

# Using R and Python in SPSS Statistics

## **Curtis Browning**

Architect – IBM SPSS Statistics

## **Jacob Stellan**

Product Manager – IBM SPSS Statistics

## **Jon K Peck**

Independent Contractor

Feb 22, 2022



# IBM SPSS Statistics

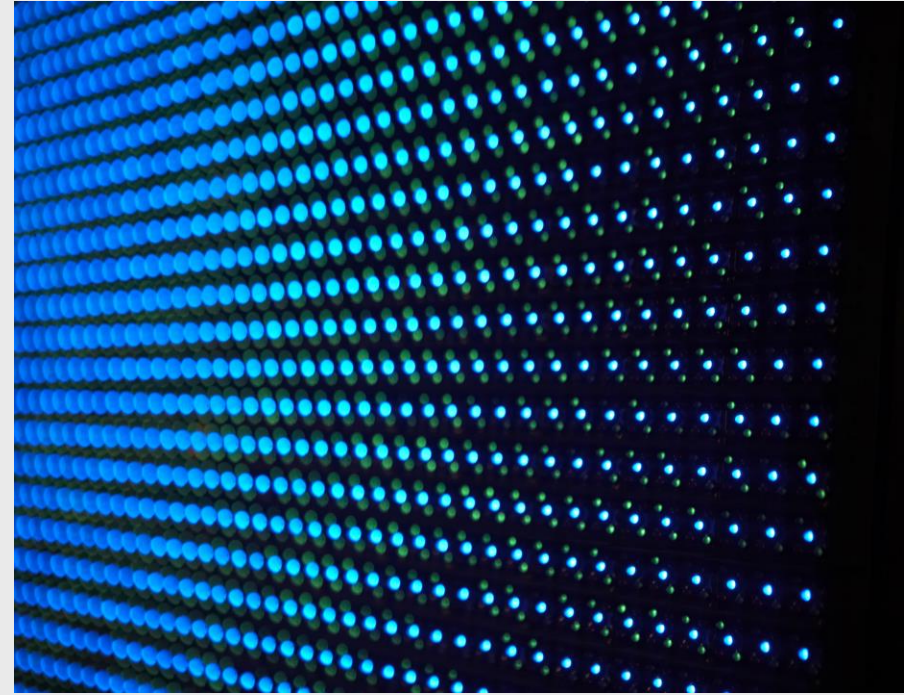
Accelerate research and analysis with the leading  
statistical software solution

# SPSS Statistics and Open Source

Why leverage open-source in SPSS Statistics?

What integrations are available to SPSS Statistics users?

How can a user integrate R/Python with SPSS Statistics?



# SPSS Statistics + Open Source

## A Powerful Combination

### ACADEMICS



ROBERT  
**The Professor**

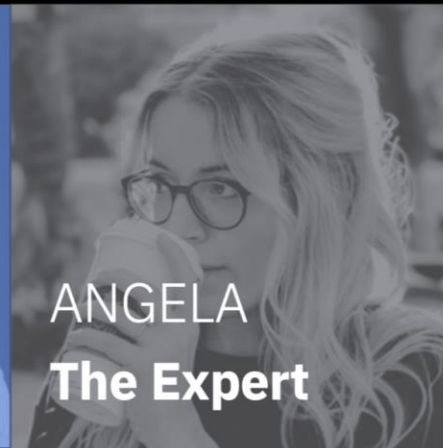
- Sociology research professor of 15+ years
- Extends syntax with R to leverage open-source statistical packages
- Prefers SPSS Statistics to an IDE for reproducibility and built-in high quality outputs
- Creates Custom Dialog Boxes to share with students

### BUSINESS



CARL  
**The Casual User**

- Analyst with no data science background at a small consulting firm
- Utilizes Extension Hub to incorporate functionality from R/Python
- Extensions look and feel like any other statistical package in SPSS Statistics



ANGELA  
**The Expert**

- Experienced quantitative market researcher for a large tech company
- Develops open-source based extensions to share with analysts
- Custom extensions increase efficiency of commonly run procedures

# Programming with SPSS Statistics

IBM SPSS Statistics as a statistical software program contains a set of powerful programming and extension capabilities that users familiar with R and Python can leverage.

Non-technical users also benefit by functionality added by those interested in extending the value of the software. This includes:

- Create and deploy program extensions in R and Python
- Execute R and Python programs from within SPSS Statistics with `BEGIN PROGRAM ... END PROGRAM`
- Driving SPSS Statistics from external R language or Python programs
- Scripting in Python (also in Basic on Windows)
  - Execute scripts internally or externally
  - Associate Autoscripts that execute when new output is generated

# Execute Code from Command Syntax

Users can execute a Python or R language program by embedding code within a BEGIN PROGRAM ... END PROGRAM block in command syntax.

The command syntax can be run from the IBM SPSS Statistics client or from the IBM SPSS Statistics Batch Facility. \*

Python and R code has access to the underlying Statistics engine via exposed Statistics library interfaces.

\* The IBM SPSS Statistics Batch Facility is a separate executable provided with IBM SPSS Statistics Server.

To embed Python 3 code within Statistics syntax, use:

```
BEGIN PROGRAM  
PYTHON3 ... END  
PROGRAM.
```

To embed R language code within Statistics syntax, use:

```
BEGIN PROGRAM R ...  
END PROGRAM.
```

```
BEGIN PROGRAM [{PYTHON3  }].  
                {R        }  
programming language-specific statements  
END PROGRAM.
```

# Driving SPSS Statistics from External Programs

In this mode, you still have access to all of the functions in the Python and R Integration Packages for IBM SPSS Statistics, but you can develop and test your programs with the development environment of your choice.

SPSS Statistics ships with automatically installed versions of Python 3 and the R language. These are part of a virtual environment that each user on a given machine has exclusive access to.

- Python programs make use of the interfaces exposed by the Python **spss** and the **SpssClient** modules. Start by a call to the **spss.StartSPSS** function.
- Users can utilize the built-in **IDLE** Python editor or use their own preferred editor.
- R programs include the **spsstatistics** library, followed by a call to the **spsspkg.StartStatistics** function.

# Running SPSS Statistics from Python

Users can drive the SPSS Statistics program from their preferred Python development environment.

The image at right shows the Wing Python development environment being used to drive Statistics via the Statistics Python libraries.

The screenshot displays the Wing Python IDE interface. The top pane shows a Python script with line numbers 1039 to 1051. The script defines a `CasingDict` class and a `VariableDict` class. The bottom pane shows the Python Shell with the following commands and output:

```
Python 3.9.1 (tags/v3.9.1:1e5d33e, Dec 7 2020, 17:08:21) [MSC v.1927 64-bit]
Type "help", "copyright", "credits" or "license()" for more
1>>> import spss, spssaux
2>>> spssaux.OpenDataFile("c:/spss28/samples/english/employee data.sav")
3>>> spss.Submit()
```

The output shows the execution of the `Submit` command, which triggers a series of SPSS commands: `StartDataStep`, `StartProcedure`, `StartSPSS`, `StopSPSS`, `StrError`, and `Submit`. The `Submit` command is highlighted in the output list.

# Running SPSS Statistics from R

Users can run an R language development environment to drive SPSS Statistics from their favorite R language IDE.

The image at right shows SPSS Statistics being driven by the RStudio R language development environment.

```
R version 4.0.5 (2021-03-31) -- "Shake and Throw"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

* Project 'C:/Users/jkpec/AppData/Roaming/IBM/SPSS Statistics/28/R40' loaded. [renv
0.13.2]
> library(spssstatistics)
> spsspkg.startStatistics()
> spsspkg.version()
$platform
[1] "x86_64-w64-mingw32"

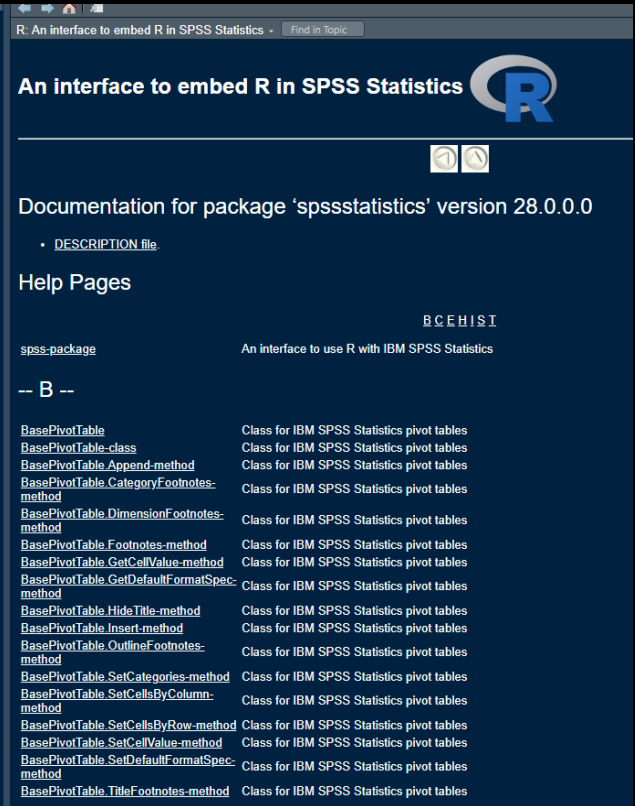
$search
[1] "x86_64"

$os
[1] "mingw32"

$system
[1] "x86_64, mingw32"

$version.string
version
"28.0.0"

> |
```





# The Scripting Facility in SPSS Statistics

The scripting facility allows you to automate tasks, including:

- Opening and saving data files.
- Exporting charts as graphic files in a variety of formats.
- Customizing output in the Viewer.
- Creating and executing autoscripts
- Startup scripts that execute when SPSS Statistics starts

## Default Script Language

- Determines the script editor that is launched when new scripts are created .
- It also specifies the default language whose executable will be used to run autoscripts.
- On Windows, the default script language is Basic.
- Change the default language from the Scripts tab in the Options dialog.

# Scripting in Python (and in BASIC on Windows)

## **Run from SPSS Statistics.**

- Run from Utilities>Run Script or from the Python script editor which is launched when opening a Python file (.py) from File>Open>Script.
- Operate on the SPSS Statistics client that launched the editor
- Use the SCRIPT syntax command to execute scripts from within or outside SPSS Statistics

## **Run from an External Python Process.**

- A Python IDE that is not launched from SPSS Statistics, or the Python interpreter.
- Attempts to start or connect to an existing SPSS Statistics client.
- By default, the newly started Data Editor and Viewer are invisible for the new client. You can choose to make them visible or work in invisible mode with datasets and output documents.

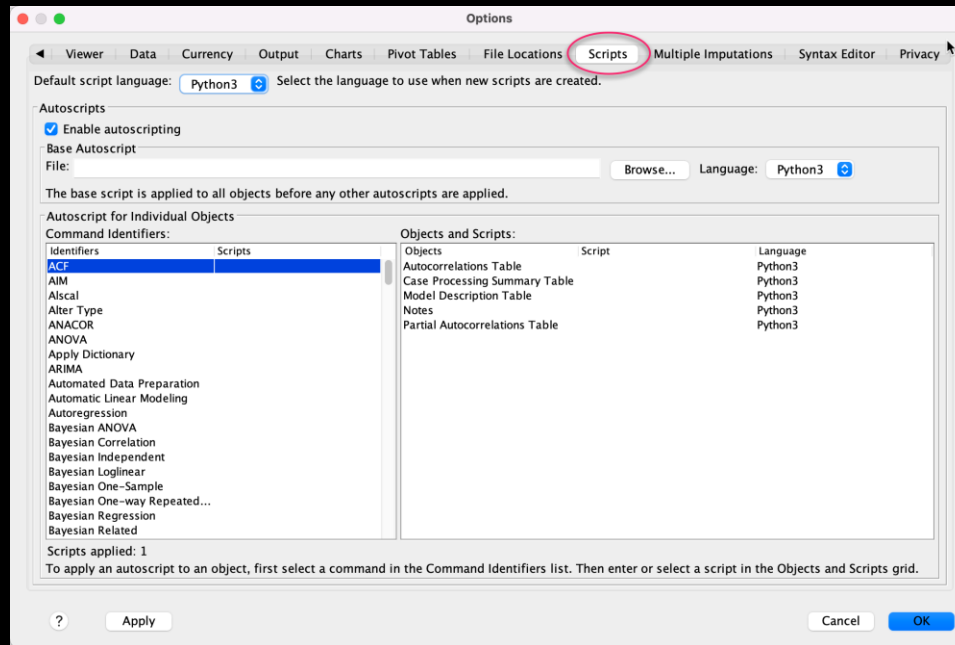
# Autoscripts

Autoscripts are scripts that run automatically when triggered by the creation of specific types of output from selected procedures.

- Events: New tables, notes, or warnings
- Associated from Output or from Options

## Autoscripts in SPSS Statistics Options

- Select the default autoscript language
- Manage script associations
- Create a Base Autoscript that applies to all new output



# R and Python Extensions

Among the most powerful features in SPSS Statistics is the ability to add functionality to the program via Extensions.

SPSS Statistics ships with a variety of extensions that are provided internally.

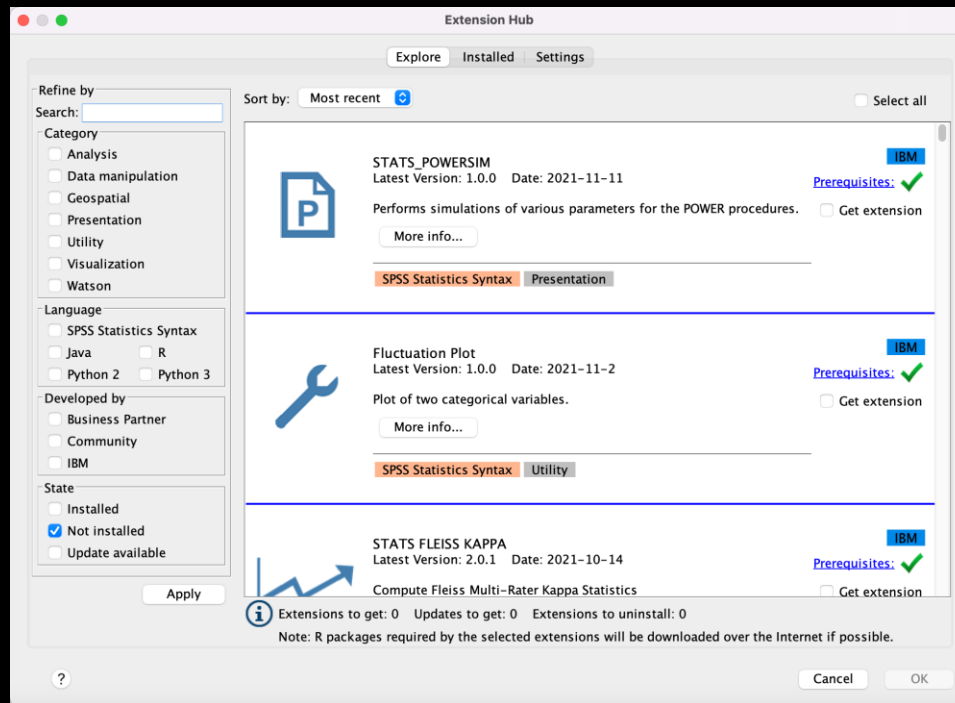
With SPSS Statistics Extensions one can:

- Customize SPSS Statistics by adding R and Python extensions that are driven by syntax
- Access datasets from the central system
- Incorporate open-source algorithms
- Generate standardized output
- Modify existing output
- Design and deploy custom user interfaces that emit syntax
- Share and deploy extensions for other SPSS Statistics users

# R and Python Extensions (continued)

Users install and uninstall extensions via the **Extension Hub**:

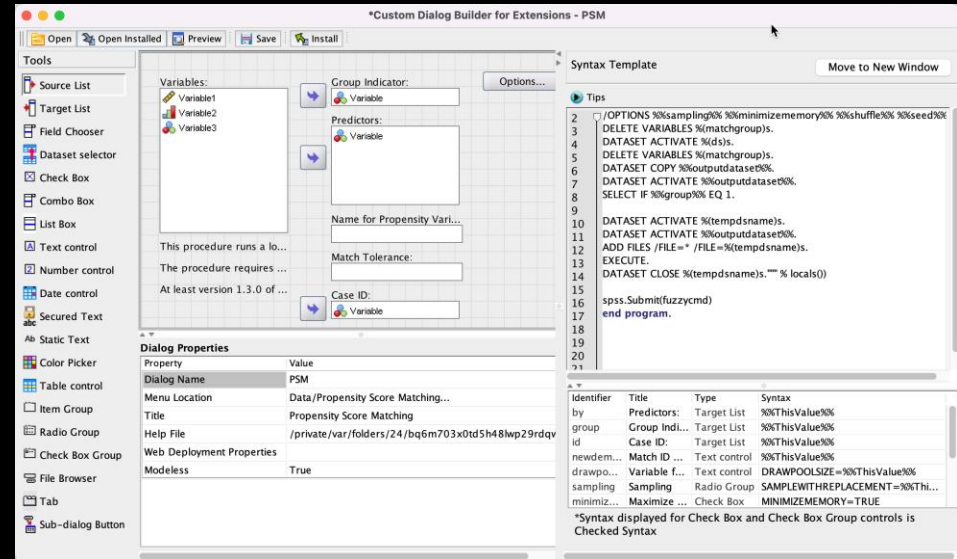
- An interface for searching for, downloading, and installing public extensions
- Accessed from **Extensions > Extension Hub**
- Local extensions can be installed from **Extensions > Install Local Extension Bundle**



# R and Python Extensions (continued)

Users create specialized user interfaces for extensions via the Custom Dialog Builder:

- Drag and drop controls onto the production canvas
- Associate controls with the properties that drive an extension
- Generate the syntax for driving extensions
- Live preview and syntax generation
- Deploy or save directly
- No GUI programming required!



# **SPSS Statistics Demo**

# Resources

Watch all Tech talks: <https://ibm.biz/spsstechtalk>

GradPack (Students / Faculty): [Academic discounts](#)

Take the no cost [30-day trial](#)

Interactive Product Tour: [Start your tour here](#)

Join the SPSS Statistics Community: [Get connected](#)

Use cases: [How industry experts use SPSS Statistics](#)

Blog: [Learning resources for SPSS Statistics](#)

Extensions: [Writing Extension Commands](#)

Solution brief: [SPSS Statistics and Open Source - a powerful combination](#)



# Next steps

Purchase SPSS Statistics



**[ibm.com/spssstatistics](https://ibm.com/spssstatistics)**

**Try the Free Edition**

**[ibm.biz/SPSSStats](https://ibm.biz/SPSSStats)**



# Trademarks and notes

© IBM Corporation 2022

IBM, the IBM logo, ibm.com, and SPSS are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “[Copyright and trademark information](http://www.ibm.com/legal/copytrade.shtml)” at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Other company, product, and service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.