

1. Download the [1994-census-summary.sql](#) and [1994-census-summary.data](#) files from the course website, and load the data into sqlite using the .read command. For example:

- `sqlite> .read data/1994-census-summary.sql`

2. Write SQL statements to answer the following questions:

- What is the average age for each workclass? Expected output:
  - Federal\_gov,43.0
  - Local\_gov,42.0
  - NA,41.0
  - Never\_worked,21.0
  - Private,37.0
  - (etc.)
  -
- What is average number of years of education by occupation? Expected output:
  - Adm\_clerical,10.0
  - Armed\_Forces,10.0
  - Craft\_repair,9.0
  - Exec\_managerial,11.0
  - Farming\_fishing,9.0
  - (etc.)
- What is the average age by occupation, listed in order of average age? Expected output:
  - Armed\_Forces,30.0
  - Handlers\_cleaners,32.0
  - Other\_service,35.0
  - Adm\_clerical,37.0
  - Sales,37.0
  - (etc.)
  -
- What is the average of years of education by both workclass and sex? Expected output:
  - Federal\_gov,Female,11.0
  - Federal\_gov,Male,11.0
  - Local\_gov,Female,11.0
  - Local\_gov,Male,11.0
  - NA,Female,9.0
  - NA,Male,9.0
  - (etc.)

3. Do the following steps related to null handling in the census data:

- edit 1994-census-summary.sql to remove the 'not null' constraint on attribute 'workclass'
- reload the census data into SQLite

- change "NA" values of the attribute 'workclass' to null values, as follows:

```
sqlite> update census set workclass=null where workclass="NA";
```

- Write an SQL query to count the number of rows in the census table.
- Write an SQL query to count the number of 'workclass' values.
- What is the difference between the two count values you just found?
- Write an SQL query to count the number of rows in which workclass is null.