Using R and Tableau

Table of Contents

What is R?

R is a popular statistical language used to perform sophisticated statistical analysis and predictive analytics, such as linear and nonlinear modeling, statistical tests, time-series analysis, classification, clustering, etc.

The R-console primarily utilizes a command-line interface, but there are many GUI tools available for download to make it easier to write R programs (most of them are free). With R, users can create variables, formulas, functions, and graphs to visualize their analysis and predictions.

As a free, open-source language, there is a community of contributors who continually create new packages (extensions) for R that define advanced statistical functions that were not originally built in to R. These packages can be downloaded into R to expand R's capabilities. Most of these packages are also developed and made available for free. It is for this reason that R has become so popular and why it continues to gain functionality over time.

What are the benefits of using R?

R provides a powerful way to do statistical analysis on large sets of data. It is also free, which is a compelling factor to its growth. Because it is open source, new functions and packages are created all the time, so if you can't find a capability initially, you can search for a package that can do it or even create a package of your own.

However, there are some limitations to R as well. In order to be as flexible as it is, R leverages a command-line interface and uses its own programming language and syntax. It does require some aptitude in coding to take advantage of the many functions. Other non-free, proprietary statistics packages often have graphical user interfaces that are much more friendly and do not require programming skills.[1] These solutions are intended for those users who are not as apt to learn the R programming language, or do not need the sophisticated capabilities that are possible with R.

▶ *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.*

[1] For example, Rattle is a purpose-built GUI for data mining.

How does Tableau integrate with R?

R functions and models can now be used in Tableau by creating new calculated fields that dynamically invoke the R engine and pass values to R. The results are then returned back to Tableau for use by the Tableau visualization engine.
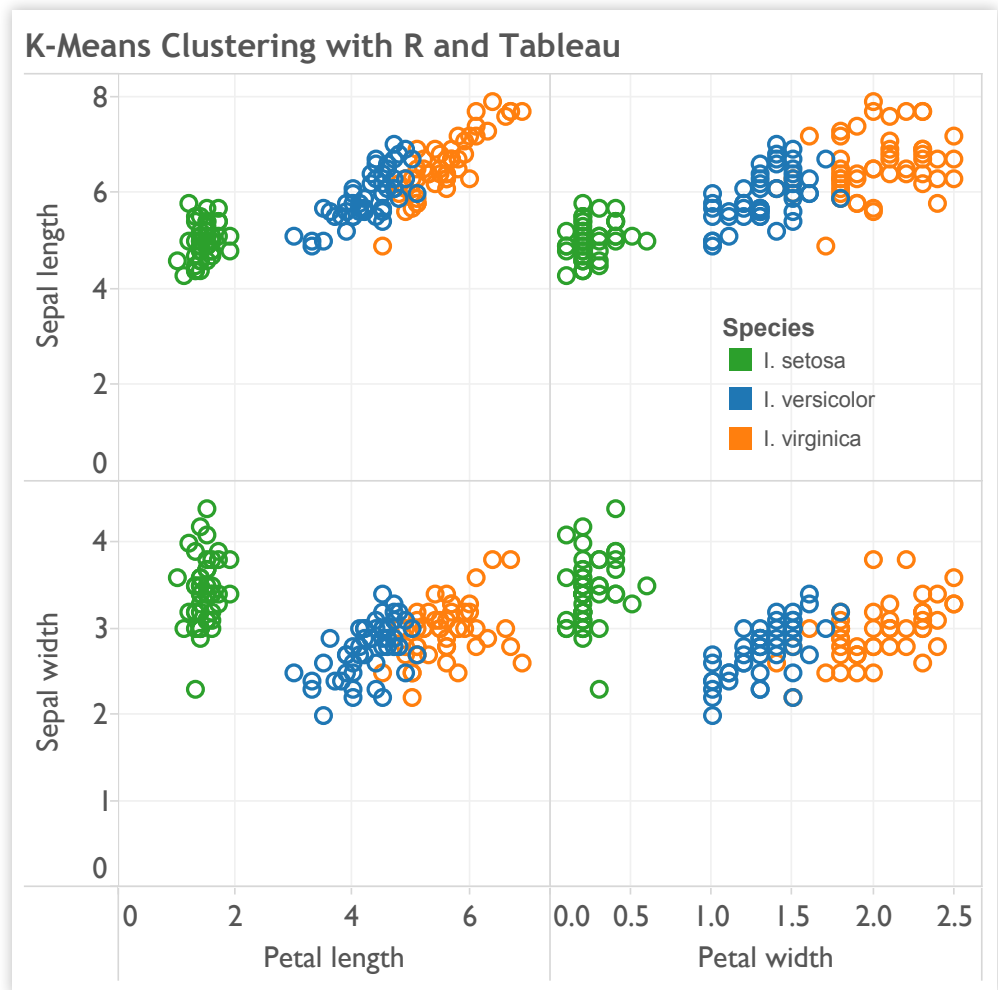
K-Means Clustering with R and Tableau

Who is this feature intended for?

This feature is primarily targeted for users who are already proficient at R. It is NOT meant for beginners with R. Anyone who wishes to use the new functions must first learn how to use R in order to leverage its capabilities in Tableau.

Users who are already proficient with R will find the integration beneficial for several reasons:

- They will be able to do statistical analysis on their Tableau data.

- They will be able to access any R package or function that has been installed in an R server which they can access.

- They will be able to take advantage of all of the visualization capabilities in Tableau to further analyze and understand their data without having to manipulate their data in R (which can be cumbersome) for the same effect.

Pre-requisites for using the feature include:

- Users must have proficiency with the R language to write the appropriate scripts and functional calls they require.

- Users must have access to an R server to access R functions from Tableau Desktop or Tableau Server.[2]

What R capabilities are available in Tableau?

Tableau Desktop:

There are four new built-in functions that are used to call specific R models and functions. The functions are:

- SCRIPT_REAL

- SCRIPT_STR

- SCRIPT_INT

- SCRIPT_BOOL

These functions are distinct only in the type of result they return: a real number, a string, an integer, or a Boolean.

The arguments you pass into each of these functions include R-language scripts and function calls. You can pass 1 or more arguments to R, which are then passed dynamically via Tableau. For instance, you can pass the sales of each customer on a viz.

When using Desktop, you will need to have an R server (via Rserve[3]) that Tableau can access to successfully call the new functions. Additionally, you can import datasets from R directly into Tableau without using functions or calculated fields.

Tableau Server:

Tableau Server must be able to communicate with your R server in order for workbooks with R function calls to work. When this is done, you can share workbooks and views with others while still incorporating the results from the R calculations. If there is no R server running that Tableau Server can access, then these workbooks will not be able to show the correct output.

Tableau Online and Tableau Public:

At the current time, Tableau Online and Tableau Public are not supporting R, so the R statistical capabilities will not be available through these services.

DESKTOP   SERVER   ONLINE   PUBLIC

[3] *What is Rserve? Rserve allows other programs to use facilities of R (see http://www.r-project.org) without the need to initialize R or link against an R library. Rserve supports remote connection, authentication and file transfer. Typical use is to integrate R backend for computation of statistical models, plots etc. in other applications.*

Does this mean Tableau can do Predictive Analytics?

Tableau provides several build-in predictive capabilities such as trending and forecasting. You can enhance your forecasting by showing prediction bands with adjustable levels of certainty, as well as forecast multiple measures per pane. In addition, users can use forecasting models with fiscal calendars and/or multiplicative models to give more precision in forecasts.

Several R packages provide advanced predictive modeling capabilities, which Tableau customers can leverage with R integration.

What R capabilities are not available in Tableau?

Tableau can only integrate with R through the calculated functions described above. This means there are several core capabilities of R that users will not be able to utilize directly in Tableau itself.

- You cannot export data from Tableau into R directly to run a new model outside of using the SCRIPT_* functions.

- Visualizations created in R cannot be imported directly into Tableau. However, image files of R visualizations or URLs to R visualizations may be used in Tableau dashboards.

How do I start using Tableau with R?

For users who are already familiar with R and its capabilities, it is fairly simple to establish the connection between R and Tableau. The instructions below are for new installations using the open-source version of R. Other options may be available using other packages, such as those from Revolution Analytics.

1. **Download and Install R.** Click here to find the file and instructions on downloading R.

2. **Download and Install Rserve.** You will need to install an Rserve for Tableau to connect to in order to utilize the new script functions. In the R console, enter the following commands:

    ```
    install.packages("Rserve")
    library(Rserve)
    Rserve()
    ```

3. **Connect Tableau to the R Server.** Once Rserve is installed, open Tableau Desktop and follow the steps below or watch this video on how to integrate R and Tableau.

    a. Go to the Help menu and select "Manage R Connection".

    b. Enter a server name of "Localhost" (or "127.0.0.1") and a port of "6311".

    c. Click on the "Test Connection" button to make sure everything runs smoothly. You should see a successful message. Click OK to close.

4. **Start using the R scripts in Tableau.** Now you will be able to create new calculated fields in Tableau Desktop that utilize the SCRIPT_* functions to make R functional calls.

About Tableau

Tableau Software helps people see and understand data. Tableau helps anyone quickly analyze, visualize and share information. More than 15,000 customer accounts get rapid results with Tableau in the office and on-the-go. And tens of thousands of people use Tableau Public to share data in their blogs and websites. See how Tableau can help you by downloading the free trial at

www.tableausoftware.com/trial.
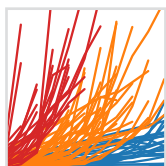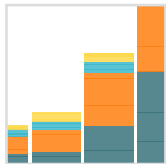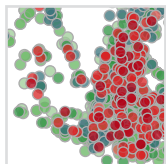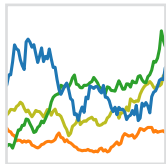
Are there more resources I can utilize to learn R?

Yes, there are several additional resources that may be helpful in getting started with R and leveraging its capabilities in Tableau.

Learning R

The R Project for Statistical Computing

R Tutorial: An R Introduction to Statistics

Intro to R by Google Developers

R Training Tutorials

Coursera: Data and Statistics Courses, and Computing for Data Analysis (Jan 2014)

Book: Software for Data Analysis

Stack Overflow Q&A

R Scripts Repository

Blogs about what people are doing with R

Tableau and R (Blogs)

TableauLove: How to configure Tableau Server to use R

Tableau Blogs: Tableau 8.1 and R

Jennifer Underwood: How to visually analyze R predictive models in Tableau

Data Plus Science: Geocoding addresses directly in Tableau 8.1 using integration with R