

MAESTRÍA EN ESTADÍSTICA APLICADA/ MANEJO DE DATOS

Unidad 3: Recopilación, integración y manipulación y almacenamiento de datos

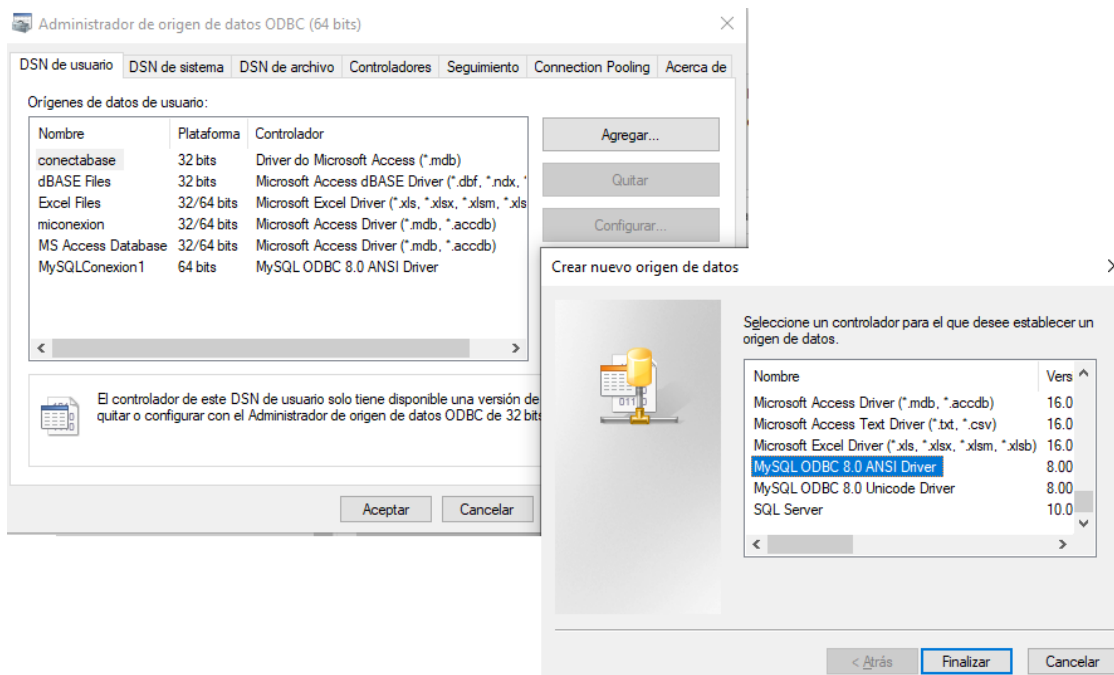
3.1. Conexión con bases de datos desde software estadístico

Una Base de Datos es una colección de datos relacionales, desde donde podemos organizar y almacenar datos. Constituye una fuente común desde dónde podemos encontrar datos. Actualmente existen diversos Sistemas de Gestión de Bases de datos (SGBD/DBMS) para trabajar con bases de datos relacionales, tales como SQLite, MySQL, SQL Server, MariaDB, PostgreSQL, Oracle, entre otros. Con el objetivo de poder extraer los datos y procesarlos desde Rstudio, existen varios paquetes y mecanismos que permiten la conexión con el SGBD.

A continuación, se presenta algunas formas de poder establecer la conexión entre MySQL y Rstudio.

1) Mediante Usuario DSN

Una forma sencilla para establecer la conexión es mediante el estándar de acceso ODBC (Open DataBase Connectivity), desde el cual se pueden crear un usuario con conexión a una fuente de datos DSN (Data Source Name). Para la creación del usuario se lo realiza desde la herramienta administrativa Orígenes de Datos ODBC, el tipo de origen debe ser MySQL ODBC.



Usuario DSN.

Los parámetros que se configuran en el usuario DSN son los siguientes: Data Source Name: [Nombre de la conexión] Port: [Puerto por defecto 3306] User: [Nombre de usuario, por defecto root] Password: [Contraseña ingresada en la instalación de MySQL] Database: [Nombre de la Base de Datos]

The screenshot shows the 'MySQL Connector/ODBC' window with the 'Connection Parameters' tab selected. The 'Data Source Name' is 'MySQLConexion1'. The 'Description' is 'Conexion prueba a base de datos MySQL'. Under 'TCP/IP Server', the 'Port' is set to '3306'. The 'User' is 'root' and the 'Password' is masked with dots. The 'Database' is set to 'world'. There is a 'Test' button next to the database dropdown. At the bottom, there are buttons for 'Details >>', 'OK', 'Cancel', and 'Help'.

Parámetros DSN.

Una vez configurado el usuario DSN, en Rstudio se debe instalar el paquete RODB y luego utilizar la función `odbcConnect` y establecer los parámetros correspondientes, `odbcConnect` ("Nombre del usuario DSN", uid = "root", pwd="contraseña"). Con la conexión establecida, se aplica el comando `sqlQuery` para ejecutar alguna sentencia SQL. A continuación, se muestra un ejemplo de conexión:

```
#if (require("RODBC")) install.packages("RODBC") #instala
ción de paquete RODBC
library(RODBC) #referen
cia a La Librería RODBC
con<-odbcConnect("MySQLconexion1", uid = "root", pwd="Matias2710") #Esta
blecer la conexión con el DBMS
df_ciudad<-sqlQuery(con,"Select * from city") #Ejec
utar una sentencia SQL
head(df_ciudad,5)
```

| ## | ID | Name | CountryCode | District | Population |
|------|----|----------|-------------|----------|------------|
| ## 1 | 1 | Kabul | AFG | Kabul | 1780000 |
| ## 2 | 2 | Qandahar | AFG | Qandahar | 237500 |

| | | | | | | |
|----|---|---|----------------|-----|---------------|--------|
| ## | 3 | 3 | Herat | AFG | Herat | 186800 |
| ## | 4 | 4 | Mazar-e-Sharif | AFG | Balkh | 127800 |
| ## | 5 | 5 | Amsterdam | NLD | Noord-Holland | 731200 |

Descripción de la Sintaxis: El objeto “con” mantiene la conexión con el DBMS mediante el usuario DSN denominado MySQLconexion1. De la base de datos “world”, se extraen todos los registros de la tabla city y se lo almacena en un Dataframe denominado ciudad, finalmente se muestran los datos.

Actividad 1: *Utilizando la misma conexión, extraiga los datos de Código, Nombre, Continente y Población de la tabla country y almacénelo en una Dataframe denominado df_country.*

2) Conexión string ODBC

Otra forma de conectarse con la base de datos MySQL es utilizar la connection string del paquete ODBC. La sintaxis se muestra a continuación:

```
#install.packages("odbc")
require("odbc")

## Loading required package: odbc

## Warning: package 'odbc' was built under R version 3.6.3

con2<-dbConnect(odbc::odbc(),
                .connection_string="Driver={MySQL ODBC 8.0 Unicode Driver
};",
                Server="localhost", Database="world", UID="root", PwD="Ma
tias2710",
                Port=3306)
df_ciudad<-dbGetQuery(con2, statement = "select * from city")
View(df_ciudad)
```

Descripción de la Sintaxis: El objeto “con2” mantiene la conexión con el DBMS mediante el método connection_string. Se enlaza con la base de datos “world”, luego mediante el comando dbGetQuery se ejecuta la sintaxis SQL para extraer todos los registros de la tabla city y se lo almacena en un Dataframe denominado df_ciudad, finalmente se muestran los datos.

Actividad 2: *Utilizando la conexión con ODBC, extraiga los datos de la tabla countrylanguage, de aquellos países cuya población hable el lenguaje en más de un 70%, y almacénelo en una Dataframe denominado df_language.*

Existen otras formas de establecer conexión con MySQL, entre las cuales se tiene, el paquete odbc con driver MySQL o el paquete RMariaDB.