Working with real world datasets

* Many real data sets are available as a .CSV file
* .CSV COMMA separated values
* Use double quotes to specify mixed-case column names.
  + **Select “Id” from DOGS;**
* By default, spaces are mapped to undescores:
  + Name of Dog = Name\_of\_Dog
  + **Select “Id”, “Name\_of\_Dog”, “Breed\_\_Dominant\_breed\_if\_not\_pure\_breed\_” from dogs** \*\*Note the double underscore between breed and dominant.
* Using quotes in jupyter notebooks
  + **selectQuery = ‘select “Id” from dogs’**
  + **selectQuery = ‘select \* from dogs where “Name\_of\_dog”=\’Huggy’ ‘**
  + USE backslash “\” to split the query into multiple lines.
    - **%sql select “Id”, “Name\_of\_Dog”, \**

**From dogs \**

**Where “Name\_of\_Dog”=’Huggy’**

* + USE %%sql in the first row of the cell
    - **%%sql**

**Select “Id”, “Name\_of\_dog”,**

**From dogs**

**Where “Name\_of\_dog”=’Huggy’**

* Restricting the number of rows retrieved
  + **Select \* from census\_data LIMIT 3**

Getting Tables and Column Details

* Getting list of table names
  + DB2 = SYSCAT.TABLES
  + SQL Server = INFORMATION\_SCHEMA.TABLES
  + ORACLE = ALL\_TABLES OR USER\_TABLES
* On DB2 (getting all tables)
  + **Select \* from syscat.tables**
  + **Select TABSCHEMA, TABNAME, CREATE\_TIME**

**From syscat.tables**

**Where tabschema= ‘db2 user name’**

* Getting list of columns names in database
  + **Select \* from syscat.columns where tabname = ‘DOGS’**
  + To Specific column properties
  + **Select distinct(name), coltype, length**

**From sysibm.syscolumns**

**Where tbname = ‘table name’**