

Revolution R Enterprise Tutorial for Hortonworks Sandbox

Introduction

Summary

In this tutorial the user will be introduced to Revolution R Enterprise and how it works with the Hortonworks Sandbox. A data file will be extracted from the Sandbox using ODBC and then analyzed using R functions inside Revolution R Enterprise.

About Revolution R Enterprise

Revolution R Enterprise is a production grade analytics software product built by combining the powerful open source R statistics language with commercial enhancements and professional support necessary for enterprise deployment.

Prerequisites:

- [Hortonworks Sandbox](#)
- Revolution R Enterprise (RRE) 6.2 or later – request evaluation of RRE [here](#).
- Download [data](#) before starting this tutorial
- Insure that the data file “[airline.csv](#)” is pre-loaded into the Hortonworks Sandbox

Overview

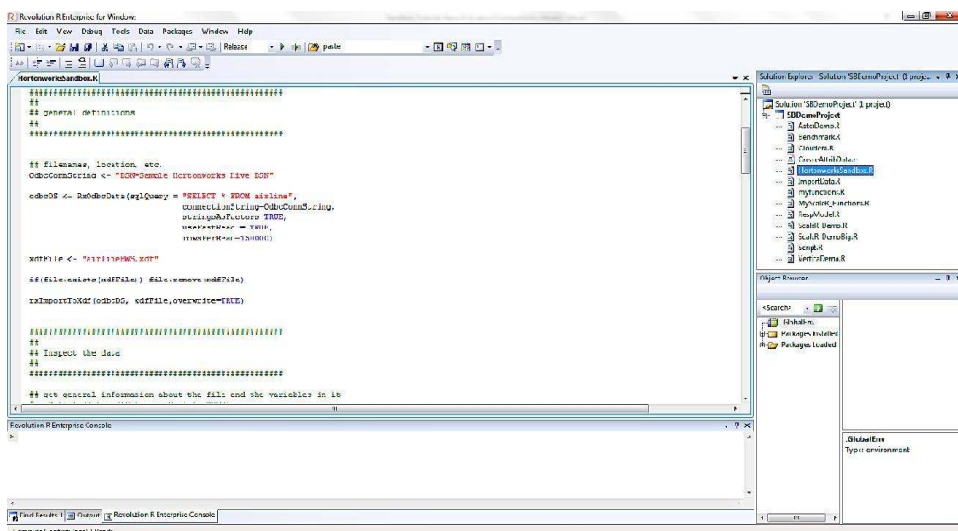
The main steps needed to perform the procedure are:

1. Start RRE and open “HortonworksSandbox.R”.
2. Execute the R functions, either one at a time, as a block, or the entire script.
3. Review results in the console window and object browser.

Step 1 – Start RRE and open Hortonworks.R

In this section you will launch Revolution R Enterprise (RRE) and open your HortonworksSandbox.R source file.

1. Launch Revolution R Enterprise (RRE) from your Start Menu. Inside RRE, go to File->Open->File to open HortonworksSandbox.R and import it into your R project.

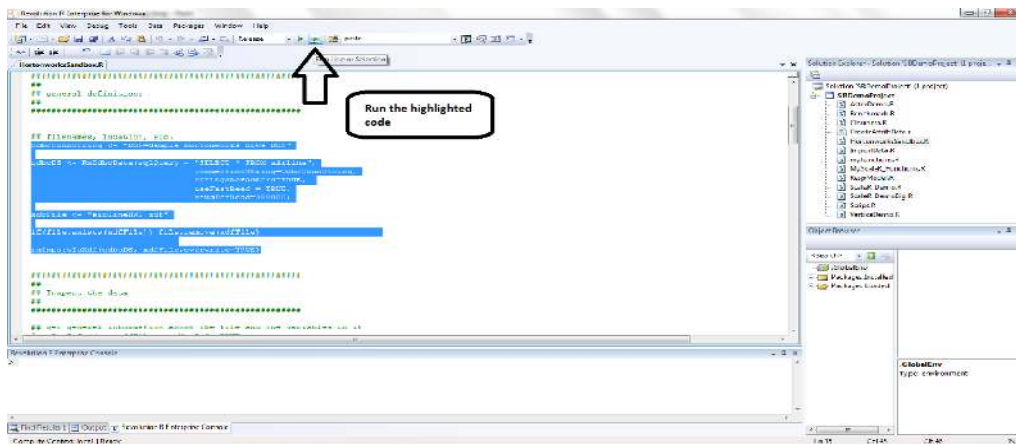


With RRE running and the HortonworksSandbox.R script loaded, we can see the R script in the upper left corner of the RRE Interactive Development Environment (IDE). To the right of the source code is the Solution Explorer – which contains the R project and all source files that belong to the project. Below the Solution Explorer is the Object Browser, which allows you to see all R packages that are installed and loaded, as well as all R objects that you create when you execute your code. To the left of the Object Browser is the RRE Console, which shows the results when you execute R functions.

Step 2 – Execute the R functions, either one at a time, as a block, or the entire script

In this section, you will execute R function(s) from within RRE.

1. There are two rightward-pointing arrows above the source code. The one on the left allows you to run your script from top to bottom all at once. The arrow on the right allows you to run your selected code, either one line at a time or a block of highlighted code at a time.

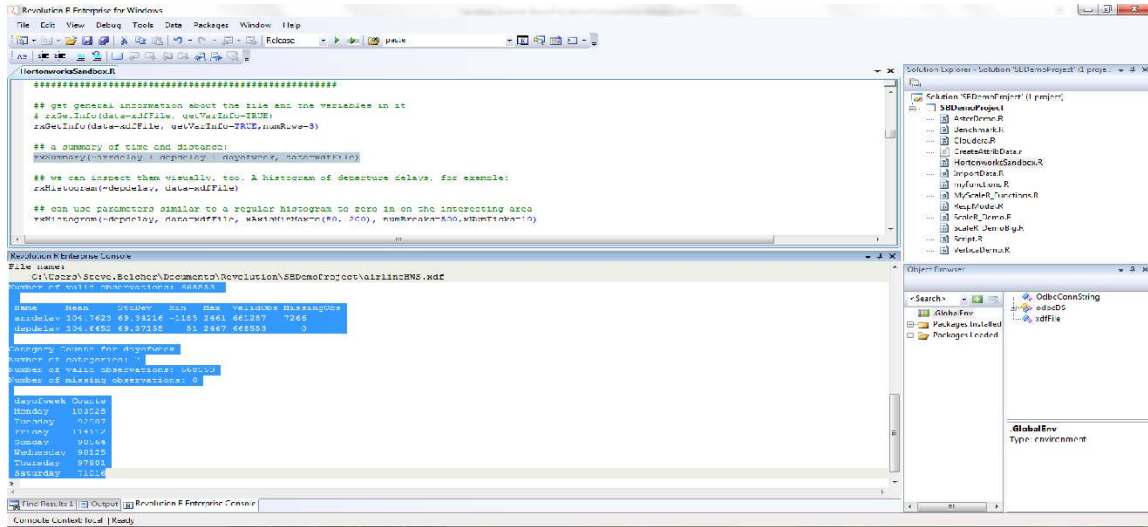


Having executed some R code, you can now examine the results.

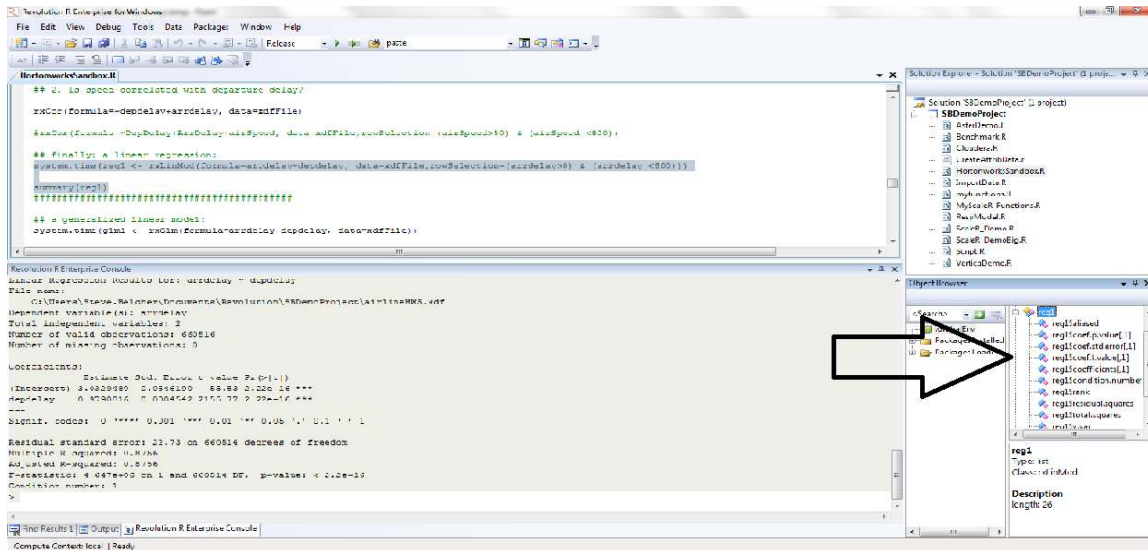
Step 3 – Review results in the console window and object browser

In this section, you will examine the results of the R code that you just executed.

1. When you execute the rxSummary function, you will see basic statistics on the 3 variables that are summarized in the RRE Console window.



2. You can also see information about the R objects that you created by running your code. The objects are found in the lower right side of the IDE in the Object Browser window. In the picture below, a linear regression model object is being displayed in the Object Browser.





You have just executed an R script to extract a file from the Hortonworks Sandbox and analyzed the data in Revolution R Enterprise. Revolution Analytics and the R community provide thousands of packages and functions that are available to you to perform statistical analysis and visualization on any type of data.

Feedback

We're interested in your experience with this tutorial. Please take this [short survey](#).

About Revolution Analytics

Revolution Analytics is the leading commercial provider of software and services based on open source R, the world's most widely-used statistics software. Revolution R Enterprise brings big data scalability, performance, and cross-platform enterprise readiness to R, to meet the production needs of data scientists in data-driven industries worldwide. For more information, visit www.revolutionanalytics.com.

About Hortonworks

Hortonworks develops, distributes and supports the only 100-percent open source distribution of Apache Hadoop explicitly architected, built and tested for enterprise grade deployments. Developed by the original architects, builders and operators of Hadoop, Hortonworks stewards the core and delivers the critical services required by the enterprise to reliably and effectively run Hadoop at scale. Our distribution, Hortonworks Data Platform, provides an open and stable foundation for enterprises and a growing ecosystem to build and deploy big data solutions. Hortonworks also provides unmatched technical support, training and certification programs. For more information, visit www.hortonworks.com. The Hortonworks Sandbox can be found at: www.hortonworks.com/sandbox.