



***Using the R Language with
Infobright***

**Infobright
47 Colborne Street, Suite 403
Toronto, Ontario M5E 1P8 Canada
www.infobright.com
www.infobright.org**

How To Use The R Language with Infobright

R is a well-known programming language originally created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand. In fact, the R name comes from the first letter of each name of the authors. Popular amongst statisticians for handling a wide variety of statistical and data analysis, R is a true computer language that is highly extensible through the use of user submitted libraries or packages. In addition, the popular machine learning software Weka has added support for R in order to give the Weka environment statistical data handling capabilities.

Connecting R to the Infobright® analytic database offers a powerful combined solution for handling deep analysis and statistical computations over the data contained in an Infobright database. Let's now examine one of the most popular methods for connecting R to Infobright, which is by using the RODBC package.

RODBC is an R package originally developed by Michael Lapsley (University of London) and further extended and developed by Brian Ripley (University of Oxford) to enable ODBC connectivity from R to a wide range of database solutions including Infobright.

Installation:

For this guide, we use the Windows version of RODBC from the CRAN distribution for R found here:

<http://cran.r-project.org/web/packages/RODBC/index.html>

In addition to this distribution you will find several helpful references to further assist you in setting up and installing RODBC and R, including setup instructions for ODBC connectivity on Windows and other operating systems. For the purposes of this "How To" guide, we assume you have already installed Infobright on Windows. If you haven't, you can download Infobright Community Edition at <http://www.infobright.org/Download/ICE/>.

Once the RODBC distribution and package has been installed, it is important to ensure that RODBC version 2.10.1 is installed in the Library directory found under the R directory, usually in the Program Files on Windows as shown below in this directory structure:

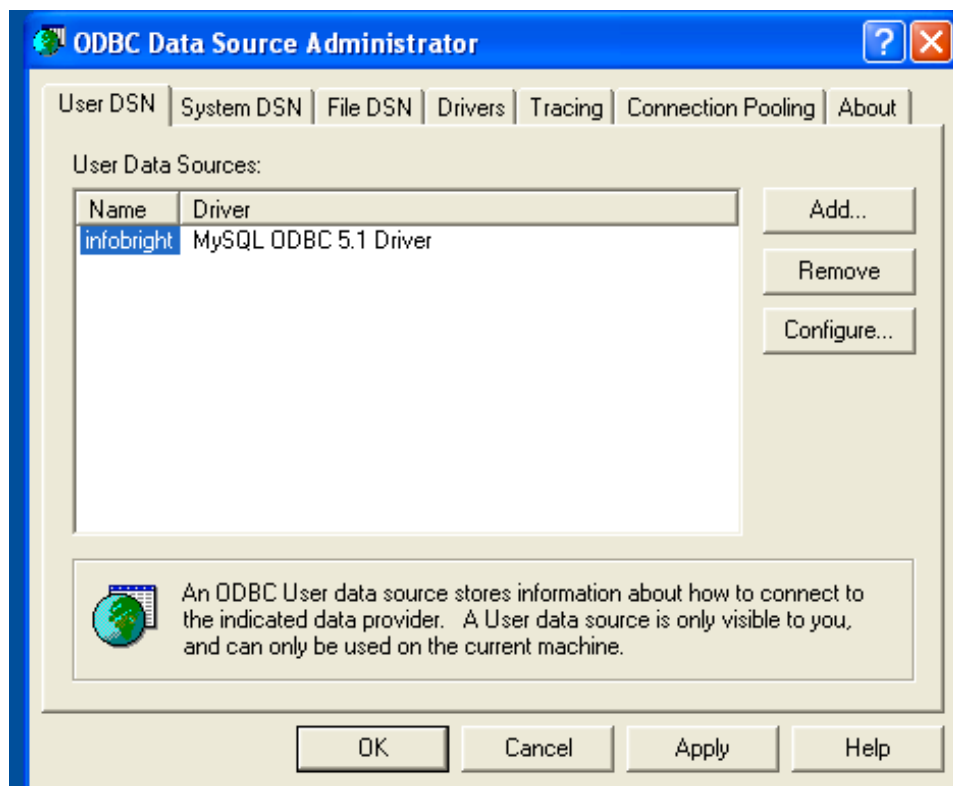
```
Program Files
  R
    R-2.10.1
      Library
        RODBC
```

You will find many other packages installed under the Library subdirectory but be sure that the RODB subdirectory is contained under this Library subdirectory otherwise R will not be able to load the RODB library.

Beyond ensuring that RODB is downloaded and installed and setup in the Library directory, it is required that you also have the ODBC Driver Manager installed, normally included with your Windows installation, along with the ODBC driver for MySQL installed. Accessing the Infobright database is achieved through the use of the MySQL ODBC driver, which can be downloaded from this link:

<http://dev.mysql.com/downloads/connector/odbc/5.1.html#downloads>

The following screen shot shows the ODBC Data Source Administrator indicating that the MySQL ODBC 5.1 Driver and a data source name called "infobright" is installed:

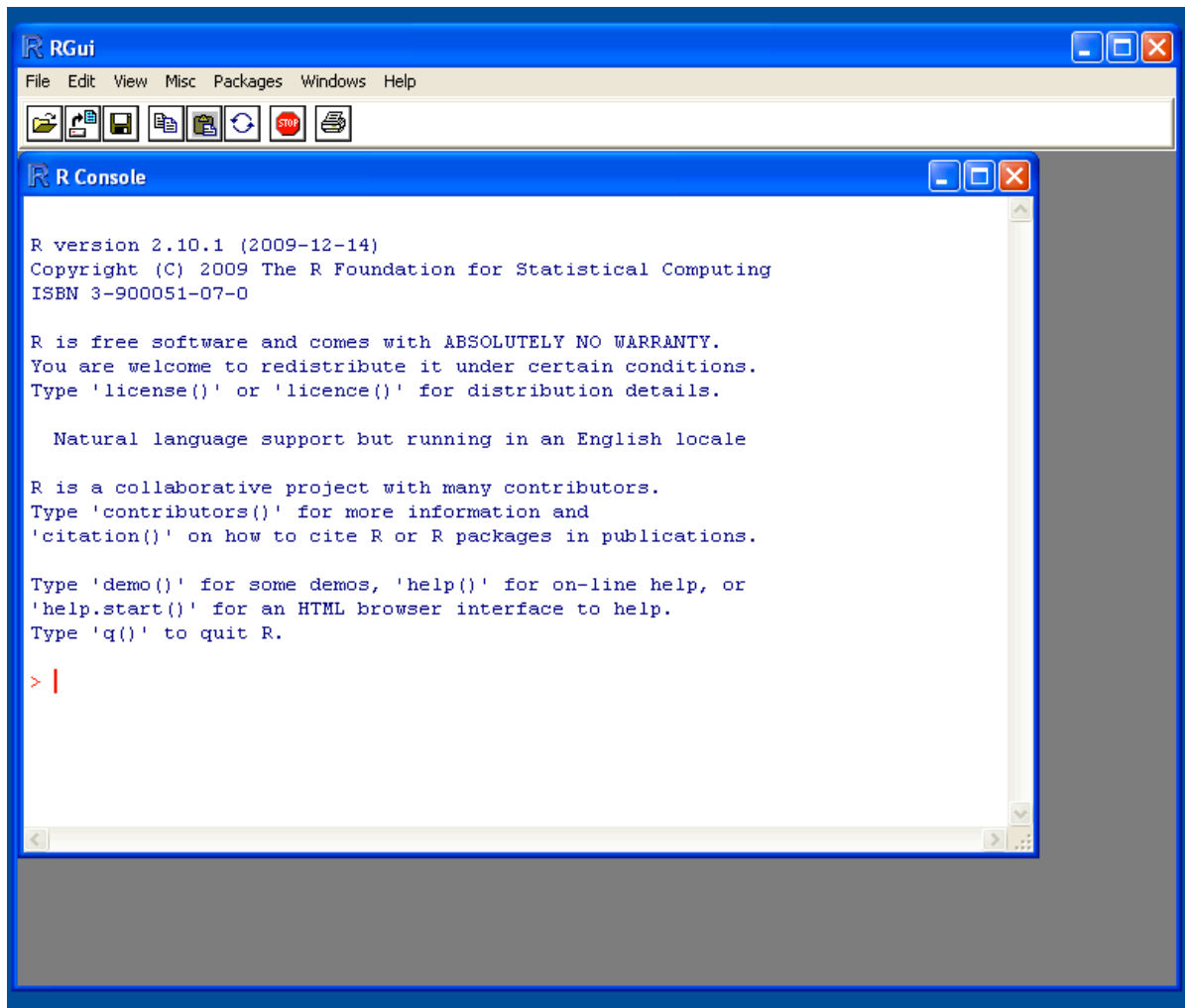


When using R with RODB and connecting to Infobright, you will reference this User DSN. You can call the DSN anything you would like. In this case we named it "infobright".

Now that we have the MySQL ODBC driver installed and a user DSN setup named infobright along with the RODB package installed in the R library directory, we can get started interfacing to Infobright using R.

Using Infobright with R

As part of the CRAN distribution of R, noted above, there is a GUI based front-end tool called RGui along with a console interface called R Console. After installing R for Windows, you will find an R menu under your All Programs under the Start Menu. Clicking on the R icon will start RGui and R Console as shown below:



You should now see the command prompt labeled with > as shown above.

At this point you should now be able to issue commands to the R Console.

Connecting to Infobright is accomplished by first executing the Library function which will load the RODBC package.

```
> library(RODBC)
```

Once the RODB Library is loaded, you can connect to the Infobright database.

The following command will establish the database connection to the "infobright" DSN that you set up during the ODBC configuration process.

```
> con <- odbcConnect("infobright")
```

Alternatively you could also use the `odbcConnect("infobright", uid = "user", pwd = "****")` command with "user" and "****" as the userid and password for connecting to the database.

Now that we have established our connection to the Infobright database, we can begin issuing RODB statement. For example, the "con" variable from our `odbcConnection` call above holds the connection to the database. Once this is established we can now reference that connection in the RODB calls.

To view the tables in the database, one could issue the `sqlTables(con)` call. For the purposes of this example, we are using a car sales database. To see the tables in this database we issue the command:

```
sqlTables(con)
```

When executing in the R Console, you will see results something like the results below with a listing of the tables in the database you are connecting to that you specified when setting up your ODBC connection.

```
> con <-odbcConnect("infobright")
> sqlTables(con)
  TABLE_CAT TABLE_SCHEM          TABLE_NAME TABLE_TYPE
1
2          dim_cars          TABLE
3          dim_cars_myisam      TABLE
4          dim_dates          TABLE
5          dim_dates_myisam    TABLE
6          dim_dealers        TABLE
7          dim_dealers_myisam  TABLE
8          dim_msa            TABLE
9          dim_msa_myisam      TABLE
10         dim_sales_area      TABLE
11         dim_sales_area_myisam TABLE
12         dim_vehicles        TABLE
13         dim_vehicles_myisam TABLE
14         fact_sales          TABLE
15         fact_sales_myisam   TABLE
16         fact_sales_wide     TABLE
17         fact_sales_wide_myisam TABLE
```

R users frequently use or reference what is called a data frame. Using RODBC, you can use the `sqlFetch()` command to load a table into a data frame. In this case we are loading our "dim_cars" table into the frm data frame.

```
frm <-sqlFetch(ch,"dim_cars")
```

Summary

Now that we have shown you how to set up your Windows environment with R, RODBC and Infobright, you have the basic information for establishing a connection between R and Infobright and the ability to issue both R and RODBC commands that interact between the R language and the data in the Infobright analytic database.

For further information on using R with Infobright, we recommend reading the RODBC vignette that provides more detail and information about the database connectivity option for RODBC. All this along with other documentation can be found at this link:

<http://cran.r-project.org/web/packages/RODBC/index.html>

As you use R with Infobright we expect you to see blazing fast performance, as Infobright's columnar analytic database is tuned to the needs of users requiring very fast responses.

To find out more details about Infobright's capabilities, we recommend visiting our website at this link:

<http://www.infobright.com> or <http://www.infobright.org> (our open source community)

You can also contact us via email at info@infobright.com or at one of our locations below:

INFOBRIGHT
403-47 Colborne St
Toronto, ON M5E1P8 Canada
HQ: 416-596-2483
Sales: 630-297-4081
