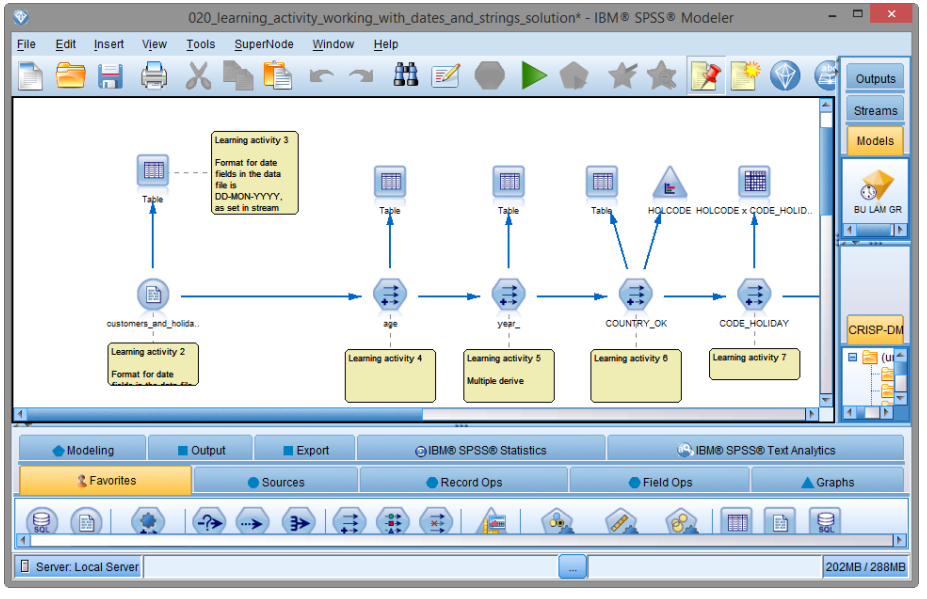
Background mode runs production jobs on a server



Server side batch engines



statisticsb

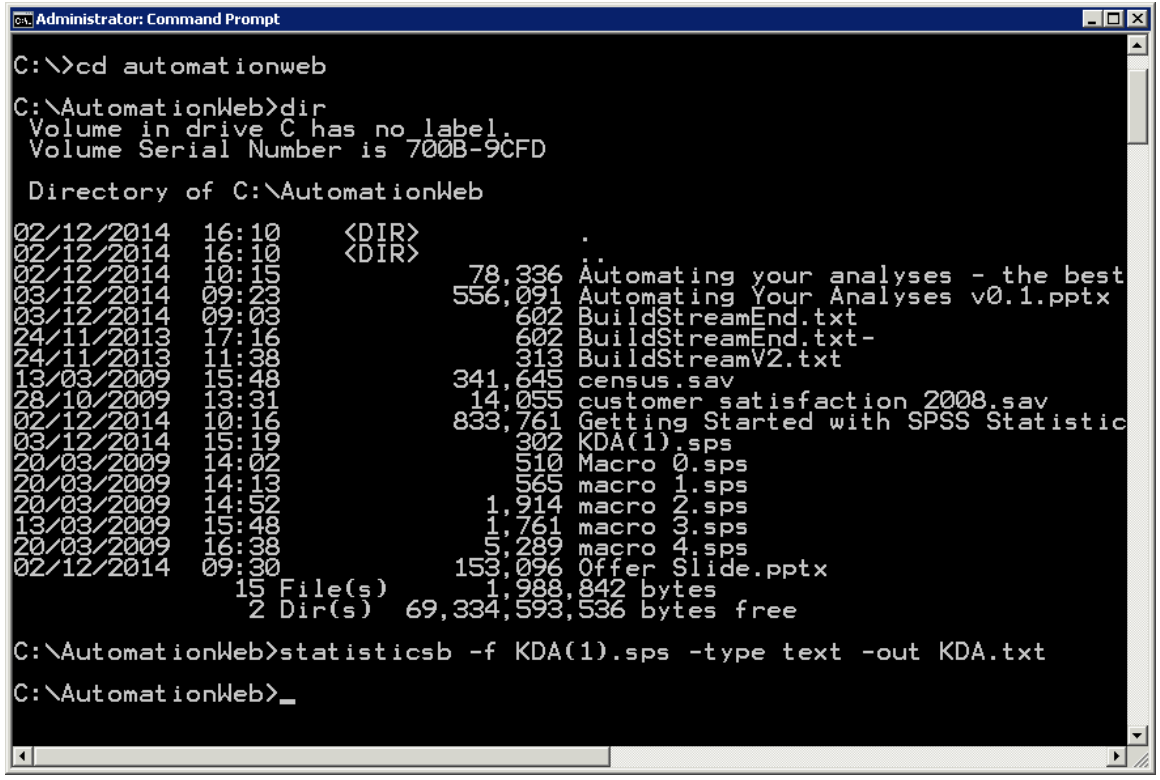


clemb

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

On the server side

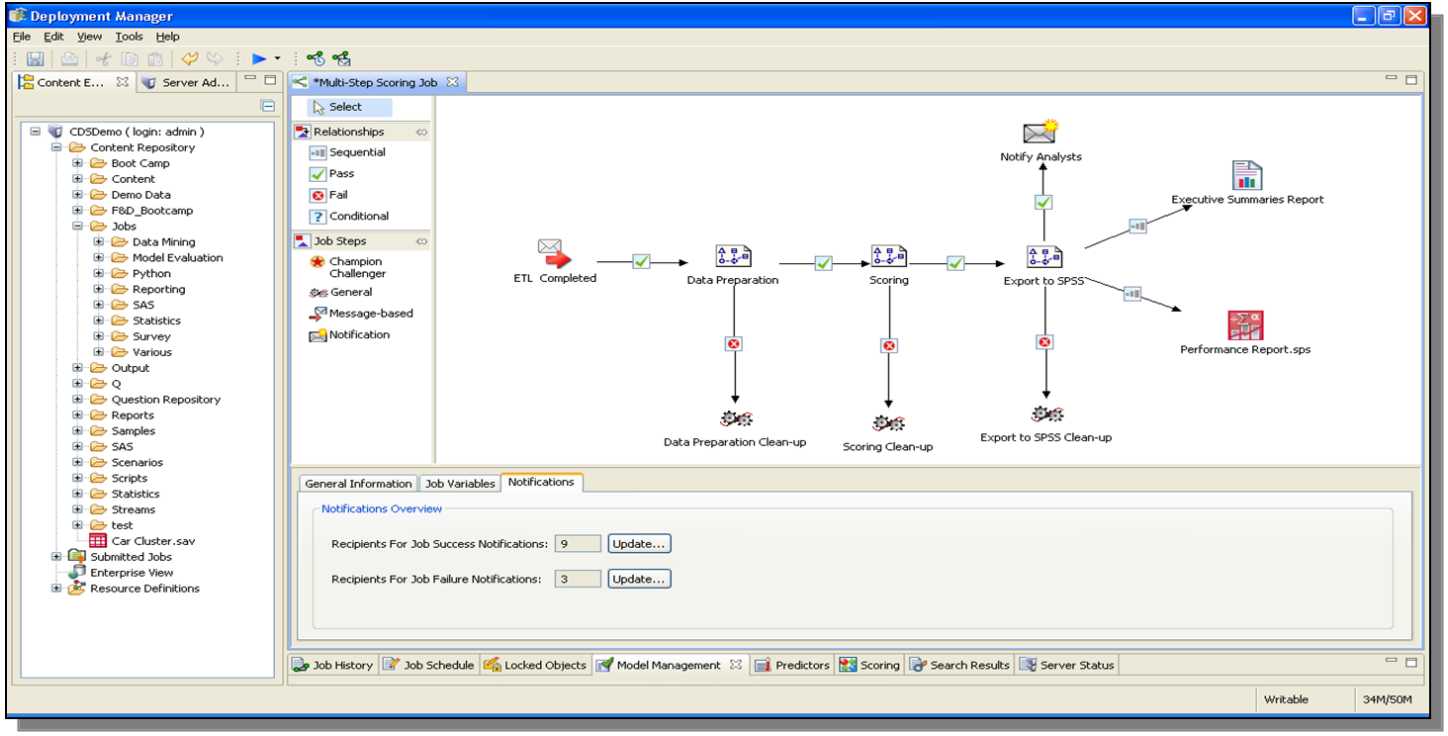


Batch jobs can be scheduled to run using the Windows Task Scheduler

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 (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
```

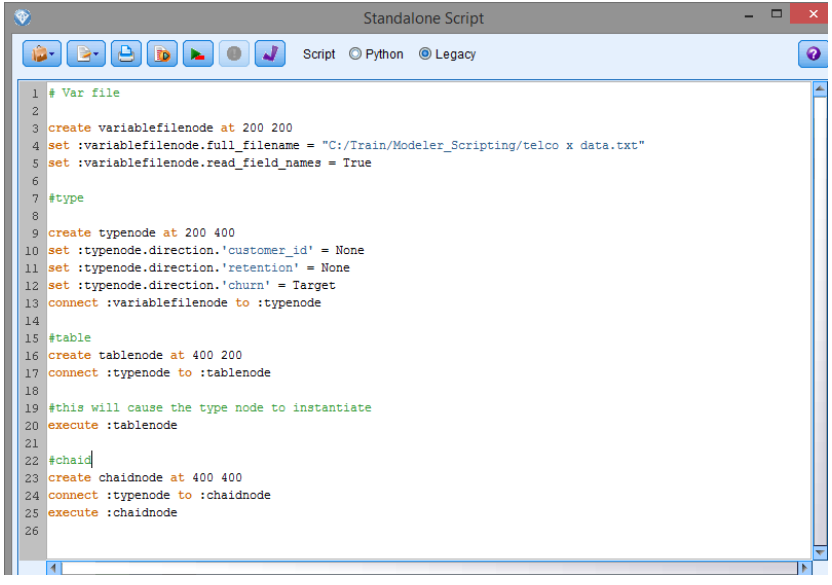
This script looks inside a correlation table
Identifies statistically significant correlations

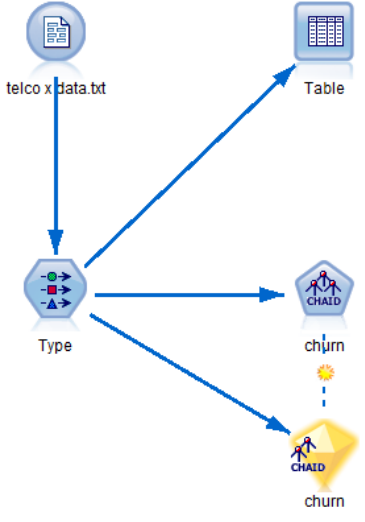
Automating beyond streams - Modeler

- 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

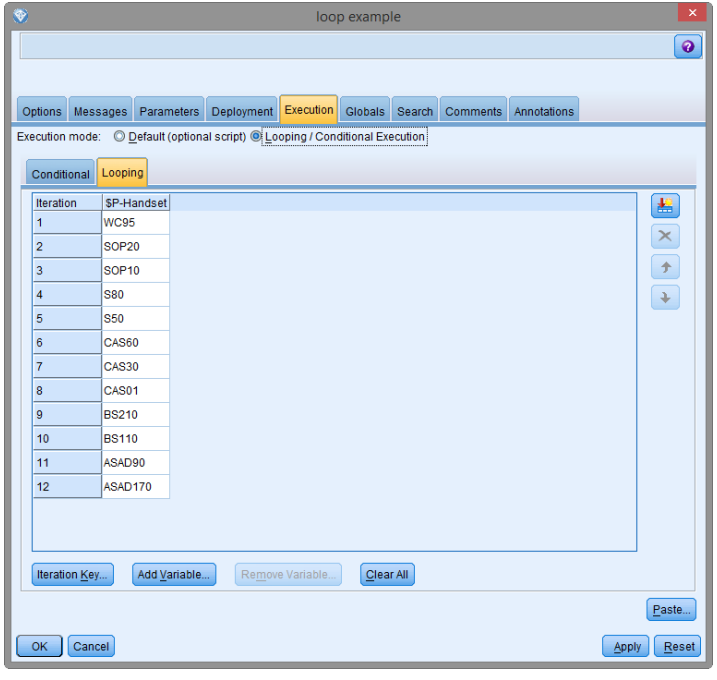


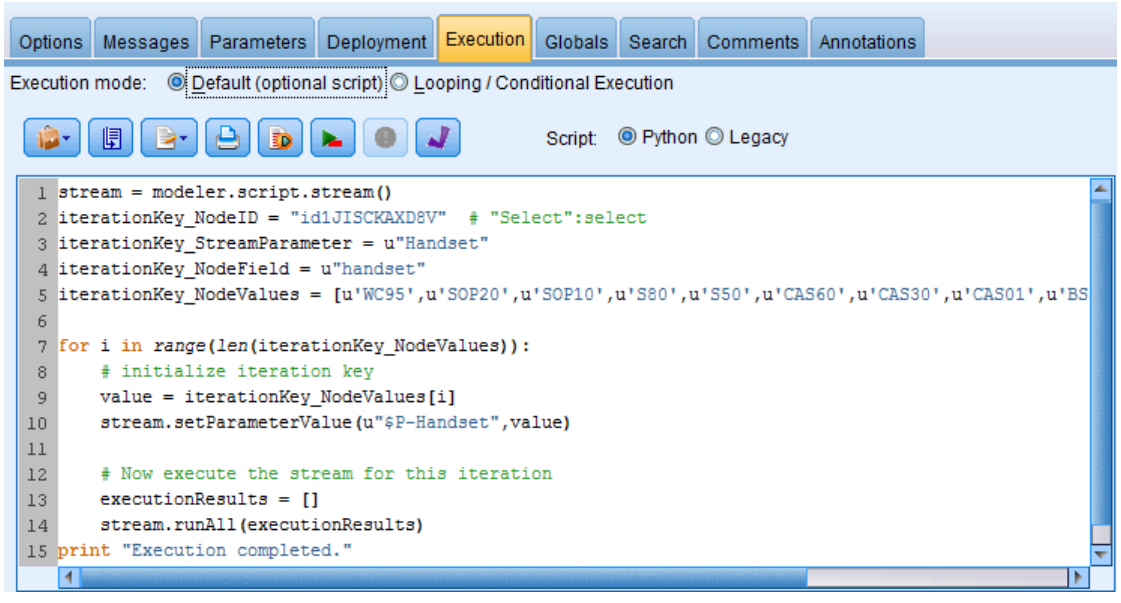


This script:

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

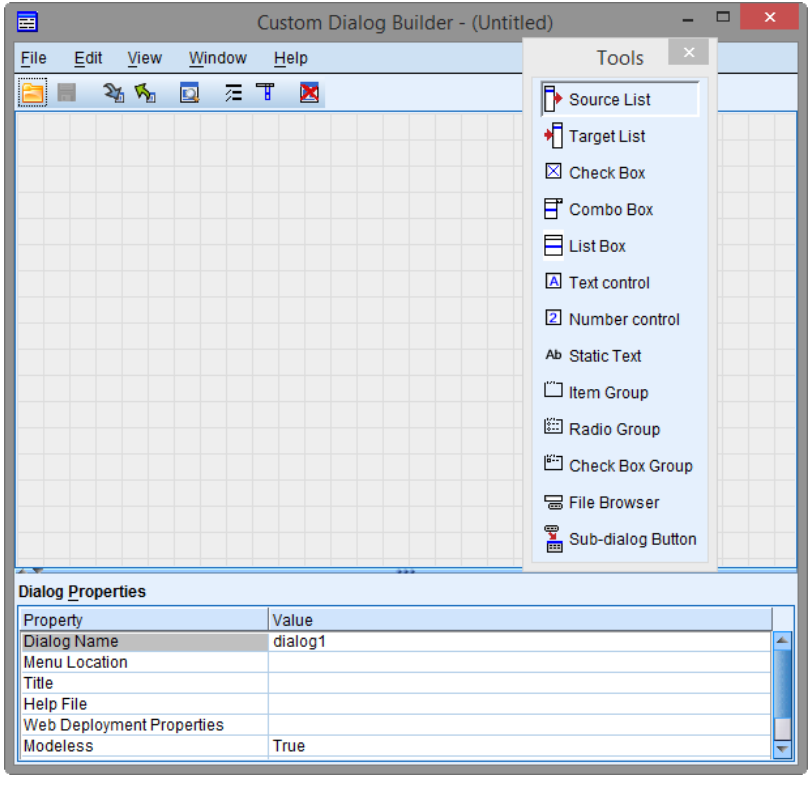
An example python script - looping





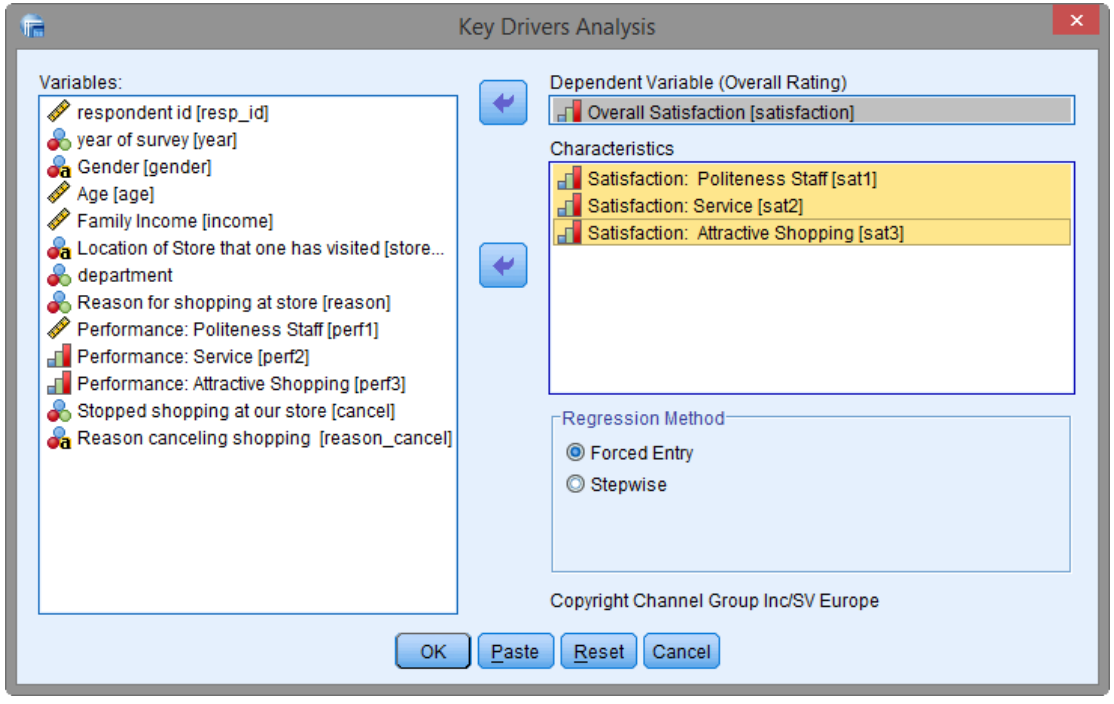
In the **looping and conditional execution** option in Modeler we can **Paste** Python syntax to execute the operation

Extensibility



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

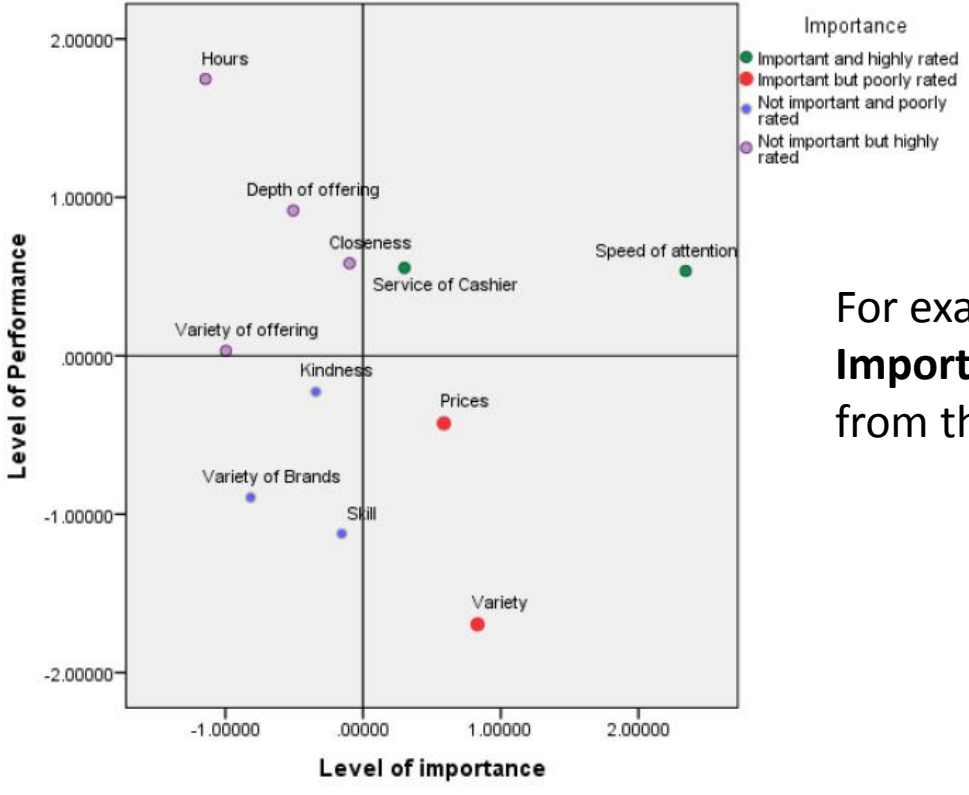
A KDA extension



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

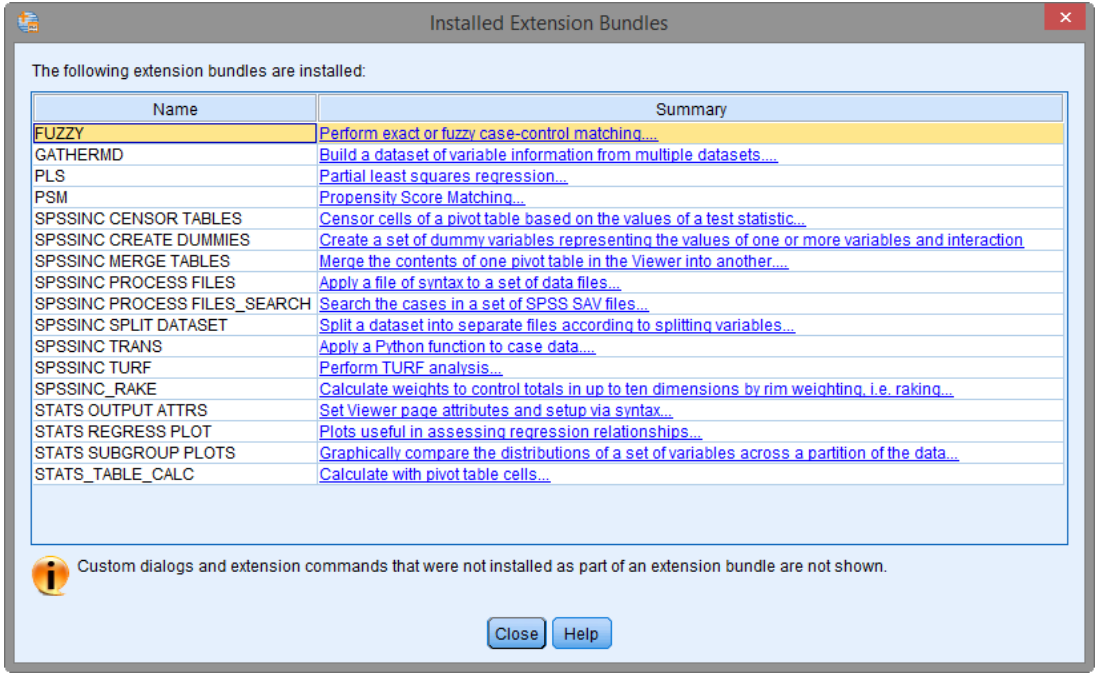
It simplifies several steps beyond the KDA syntax that we ran earlier

A KDA extension



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

Extension Bundles

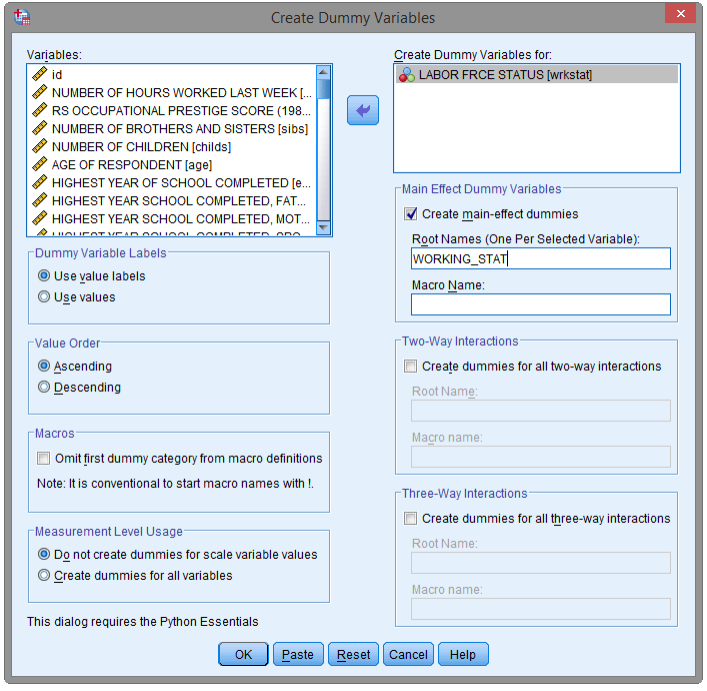


Typically written in Python

Check out the SPSS Developer Central for more resources

www.ibm.com/**spss**/dev**central**

The Create Dummy Variables extension



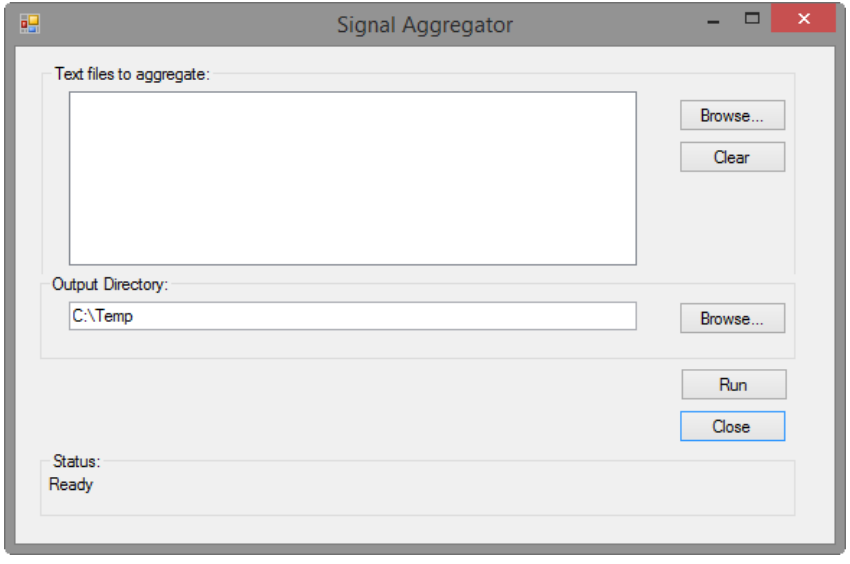
Variable Creation

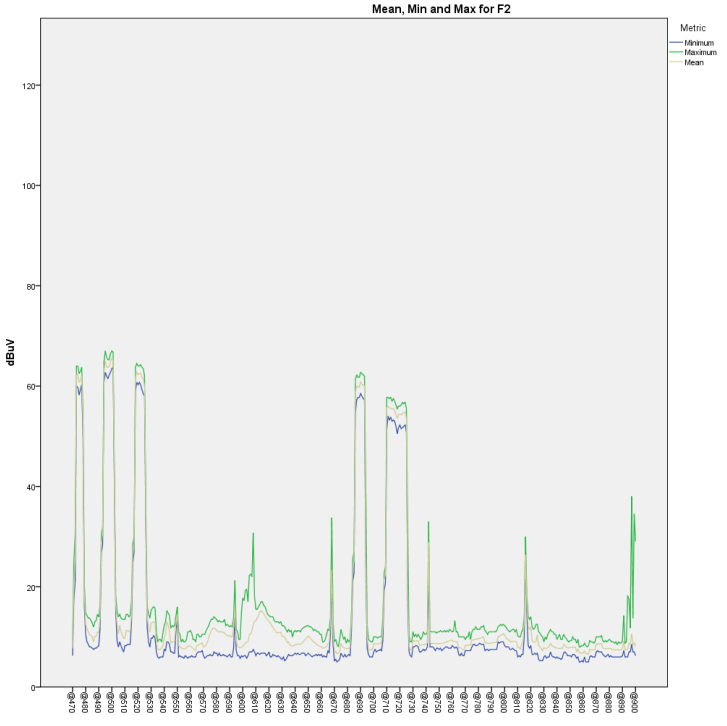
	Label
WORKING_STAT_1	wrkstat=WORKING FULLTIME
WORKING_STAT_2	wrkstat=WORKING PARTTIME
WORKING_STAT_3	wrkstat=TEMP NOT WORKING
WORKING_STAT_4	wrkstat=UNEMPL, LAID OFF
WORKING_STAT_5	wrkstat=RETIRED
WORKING_STAT_6	wrkstat=SCHOOL
WORKING_STAT_7	wrkstat=KEEPING HOUSE
WORKING_STAT_8	wrkstat=OTHER

Contents

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

Automating from the outside



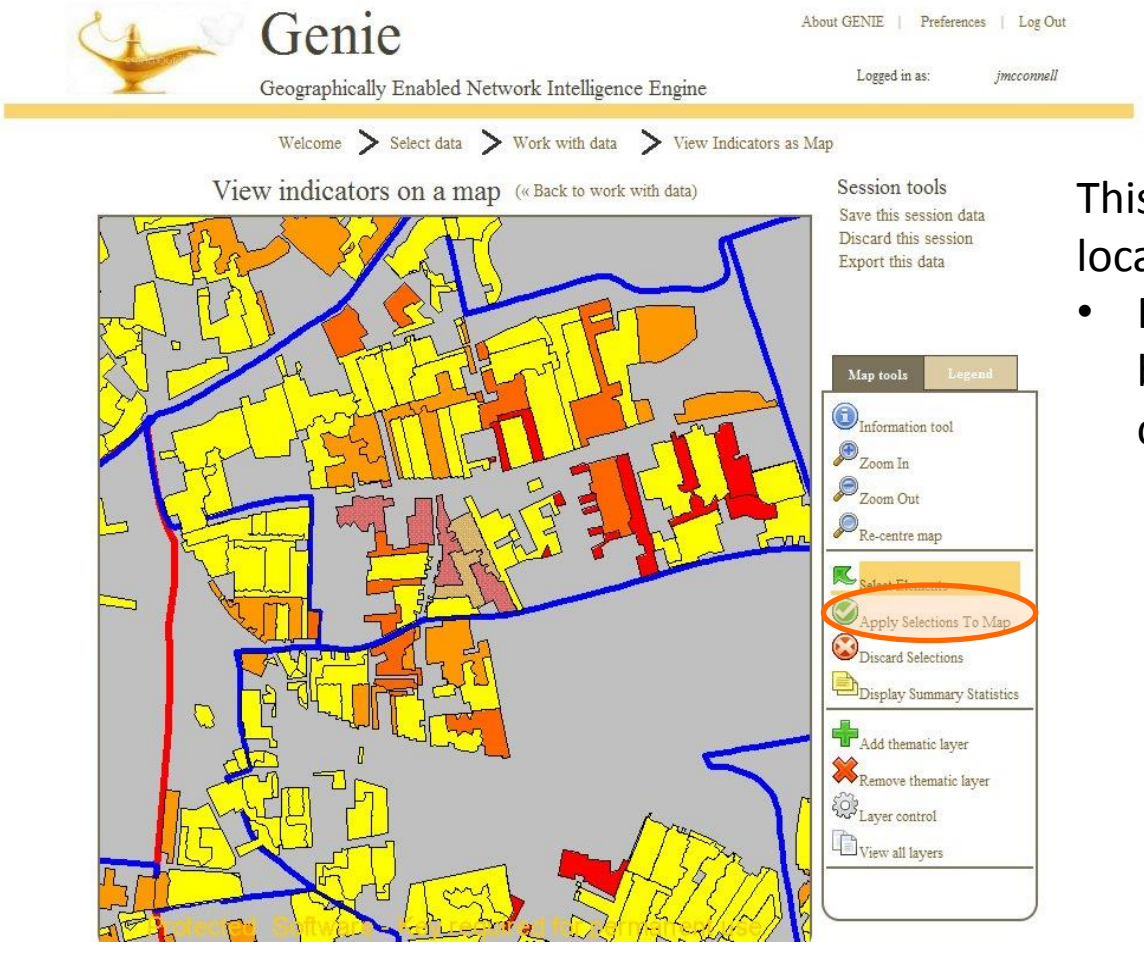


This UI runs a standalone app

a) Reads and cleans data coming form 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

- 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**

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