

# Add Some Spark to Your Analysis – Sparklyr: Exercise 1

☒ In this exercise set, we will see how to work with Big data in spark using sparklyr. Please read the documentation and download the data-set required for this exercise from [here](#). Answers to this set are available [here](#).

## Exercise 1

Install the package Sparklyr. Install spark via sparklyr with the spark version 2.2.0.

## Exercise 2

Create a spark connection with Spark Connect using the locally installed spark.

## Exercise 3

Upload the data set into the spark cluster using `spark_read_csv`.

## Exercise 4

Check the data set available in the cluster.

## Exercise 5

Sparklyr can also copy any data frame available in an R session to spark cluster. Copy the iris data-set to the spark cluster.

## Exercise 6

Sparklyr creates a name space for the data frame when you upload a csv or transfer a data frame from R to a spark cluster. You can use `tbl` command to take the meta-data from spark and use it in an R session. Do this and check the resultant tibble.

### **Exercise 7**

Select countries, the motive of the attack, city, and number of kills of the attack from the tibble.

### **Exercise 8**

Now, filter the terrorist activities which have happened after the year 2000. The country of the activity is the US.

### **Exercise 9**

Then, arrange the activities by descending year.

### **Exercise 10**

Create a new column (`large_death_toll`) which is 1 when the number of kills is more than 10 and 0 otherwise.