



Automating with IBM SPSS

John McConnell – Services

Rachel Clinton – Business Development

www.sv-europe.com

A SELECT INTERNATIONAL COMPANY

FAQ's

- Is this session being recorded? Yes
- Can I get a copy of the slides? Yes, we'll email a PDF copy to you after the session has ended.
- Can we arrange a re-run for colleagues? Yes, just ask us.
- How can I ask questions? All lines are muted so please use the chat facility – if we run out of time we will follow up with you.

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

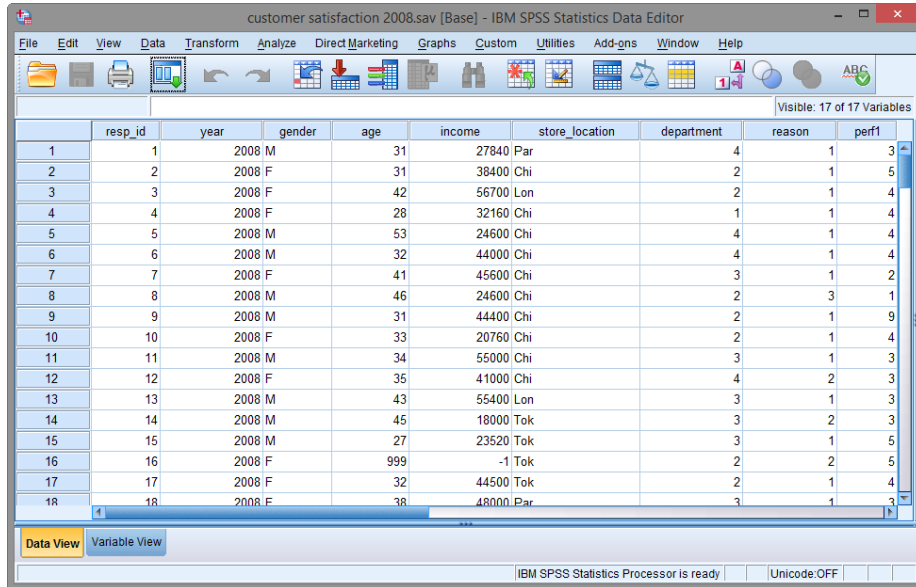


Delegation

Contents

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

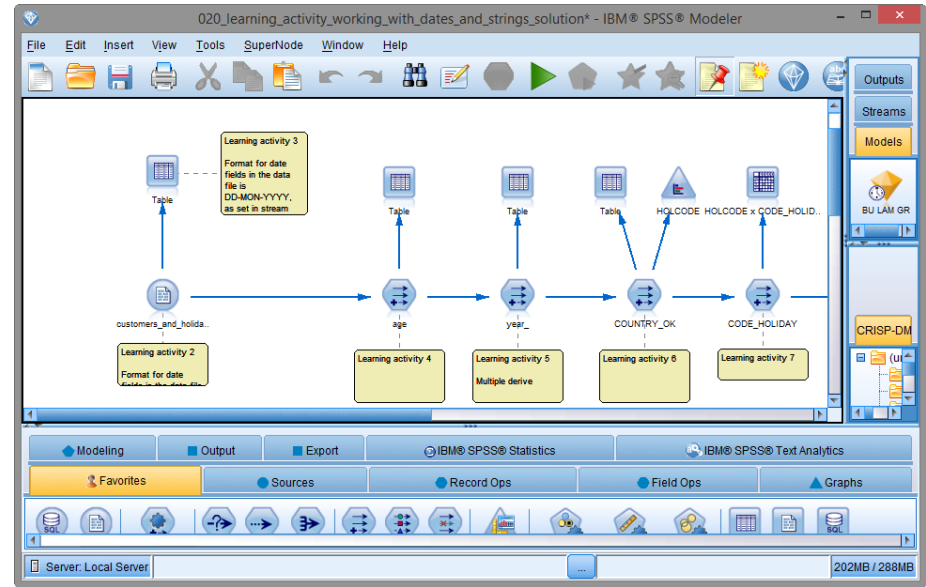
IBM/SPSS Statistics & IBM/SPSS Modeler



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

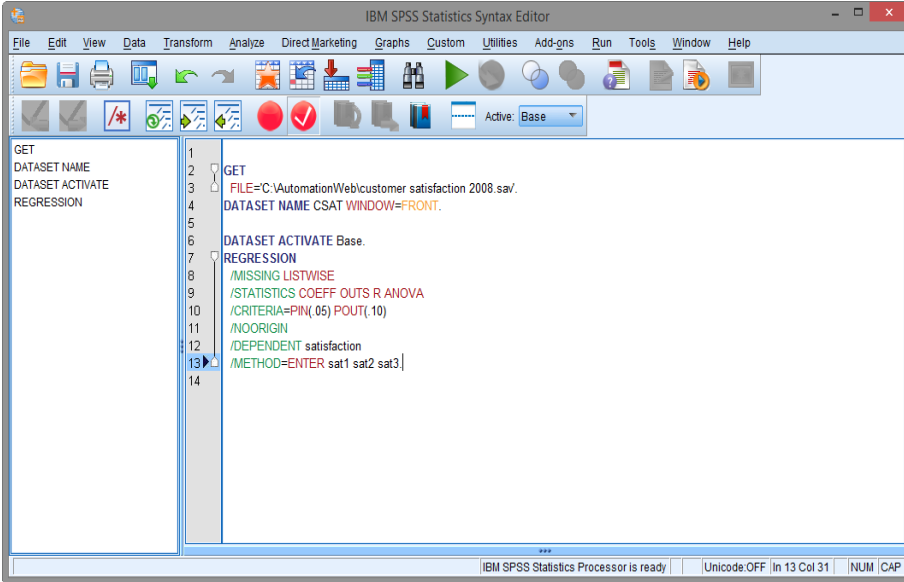
	resp_id	year	gender	age	income	store_location	department	reason	perf1
1	1	2008	M	31	27840	Par	4	1	3
2	2	2008	F	31	38400	Chi	2	1	5
3	3	2008	F	42	56700	Lon	2	1	4
4	4	2008	F	28	32160	Chi	1	1	4
5	5	2008	M	53	24600	Chi	4	1	4
6	6	2008	M	32	44000	Chi	4	1	4
7	7	2008	F	41	45600	Chi	3	1	2
8	8	2008	M	46	24600	Chi	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

Statistics

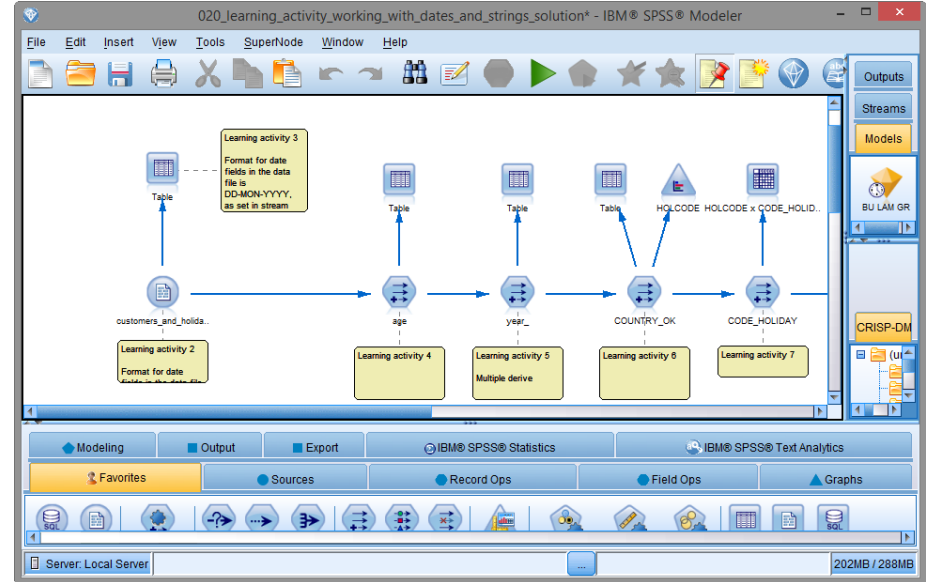


Modeler

Automation – Level 1

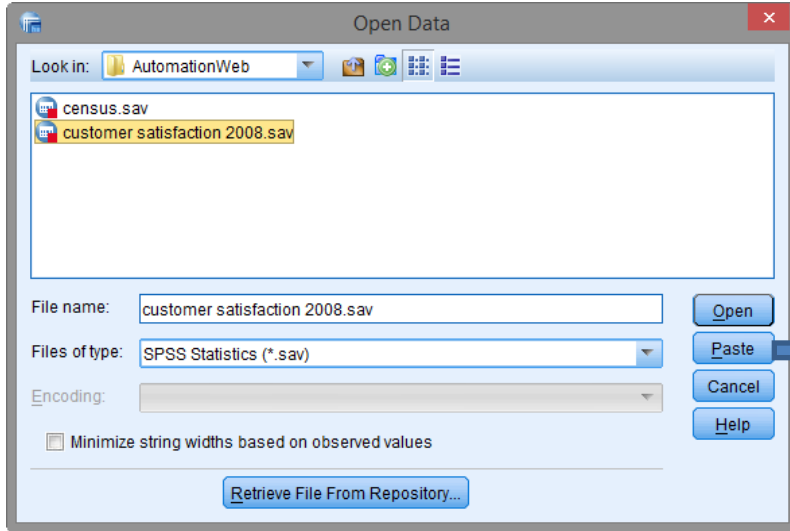


Syntax



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.

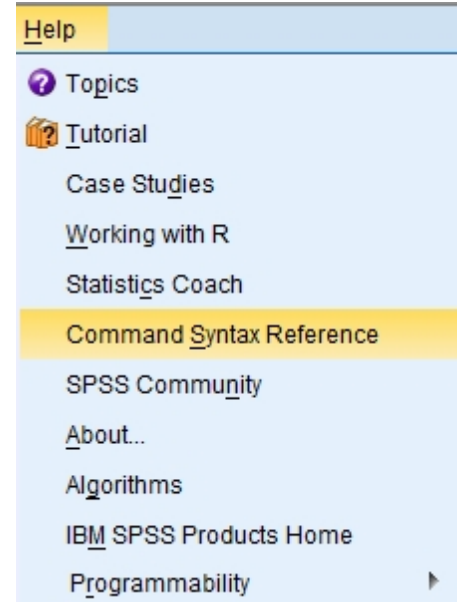
Getting help



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?

Charts Pivot Tables **File Locations** Scripts Multiple Imputations Syntax Editor

Startup Folders for Open and Save Dialogs

Specified folder

Data files: C:\Users\jmcco_000\Documents Browse...

Other files: C:\Users\jmcco_000\Documents Browse...

Last folder used

Session Journal

Record syntax in Journal

Append Overwrite

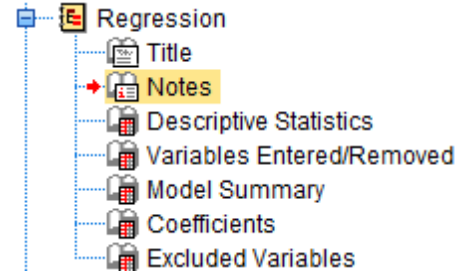
Journal file: C:\temp\statistics.jnl Browse...

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

Forgot to Paste?

Notes

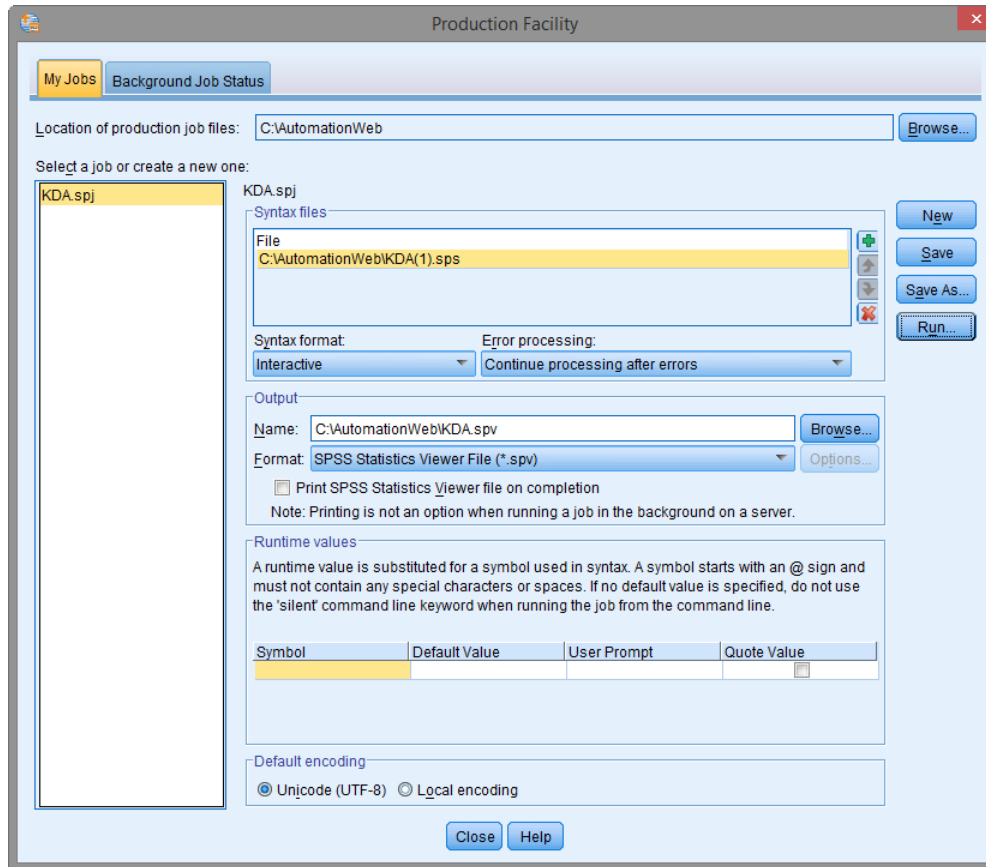
Output Created		03-DEC-2014 07:12:18
Comments		
Input	Data	C:\AutomationWeb\customer satisfaction 2008.sav
	Active Dataset	Base
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	140
Missing Value Handling	Definition of Missing	User-defined missing values are treated as 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

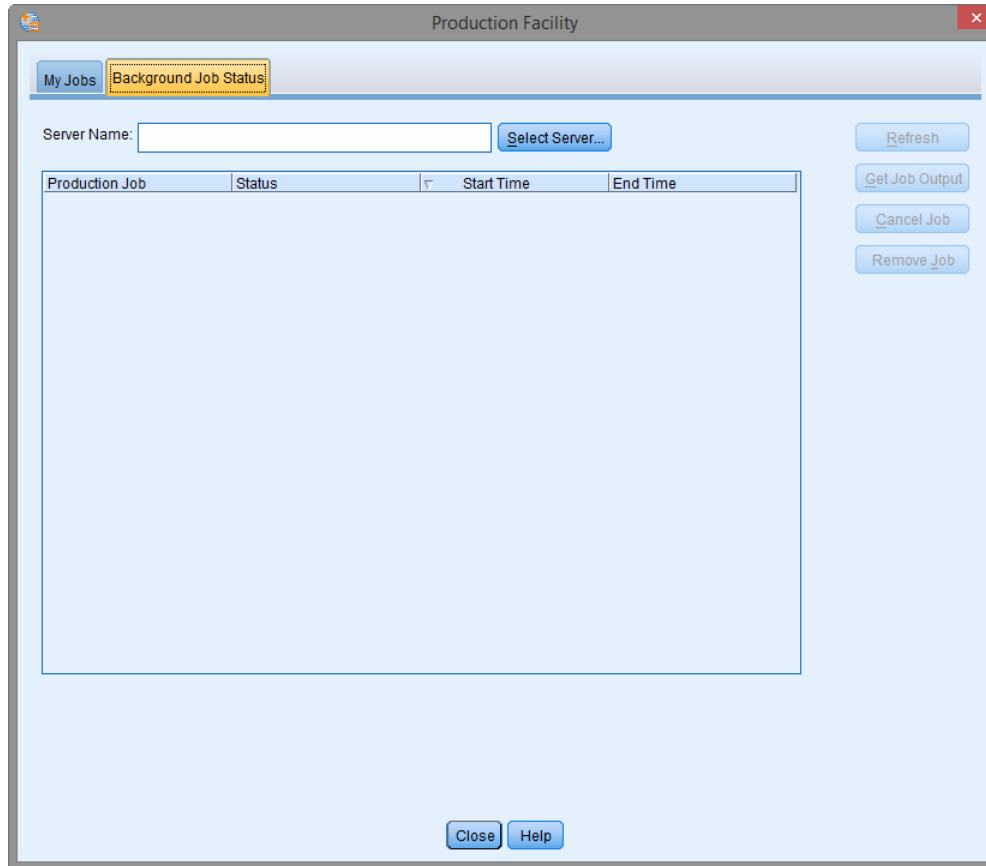


Batch running Syntax – The Production Facility

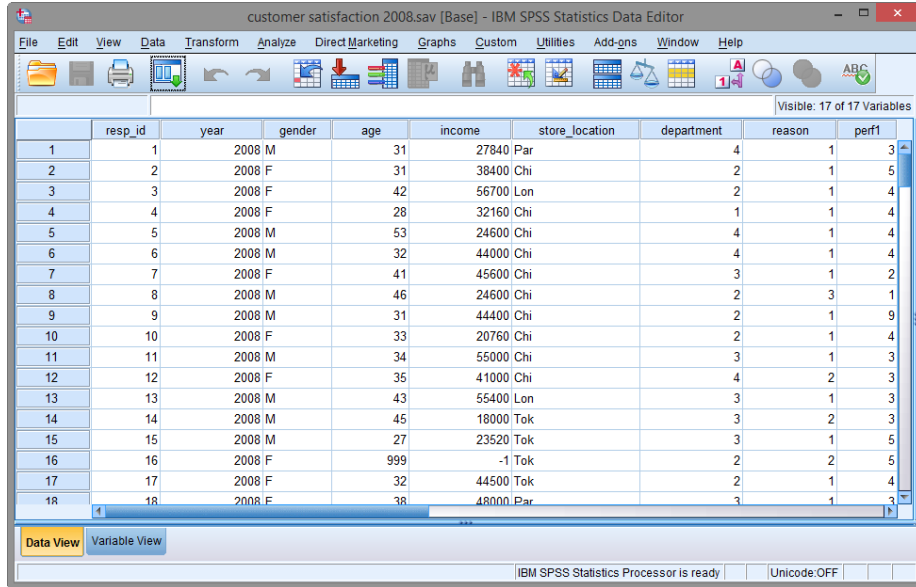


Menu path:
Utilities > Production Facility

Background mode runs production jobs on a server



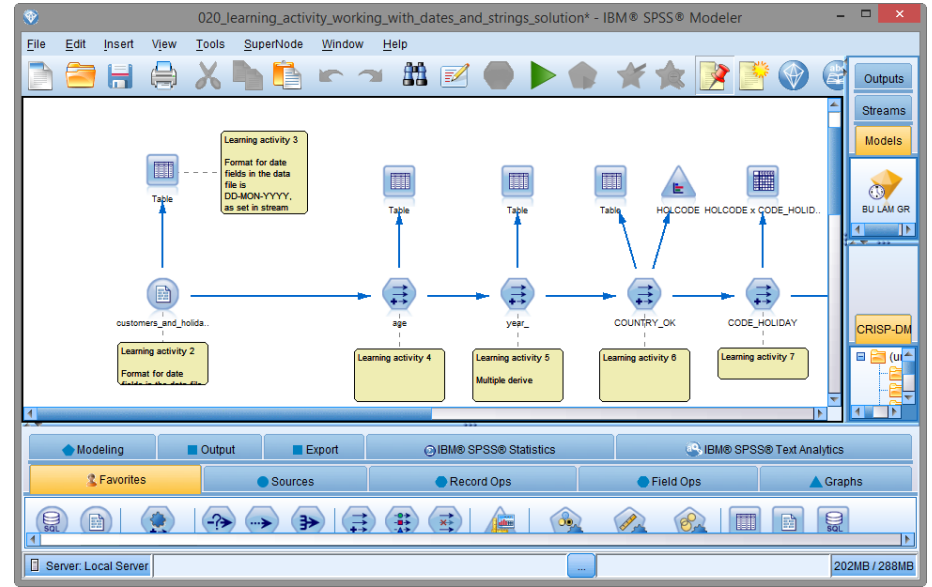
Server side batch engines



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

	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

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

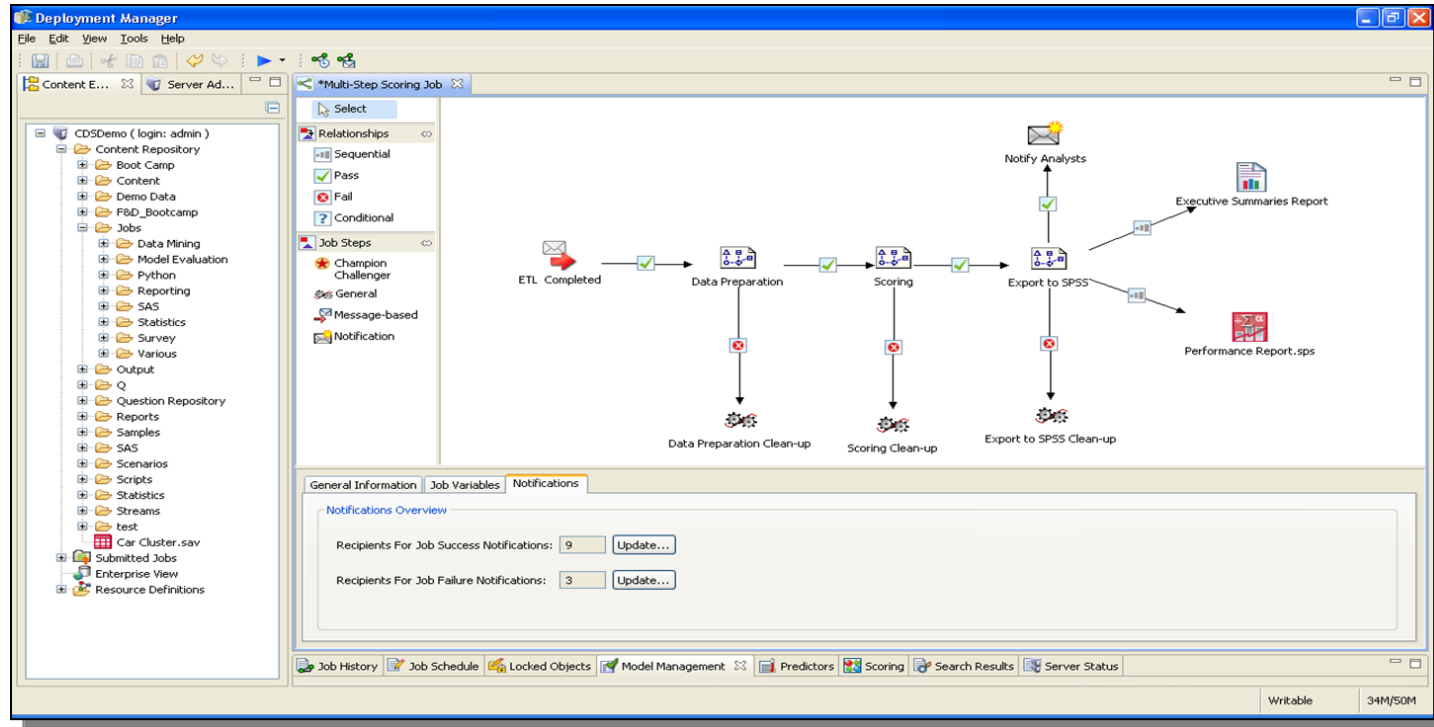
02/12/2014  16:10    <DIR>          .
02/12/2014  16:10    <DIR>          ..
02/12/2014  10:15           78,336 Automating your analyses - the best
03/12/2014  09:23          556,091 Automating Your Analyses v0.1.pptx
03/12/2014  09:03           602 BuildStreamEnd.txt
24/11/2013  17:16           602 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>_
```

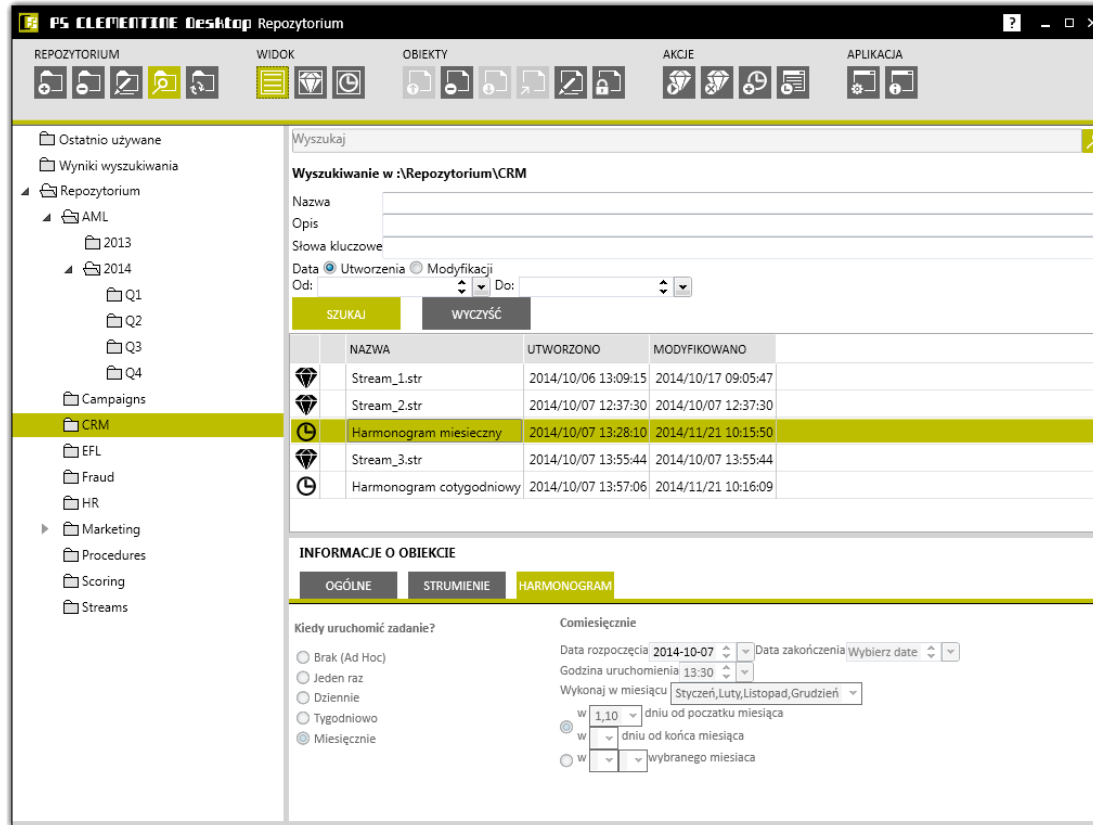
Batch jobs can be scheduled to run using the Windows Task Scheduler



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



PS Clementine Pro – “C&DS-Light” for Modeler Desktop



The screenshot displays the PS CLEMENTINE Desktop interface. The main window is titled "PS CLEMENTINE Desktop Repozytorium". The interface is divided into several sections:

- REPOZYTORIUM:** A sidebar on the left shows a tree view of the repository structure, including folders for "Ostatnio używane", "Wyniki wyszukiwania", "Repozytorium", "AML", "2013", "2014" (with sub-folders Q1, Q2, Q3, Q4), "Campaigns", "CRM" (highlighted), "EFL", "Fraud", "HR", "Marketing", "Procedures", "Scoring", and "Streams".
- WYSZUKAJ:** A search bar at the top right contains the text "Wyszukaj". Below it, the search results are displayed for the path "Wyszukiwanie w :\\Repozytorium\CRM".
- Wyszukiwanie w :\\Repozytorium\CRM:** A table listing search results with columns: NAZWA, UTWORZONO, and MODYFIKOWANO. The table contains the following data:

NAZWA	UTWORZONO	MODYFIKOWANO
Stream_1_str	2014/10/06 13:09:15	2014/10/17 09:05:47
Stream_2_str	2014/10/07 12:37:30	2014/10/07 12:37:30
Harmonogram miesieczny	2014/10/07 13:28:10	2014/11/21 10:15:50
Stream_3_str	2014/10/07 13:55:44	2014/10/07 13:55:44
Harmonogram cotygodniowy	2014/10/07 13:57:06	2014/11/21 10:16:09
- INFORMACJE O OBIEKCIE:** A section below the table with tabs for "OGÓLNE", "STRUMIENIE", and "HARMONOGRAM" (selected). It contains configuration options for the task, including "Kiedy uruchomić zadanie?" (radio buttons for Brak (Ad Hoc), Jeden raz, Dziennie, Tygodniowo, Miesięcznie) and "Comiesięcznie" (radio buttons for monthly frequency and start/end of month).

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 SPSS
 - 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\  
!ENDDFINE.
```

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  
  
!ENDDFINE.
```

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

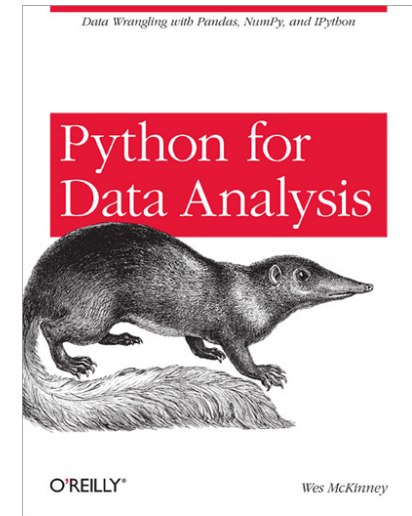
```
'Begin Description
'This file removes upper diagonal of correlation matrix and highlights
'correlations significant at the .01 level.
'End Description
Sub Main
  Dim objPivotTable As PivotTable
  Dim objItem As ISpssItem
  Dim bolFoundOutputDoc As Boolean
  Dim bolPivotSelected As Boolean
  Dim lngIndex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

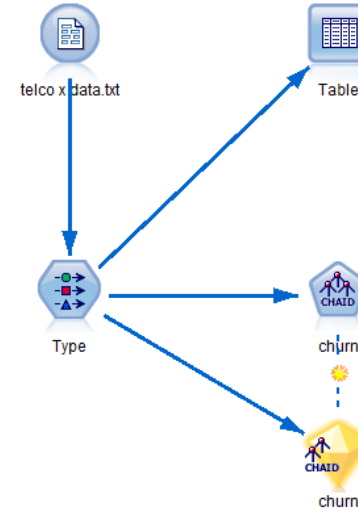
Automating beyond streams - Modeler

- 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

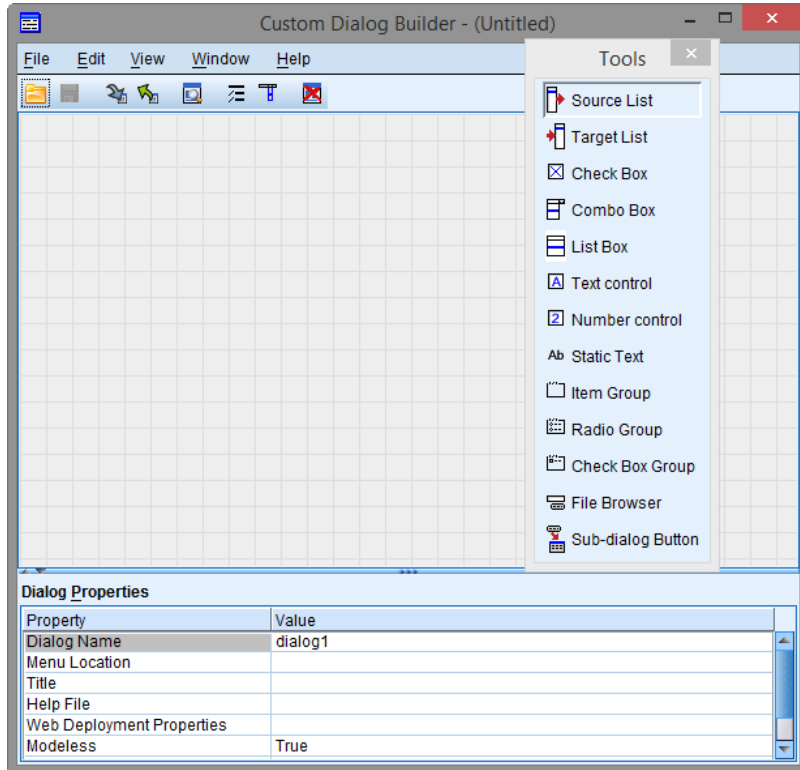
An example python script - looping

Iteration	SP-Handset
1	WC95
2	SOP20
3	SOP10
4	S80
5	S50
6	CAS60
7	CAS30
8	CAS01
9	BS210
10	BS110
11	ASAD90
12	ASAD170

```
1 stream = modeler.script.stream()
2 iterationKey_NodeID = "idiJISCKAXD8V" # "Select":select
3 iterationKey_StreamParameter = u"Handset"
4 iterationKey_NodeField = u"handset"
5 iterationKey_NodeValues = [u'WC95',u'SOP20',u'SOP10',u'S80',u'S50',u'CAS60',u'CAS30',u'CAS01',u'BS
6
7 for i in range(len(iterationKey_NodeValues)):
8     # initialize iteration key
9     value = iterationKey_NodeValues[i]
10    stream.setParameterValue(u"%P-Handset",value)
11
12    # Now execute the stream for this iteration
13    executionResults = []
14    stream.runAll(executionResults)
15 print "Execution completed."
```

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

Extensibility



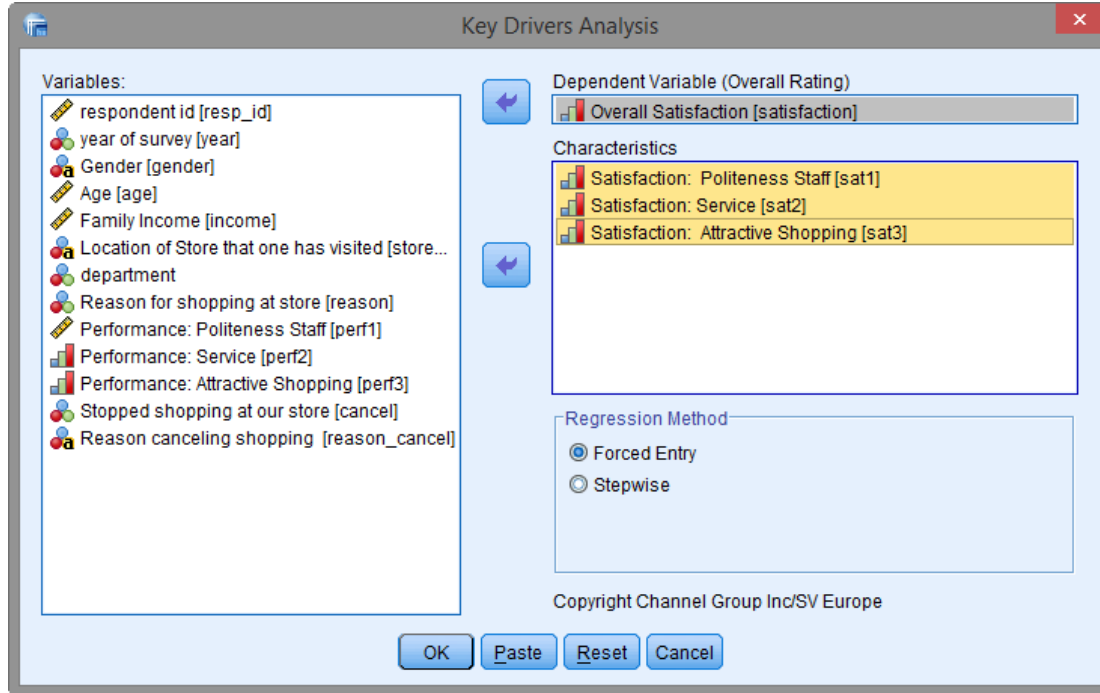
We can use the **Custom Dialog** builder in SPSS (Statistics & Modeler) to:

- Create our own User Interfaces (UIs)
 - and Automate behind them
- With Syntax, Python, R, etc.

From v17.1 of IBM/SPSS Modeler **Custom Dialogs** will support **Apache Spark**

- A number of standard Spark-ML nodes will also be available
- Requires SPSS Analytic Server

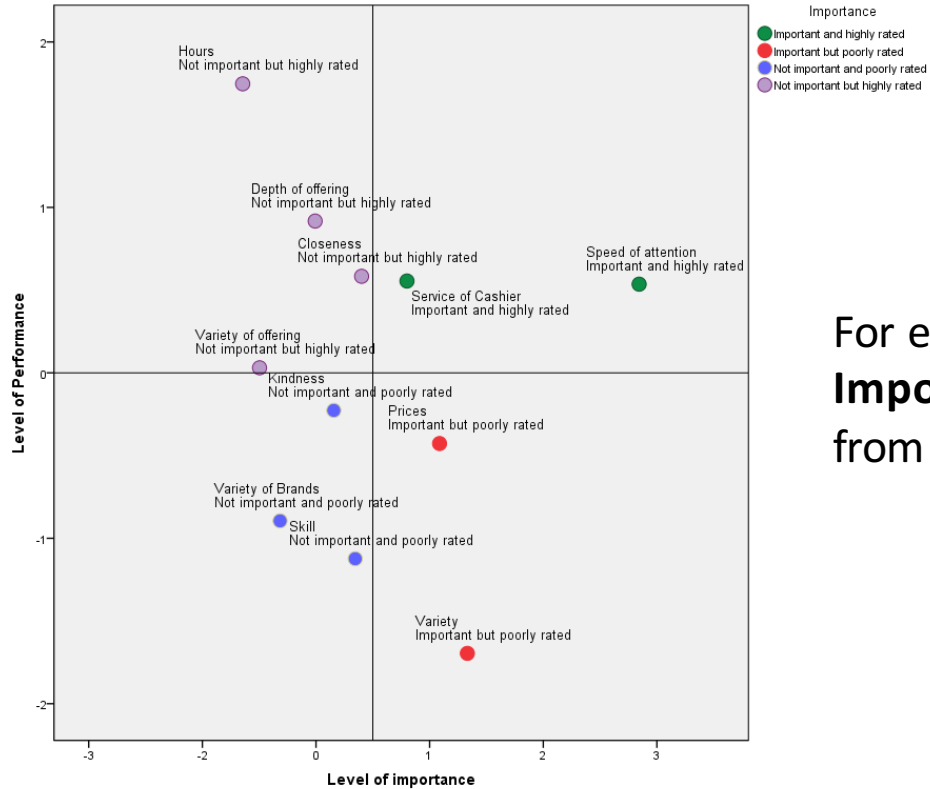
A KDA extension



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

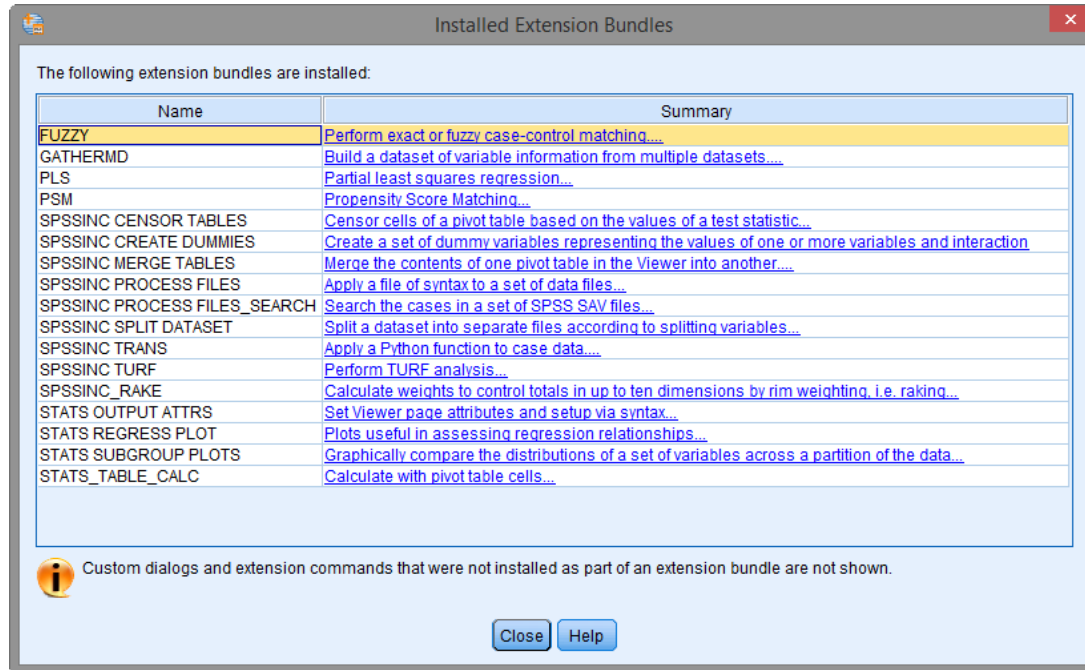
It simplifies (automated) several steps typical in a Key Driver Analysis

A KDA extension



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

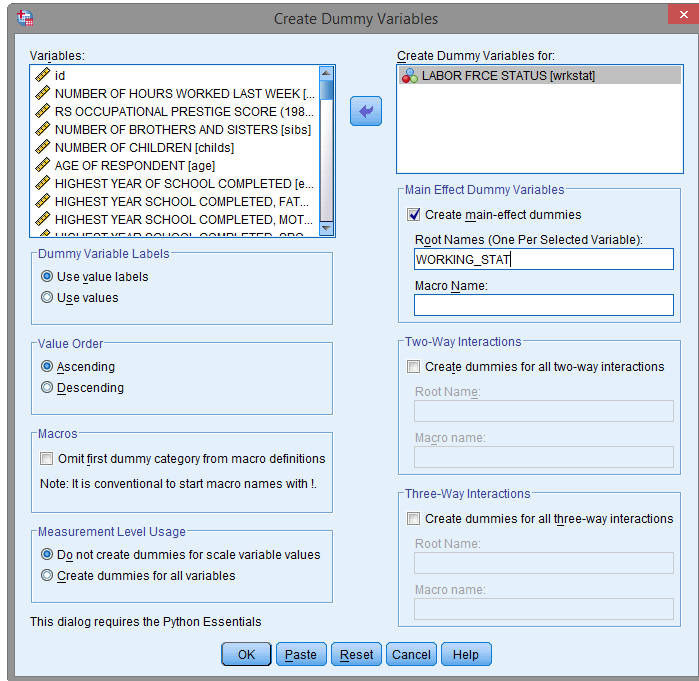
Extension Bundles



Typically written in/using Python and/or R
Check out the SPSS Developer Central for more resources

www.ibm.com/spss/devcentral

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

This is now a standard option in version 23

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/>

A SELECT INTERNATIONAL COMPANY

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)
```

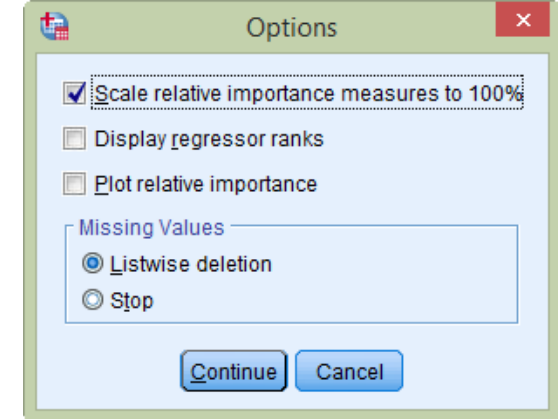
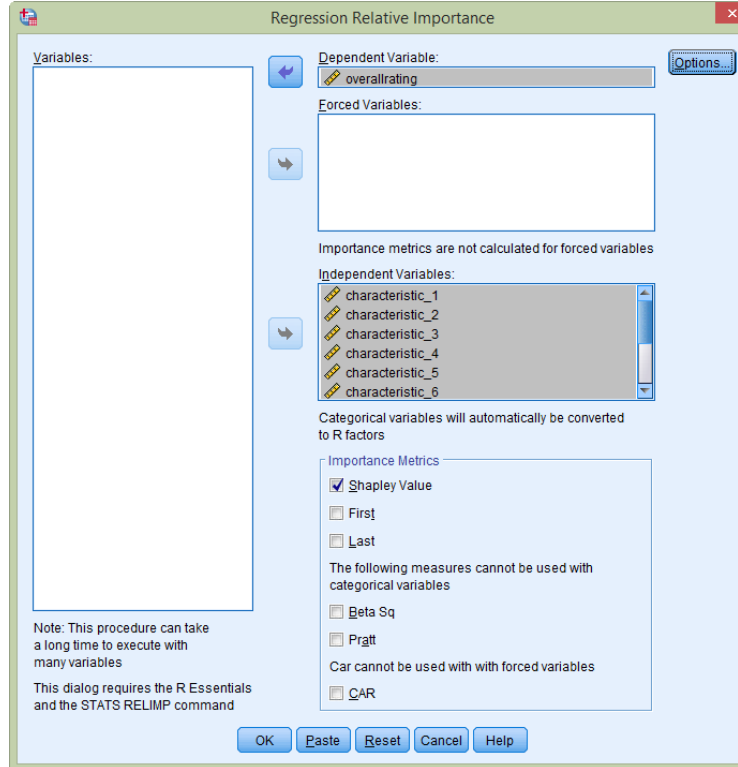
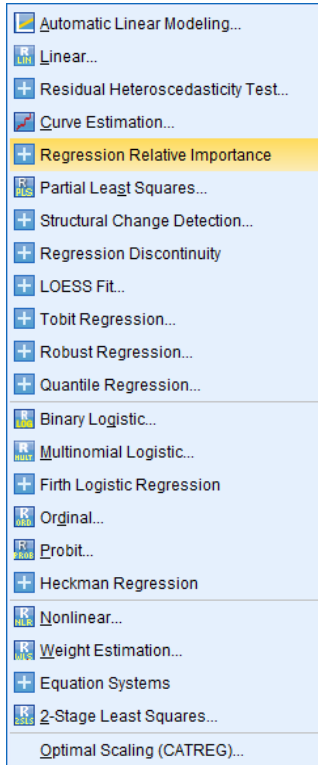
```
summary(reg1)
```

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

END PROGRAM.

Often called
Relative Weight Analysis (RWA)
Or **Shapley Regression**

In V23 Relative Importance is built-in




Predictive Extensions in the IBM Analytics Zone

http://www.ibm.com/web/portal/analytics/analyticszone/downloads/filter?cV=/AnalyticsZone_Content/AZCatalog/Predictive_Extension/

AnalyticsZone

Free Downloads, Trials & More

- FREE Trials & Downloads
- Exclusive Events
- Discussion Forums
- Top Research
- Best Practices
- Expert Blogs & more!



Home
Topics
Downloads
Blogs
Events
Forums




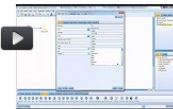
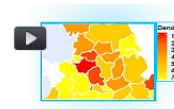

Predictive Extension

1-10 of 10 results

- All
- Trials
- Industry
- Visualizations
- Utilities
- Samples
- Products
- Solutions
- Predictive Extensions

- Category
- Geospatial
 - Model
 - Predictive Analytics
 - Process
 - Visualization

PCI Accelerators

<p>Predictive Extension</p>  <p>Get Coordinates This SPSS Modeler node allows you to get the latitude and the longitude of an address. ...Read more</p>	<p>Predictive Extension</p>  <p>Great Circle Distance Compute the shortest distance between two points on the surface of the Earth. ...Read more</p>	<p>Predictive Extension</p>  <p>Model for Random Forest This is an SPSS® Modeler 'model' node for classification and regression based on a forest of trees using random inputs, utilizing conditional inferenc ...Read more</p>
<p>Predictive Extension</p>  <p>Multicolor Plot This SPSS Modeler 'Graph' node allows you to plot different intervals of a same curve in various colors. Simply install the node, select the coordinat ...Read more</p>	<p>Predictive Extension</p>  <p>Plot Heatmaps This SPSS Modeler 'output' node allows you to plot spatial data on a density heatmap. Simply install the node, select the coordinate's fields and choo ...Read more</p>	<p>Predictive Extension</p>  <p>Plot Spatial Data This SPSS Modeler output node allows you to plot data containing Latitude and Longitude on an interactive map. You can use the same color for all poin ...Read more</p>

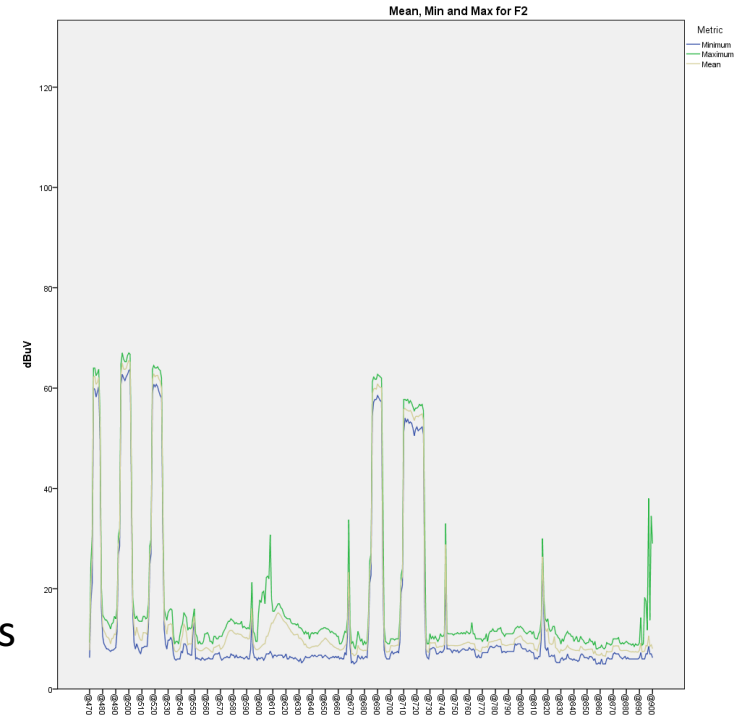
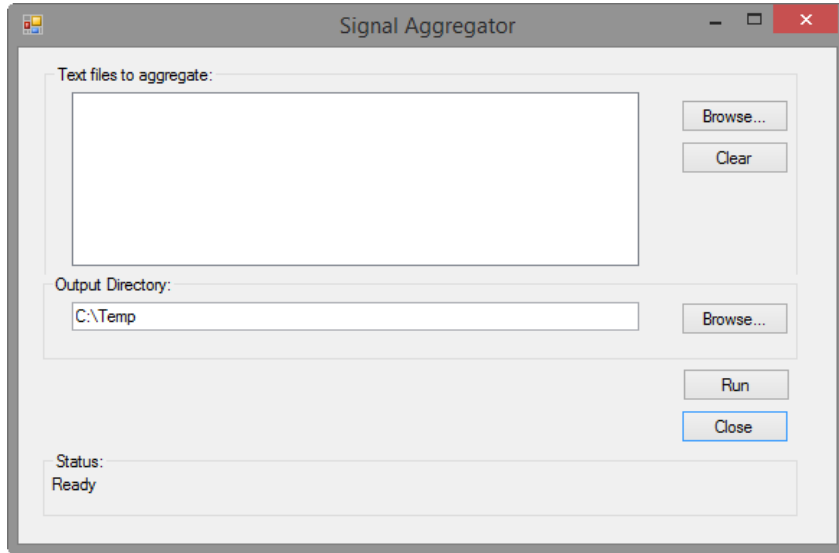
Bit.ly:
<http://ibm.co/1Khmocf>



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 for engineers



Genie

Geographically Enabled Network Intelligence Engine

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

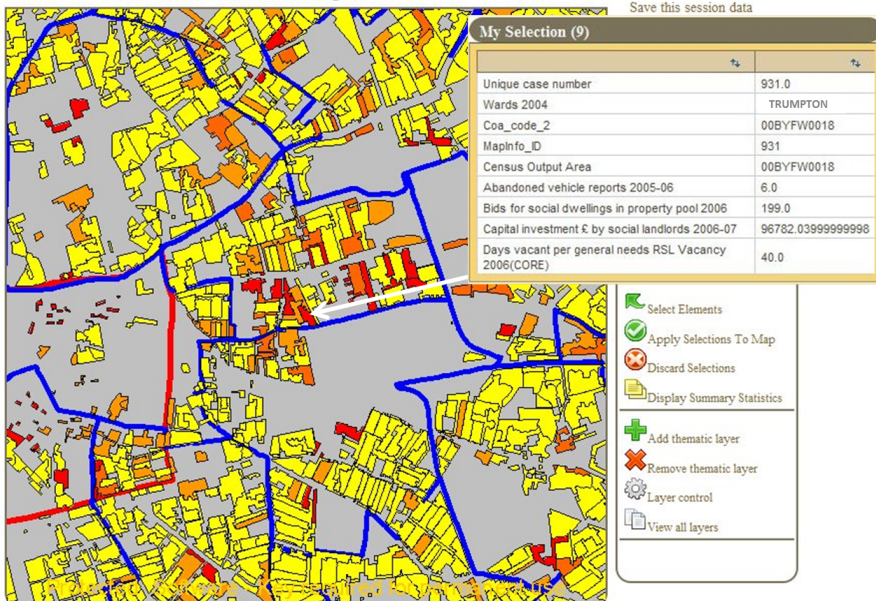
Logged in as: *jmcconnell*

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

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

Session tools

[Save this session data](#)

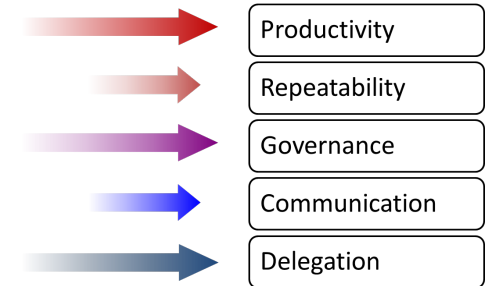


This on-line GIS app is designed for local planners

- 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