rCharts Basic Charts

Use rCharts R Package to Create Interactive Visualizations

Product: IBM® SPSS® Modeler

Extension type: Visualization

Table of Contents

Description:

This node allows you to create interactive JavaScript based visualizations as a Modeler Output. The rCharts package for R offers R users easy access to generating data visualizations that are ready to be published on the web. This extension makes it even easier to create data visualizations by allowing direct output from SPSS Modeler to a browser based chart. This extension allows the creation of bar, line, or scatter plots. It allows the specification of an x-axis variable, y-axis variable and a variable that the chart should be grouped or colored. The height and width of the plot can be modified as well as the inclusion of necessary JavaScript libraries stored locally or through a CDN.

Requirements:

- SPSS Modeler v16.0 or later
- R: http://www.r-project.org/
- 'R Essentials for SPSS Modeler' plugin: https://developer.ibm.com/predictiveanalytics/downloads/predictive-extensions

Installation:

Close SPSS Modeler. Save the .cfe file in the CDB folder of the IBM SPSS Modeler installation directory for Windows and Linux. The copy should reside in that same folder and not in a sub-folder.

For example, for Windows 7 and Modeler 17 the default location is "C:\ProgramData\IBM\SPSS\Modeler\17\CDB". If the ProgramData folder is hidden type the path manually.

Restart SPSS Modeler: the node will now appear in the Output palette.

R Packages used:

- 'devtools' package created by Hadley Wickham [aut, cre], Winston Chang [aut], RStudio [cph], R Core team [ctb] (Some namespace and vignette code extracted from base R) https://cran.r-project.org/web/packages/devtools/

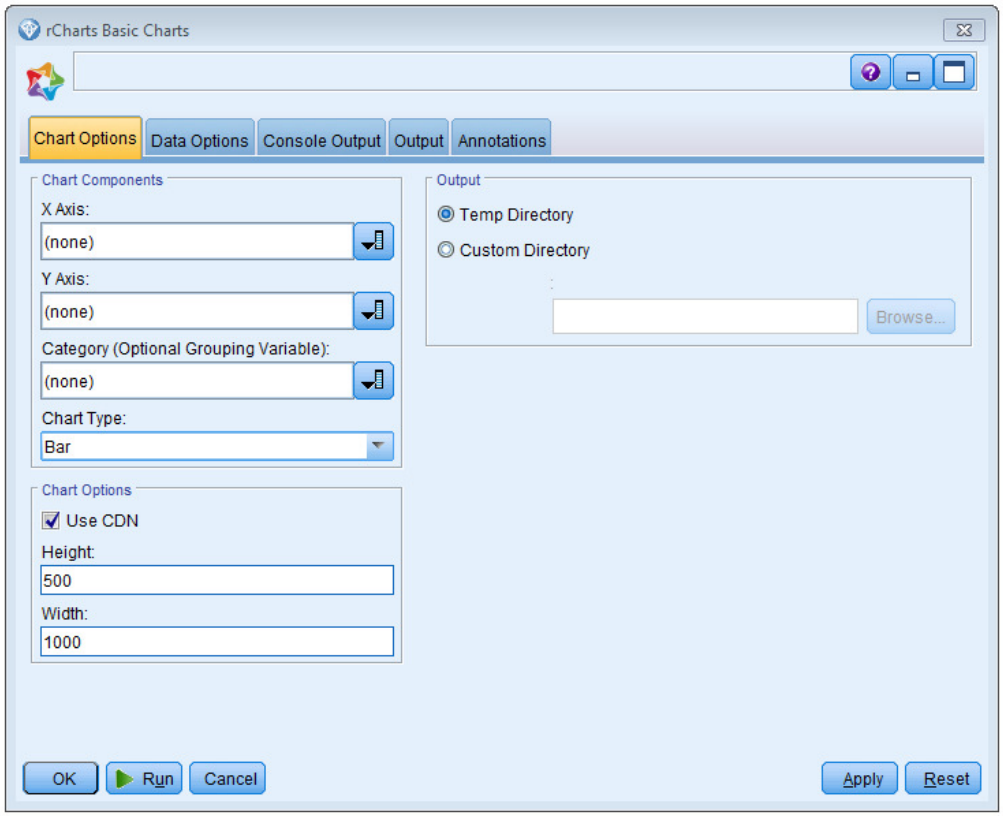
- 'RCurl' package created by Duncan Temple Lang and the CRAN team https://cran.r-project.org/web/packages/RCurl/

- 'rCharts' package by Ramnath Vaidyanathan http://ramnathv.github.io/rCharts/

User Interface

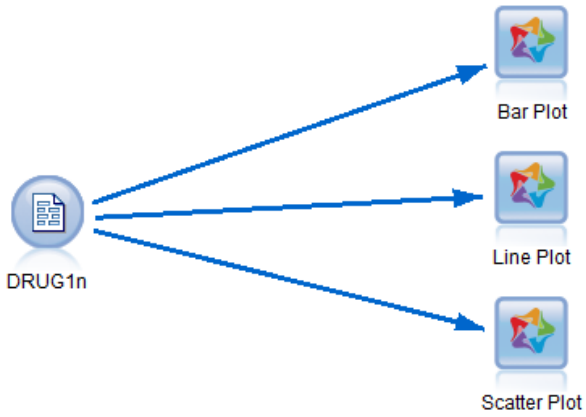
This extension only has one main dialog tab for manipulating the plot output. The dialog, shown below, has the following options.

- Chart Components
  - X Axis – this selects a field from the input dataset that will be plotted along the x-axis
  - Y Axis - this selects a field from the input dataset that will be plotted along the y-axis
  - Category (Optional Grouping Variable) – If selected then the plot will be grouped by this variable, using color. This variable is typically categorical and often is the class variable in a classification task
  - Chart Type – The current options for rCharts Basic Charts are Line, Scatter, and Bar.
- Chart Options
  - Use CDN – If this box is checked the HTML output from this extension will contain a link to a CDN (content delivery network) which hosts the necessary javascript libraries, otherwise this will be linked to a locally hosted version.
  - Height – Height in pixels of the plot
  - Width – Width in pixels of the plot
- Output
  - Temp/Custom Directory – This extension creates an HTML when creating output. If you want to only view or take the source code from the document generated a Temporary file is best, but if you want to the HTML, select a directory to save.



Example:

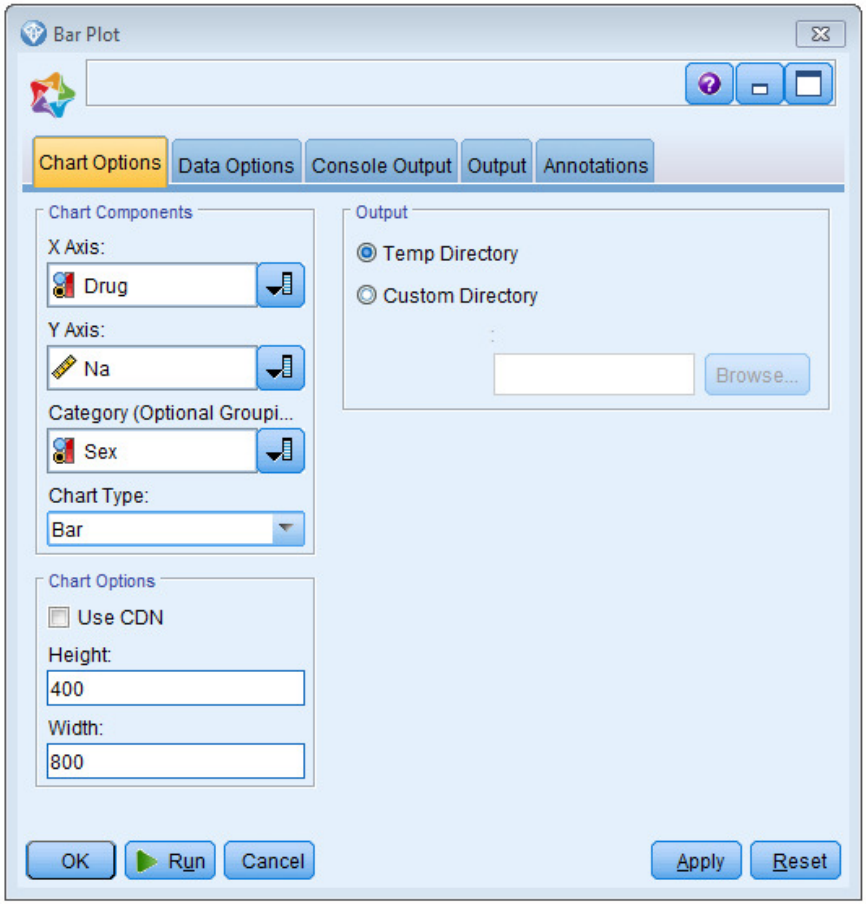
For an example use of this extension, we will use the 'DRUG1n' dataset that is available with Modeler. To find this dataset, navigate to where you installed Modeler then go to the Demos subdirectory (for my set up it is 'C:\Program Files\IBM\SPSS\Modeler\18.0\Demos'). This is a nice toy dataset that will let us get started making visualizations. You can add this to your stream by adding a Var. File node and opening this file. With this file open, we are ready to start plotting.



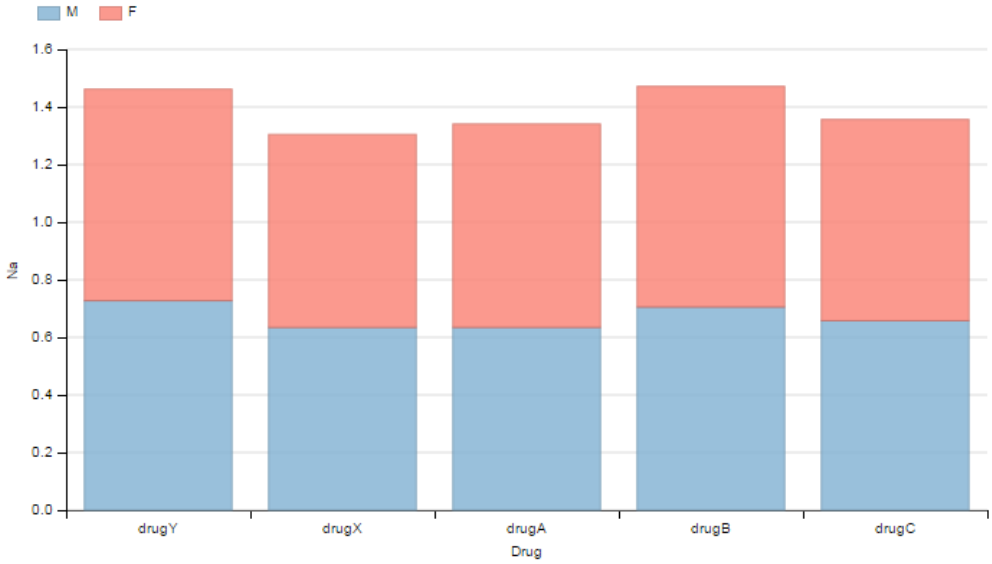
1. **Bar Plot**

   Let's start with the classic Bar Plot. First, you need to add the rCharts extension to the canvas and connect it to the DRUG1n data. Your stream will begin looking like the screenshot above. Generally for Bar Plots, we want to put a categorical variable along the x-axis for our bars, and a numeric variable on the y-axis to show the variations of that variable for each value of the category on the x-axis. We can further expand the visualization to include a grouping variable that adds another dimension for us to examine.

   Open the rCharts extension and use the following set up:



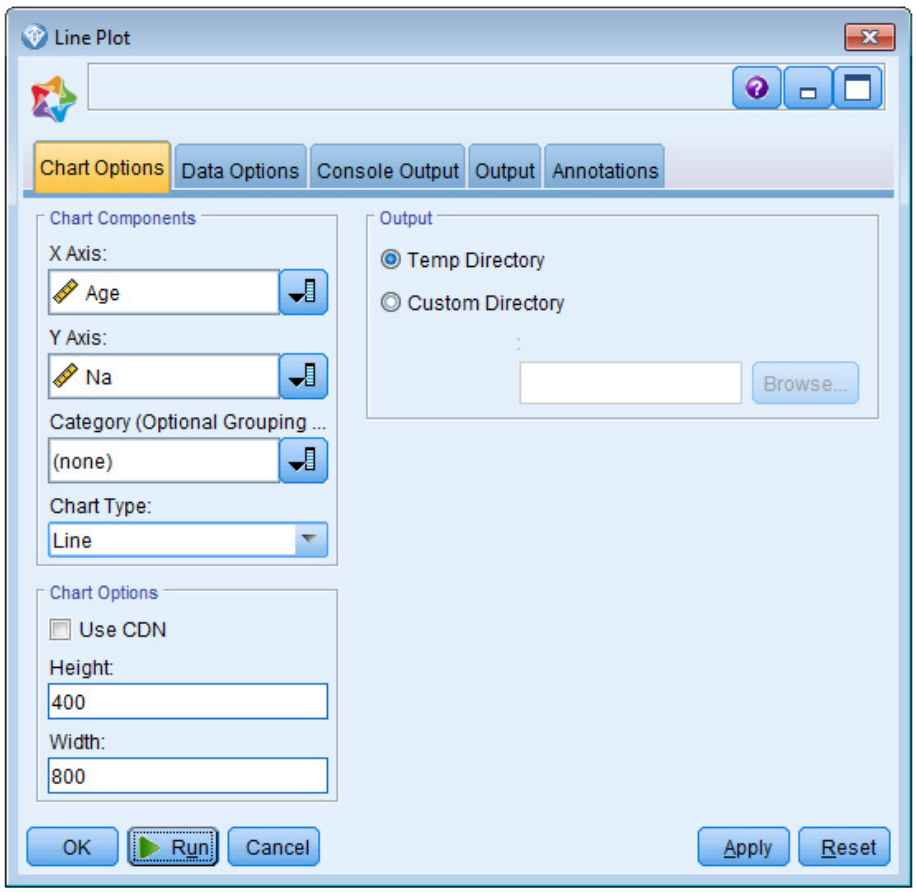
When you click “Run”, a web browser should launch and the visualization below should appear.



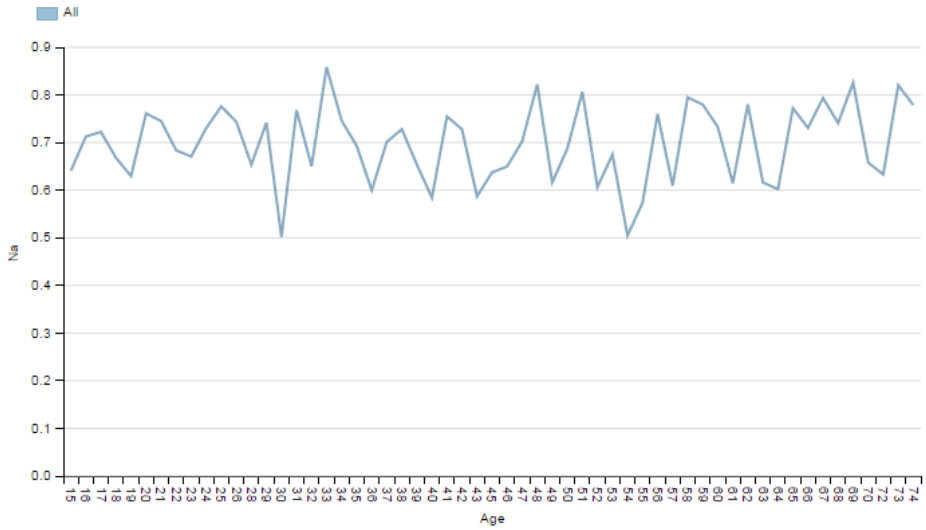
It is that easy to create an interactive web-based bar plot with this extension. If you are inclined to do more with these visualizations, take a look at the step by step example in the example directory to see how to add these visualizations into a dashboard format.

2. Line Plot

Add an rCharts node to the canvas and connect to the data set as in the previous example. Now adjust the dialog to match the screen shot below. We will leave the Category variable set as none for this example. If we chose a variable for this then we would see a line for each value for that variable. For a Line plot, it makes most sense to use numeric variables for the X and Y axis and a categorical variable for the Category.

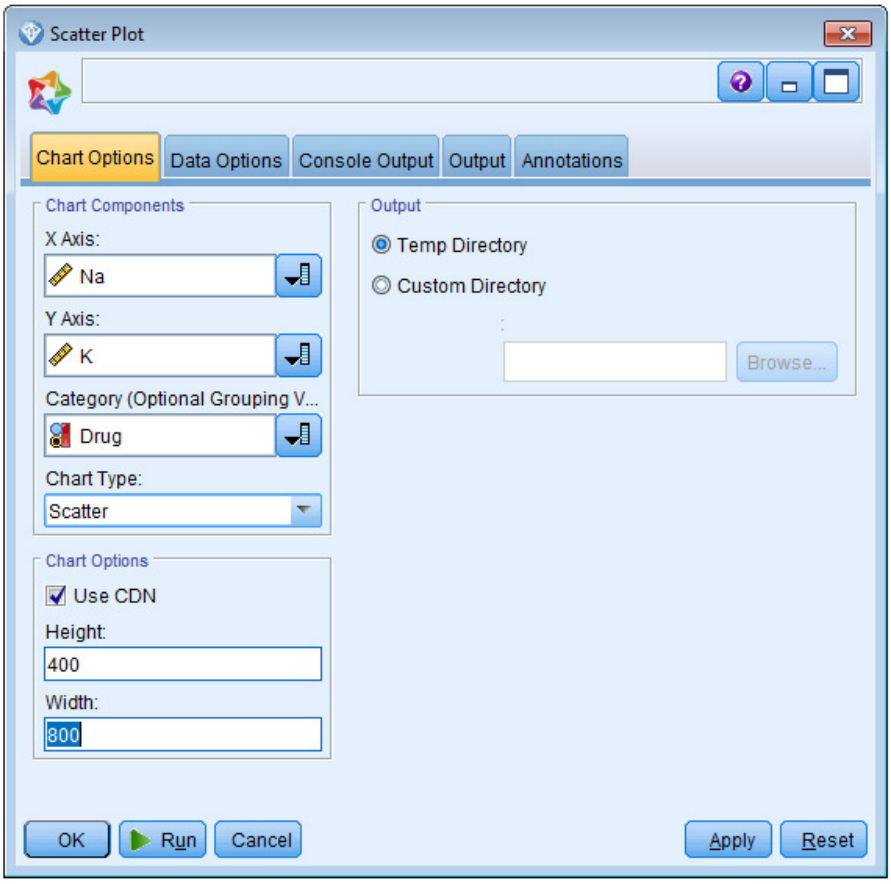


The output from this set up should look like the plot below.

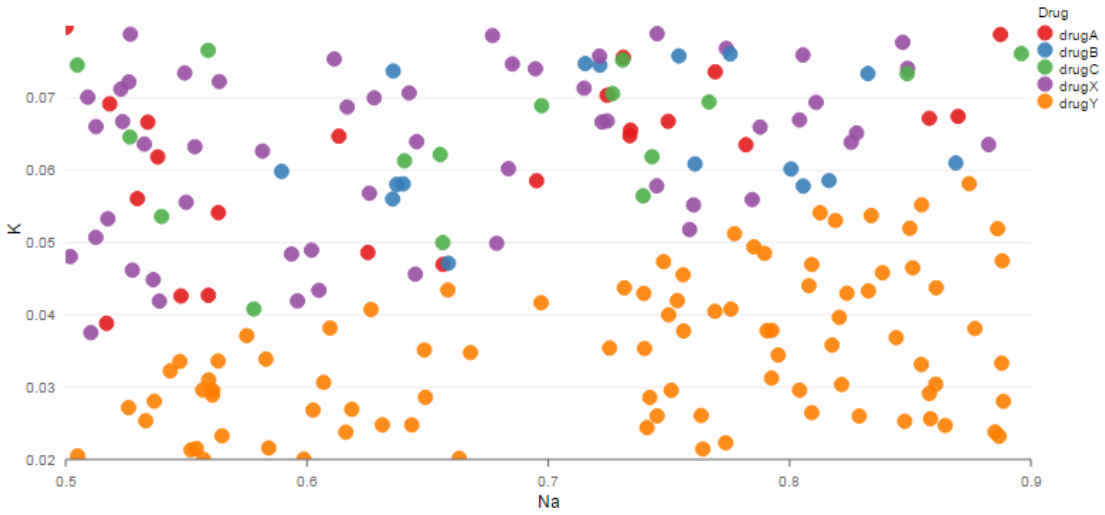


3. Scatter Plot

Add a final rCharts node to the canvas and connect to the data set as in the previous example. Now adjust the dialog to match the screen shot below. For a Scatter plot, it makes most sense to use numeric variables for the X and Y axis and a categorical variable for the Category. Let's set X to be the Na value, Y to be K, and the Category to be the Drug variable.



The output from this set up should look like the plot below.



It's that easy to create plots in rCharts with SPSS Modeler!

Important Links

Learn

- Take a look at the rCharts gallery to see everything that's possible - http://rcharts.io/gallery/
- Learn more about SPSS software.
- Visit developerWorks Business analytics for more technical analytics resources for developers.
- The Comprehensive R Archive Network is the main site for the R project and each R package. The help pages and manuals that are associated with optimx, nlmrt, and Rcgmin are detailed. Numerous references are provided.
- Read "Do I need to learn R?" (Catherine Dalzell, developerWorks, September 2013) to learn why R is a valuable tool for data analytics that was expressly designed to reflect the way that statisticians think and work.
- "Calling R from SPSS" describes how to use R code inside IBM SPSS Modeler 16.
- Read "Create new nodes for IBM SPSS Modeler 16 using R" to learn how to create new extensions easily.

Discuss

- Visit the IBM SPSS Community to share tips and experiences with other IBM SPSS developers.
- Follow developerWorks on Twitter to be among the first to hear about new resources.