Hello, my name is Armand Ruiz, and I'm going to show how to develop new extensions for IBM SPSS Modeler 16 using R. I am Armand Ruiz, an IT specialist on the data analytics Ecosystem Development team in IBM in France.

So, I'm going to show basically how to develop two different extensions for SPSS. The first one is to create a geocoding node. Geocoding is the process to find the geographic coordinates, latitude and longitude, from geographical data such as street address, zip code, et cetera.

For this we are going to use public APIs. In this case, we are going to use a Google Maps API. The advantage of doing this is that it's very fast and very easy to use inside the same workbench of SPSS.

So, first of all, we are going to develop this first node on SPSS. So, I'm not going to explain here what's the user interface or the value of SPSS, I'm going to just show how to develop these new extensions. So, this is the SPSS Modeler 16. We have a new tool called Custom Dialog Builder. So this is like the workbench to create our new extensions for SPSS Modeler.

If I open it, I have here on the left all the tools I can drag and drop into the model option. So, in this case, I'm going to select the field chooser. And I have here the properties down. So I am going to change the identifier and this is going to be the field chooser for the address.

So, I call it address and the type that I'm going to put, address field. Okay? So, then if I do edit and Script template, here it has a white space to put my R script, because the extension will be based on R code. So, I have the code here already developed. It is the same code you will find the article of developerWorks, so I just do copy/paste of my code.

In this case, what I am doing, is I am just getting the Modeler data for location and I'm creating the URL to call the Google Maps API and going to receive the result in JSON. So then the only thing I am doing is just parsing the JSON to get the latitude and longitude. And I'm creating two new fields in the data set. One is called latitude and the other one is called longitude. And then I am adding this for my data frame.

Okay? So this like this, you see I just want to put a name to this new node, it's going to be called geocode and the title will be geocoding. And I'm going to place, I'm going to say that this is a process node and I'm going to place this new node in the field operations.

And I'm going to put an icon, you have the image of the icons as well in the files of the developerWorks article. So, I just here geocode icon, so now I'm done, and I click install. And that's it. So now, if I go into field operations, I see I have my geocoding node.

So, let's try how it works. For doing this I'm going to generate some data. I go to sources, I drag and drop an User Input node. I double click in it and I'm going to create some data. In this case, the field is going to be called location, it's going to be a string, and let's create some data.

Barcelona, Spain. Madrid, Spain, I put Paris, France, Moscow Russia and also San Francisco, California. In this case, I am creating manually the data because I am testing. So, if I do preview, I can see the data generated.

So, let's geocode this and get the latitude and the longitude of these cities. I
connect the source node with geocoding. I double click the geocoding and I select the location.

Let's plot all the data in a table. So, I'll connect a table from the output palette to the geocoding node. Obviously, this node needs internet connection to work. So, that's why it's taking some time to do the request.

So, here we are. The geocoding looks like it works well. I have each of the cities, the latitude and the longitude. So, this is cool, this is very good to have it like this but what about creating another node; in this case, a Google Maps node to plot this geographical data in a Google Maps map in the Web browser.

So, here basically we're going to create an interactive map to handle this geographic data and it's going to be shown in a Web browser and we will be able to check the results of our analysis. So, we don't need to export the data to any GIS visualization tool.

So, in the same way we did before, we go again to the Custom Dialog Builder for R and here we will drag and drop two new fields: one for the latitude and another one for longitude. Say, the first one I will call...we put as identifier lat and the title latitude field. In the second one, I will call as identifier long and as a title longitude field.

In the same way as before, I do edit script template and I have here the code to do my Google Maps using R. So, I will call this node Google Maps. And as title, Google Maps.

This node is going to be obviously output node and I will place it in the output palette. Again, I have an icon, so I select the icon that you will find as well in the files of the article. So, we are done. We install the new node and here we are in the output we see we have now our Google Maps. I drag and drop, I connect it to the geocode, I double click and in the latitude fields I select the latitude and the longitude field I select the longitude, I run this, and what I'm doing here is I am geocoding the input data and I plotting it to a map.

So, like this I can really verify that the geocoding is working properly. So here it is my map, I see I have San Francisco; Madrid, Spain; Paris. And Moscow info box, I can modify this very easily.

So, let's try to do now a bit more complete scenario. I'm going to put as data source a CSV with Crime Data of City of Chicago. I have this data set in my desktop, so I am going to read this data set, it's a CSV as you can see, and if I look at review, I see I have some data about Crime and at the end, I have the latitude and the longitude.

And let's plot all these data into our Google Maps to see how it looks. So, I connect this Crime data to Google Maps, I do...select the latitude field again, the longitude field, I run this, and we've got super fast, all the crime data of the City of Chicago.

So, since it's a Web browser, I can zoom in, zoom out, and I can click each of the points and see all the information of each crime.

So, as you can see, developing new nodes for SPSS Modeler 16 using R is very easy and very useful. Hope you liked it. Thanks.

[ MUSIC ] [END OF SEGMENT]