**NEIGHBOURHOOD DATA BOSTON (US CENSUS and DEMOGRAPHICS)**

The Raw file of US census and demographics has many columns which should be reshaped, renamed and organized under multiple tables. The information of what column needs to be extracted, their alias names, the description of the columns and the tables in which the columns should be stored are stored in file neighbourhood.txt under the following columns.

* Col\_names
* Alais\_names
* Desc
* Category(table-name)

This is a tab separated file that can be opened in excel. The neighbourhood\_required\_columns.R reads all the files from year 1970 to 2010. It melts and stores them in a data frame say *data*. The code piece commented as *#Function Extract columns and Insert* extracts the required columns for various tables from this data frame and creates sql queries for renaming and transforming them as required and insert them into the database using the function postgres\_insert.

A function called postgres\_insert is created to store the user\_name and password by default and recursively insert various tables. The following tables are populated by this current code.

* Children\_and\_Seniors\_in\_poverty
* Educational\_attainment\_by\_race\_ethnicity
* Family\_income\_groups\_by\_race\_ethnicity
* HH\_Income\_by\_race\_ethnicity\_(different\_columns\_over\_the\_years)
* Housing\_cost\_burden
* Labor\_force\_status\_by\_race\_ethnicity
* Public\_Assistance
* Tenure\_by\_race\_ethnicity
* Transit\_commute\_times
* Vehicle\_ownership
* educ\_enrollment\_by\_year\_districts
* educ\_enrollment\_by\_year\_schools

The code is adaptable. It reads the table names from neighbourhood.txt. So if someone needs to extract some-other columns for a different table they can just add the column names and table names in neighbourhour.txt and run the code.

The code creates a new postgres table if the table name specified in neighbourhood.txt does not exist. However it is advisable to create the table with defined column names in postgres and then insert the table names in neighbourhood.txt. It is because the table created will not have intended datatype in postgres.

It is advised to use the below line of code in case of too many connections open error. This ideally should not happen as all connections will be closed if the code runs end to end. However if there is any break in the execution and the code stops in between the existing connections should be closed.

*lapply(dbListConnections(dbDriver("PostgreSQL")), dbDisconnect)*