# C312 Advanced Databases Course Work 2: Advanced SQL

Due in 12noon 11th November 2016

## Background Material

The **uscensus1990** database (available on both SQLServer and Postgres in DoC) is copy of data provided by US Census office from their 1990 census. Five of the database tables are illustrated below, each listed with a small fragment of the data from the table.

| county | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| state_code | fips_code | name | type | population | housing_units | land_area | water_area | latitude | longitude |
| 36 | 103 | Suffolk County | county | 1321864 | 481317 | 2360093 | 3786993 | 40.90536 | -72.679044 |
| 6 | 75 | San Francisco County | county | 723959 | 328471 | 120955 | 479639 | 37.793250 | -122.554783 |
| 55 | 29 | Door County | county | 25690 | 18037 | 1250317 | 4887822 | 45.020683 | -87.009973 |
| 55 | 59 | Kenosha County | county | 128181 | 51262 | 706605 | 1247185 | 42.582298 | -87.805528 |
| 55 | 61 | Kewaunee County | county | 18878 | 7544 | 887475 | 1921659 | 44.589317 | -87.440146 |
| 55 | 71 | Manitowoc County | county | 80421 | 31843 | 1532148 | 2336924 | 44.145467 | -87.553328 |
| 34 | 9 | Cape May County | county | 95089 | 85537 | 661007 | 945551 | 39.077466 | -74.858609 |
| 55 | 79 | Milwaukee County | county | 959275 | 390715 | 625649 | 2455935 | 42.975611 | -87.671417 |
| 55 | 89 | Ozaukee County | county | 72831 | 26482 | 600784 | 2290434 | 43.249500 | -87.501558 |
| 55 | 101 | Racine County | county | 175034 | 66945 | 862811 | 1188420 | 42.784761 | -87.755094 |

| mcd | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| state | fips | fips | name | type? | population | housing | land | water | latitude | longitude |
| code | code | subdivision | | | | units | area | area | | |
| | | code | | | | | | | | |
| 1 | 3 | 91053 | Fairhope | division | 16331 | 7361 | 172078 | 194445 | 30.466407 | -87.913337 |
| 72 | 117 | 26846 | Ensenada | barrio | 763 | 410 | 2881 | 10213 | 18.332828 | -67.284330 |
| 1 | 3 | 91152 | Foley | division | 20687 | 17587 | 453800 | 674407 | 30.292000 | -87.763677 |
| 72 | 97 | 1820 | Algarrobos | barrio | 5074 | 1649 | 4301 | 31018 | 18.209253 | -67.194724 |
| 1 | 97 | 90216 | Bayou La Batre | division | 9705 | 4580 | 216863 | 671076 | 30.301019 | -88.192562 |
| 72 | 3 | 70921 | Ri]o Grande | barrio | 864 | 292 | 2596 | 9242 | 18.395570 | -67.235495 |
| 10 | 1 | 92220 | Milford North | division | 6758 | 2938 | 168299 | 215641 | 38.998743 | -75.333951 |
| 66 | 10 | 7250 | Agat | district | 4960 | 1300 | 27192 | 48149 | 13.356057 | 144.633899 |
| 5 | 23 | 93750 | Valley | township | 749 | 660 | 18250 | 29743 | 35.496497 | -92.105746 |

| place | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| state_code | county_code | name | type | population | housing_units | land_area | water_area | latitude | longitude |
| 2 | 2025 | Amchitka | CDP | 25 | 0 | 299980 | 417405 | 51.567103 | 178.877380 |
| 60 | 60100 | Olosega | village | 201 | 47 | 2969 | 42842 | -14.201212 | -169.599688 |
| 2 | 4210 | Atka | city | 73 | 26 | 23772 | 70025 | 52.242218 | -174.205154 |
| 53 | 56304 | Priest Point | CDP | 703 | 313 | 2470 | 7336 | 48.036906 | -122.249727 |
| 2 | 13860 | Chiniak | CDP | 69 | 36 | 103269 | 192748 | 57.631863 | -152.182537 |
| 72 | 65589 | Puerto Real | comunidad | 3429 | 1206 | 1116 | 1666 | 18.072680 | -67.191123 |
| 2 | 82750 | Wainwright | city | 492 | 160 | 10557 | 30331 | 70.599953 | -160.071563 |
| 48 | 72989 | Tiki Island | village | 537 | 441 | 1679 | 1814 | 29.298700 | -94.914177 |
| 2 | 86490 | Yakutat | city | 534 | 189 | 7572 | 12124 | 59.557526 | -139.762121 |

| state | | |
|---|---|---|
| code | abbr | name |
| 1 | AL | ALABAMA |
| 2 | AK | ALASKA |
| 4 | AZ | ARIZONA |
| 5 | AR | ARKANSAS |
| 6 | CA | CALIFORNIA |
| 8 | CO | COLORADO |
| 9 | CT | CONNECTICUT |
| 10 | DE | DELAWARE |
| 11 | DC | DISTRICT OF COLUMBIA |
| 12 | FL | FLORIDA |

| zip | | | | | | |
|---|---|---|---|---|---|---|
| state_code | zip_code | zip_name | longitude | latitude | population | allocation_factor |
| 1 | 35004 | ACMAR | -86.51557 | 33.584132 | 6055 | 0.001499 |
| 1 | 35005 | ADAMSVILLE | -86.959727 | 33.588437 | 10616 | 0.002627 |
| 1 | 35006 | ADGER | -87.167455 | 33.434277 | 3205 | 0.000793 |
| 1 | 35007 | KEYSTONE | -86.812861 | 33.236868 | 14218 | 0.003519 |
| 1 | 35010 | NEW SITE | -85.951086 | 32.941445 | 19942 | 0.004935 |
| 1 | 35014 | ALPINE | -86.208934 | 33.331165 | 3062 | 0.000758 |
| 1 | 35016 | ARAB | -86.489638 | 34.328339 | 13650 | 0.003378 |
| 1 | 35019 | BAILEYTON | -86.621299 | 34.268298 | 1781 | 0.000441 |
| 1 | 35020 | BESSEMER | -86.947547 | 33.409002 | 40549 | 0.010035 |
| 1 | 35023 | HUEYTOWN | -86.999607 | 33.414625 | 39677 | 0.00982 |

The state table contains all states and some territories of the USA, which for the purpose of this exercise will be all referred to as states. Each state is divided into counties or adminstratively equivalent units, which are stored in the county table. The type column of county identifies the type of adminstrative unit. Counties are further divivided into **minor civil divisions** (**mcd**) or adminsitratively equivalent ares held in the mcd table, and again each is associated with the type of unit held.

## Submission

To gain full marks, answers to the following questions should make full use of ANSI SQL commands to write compact and efficient queries, and be laid out such that structure of the query is clear. The queries must also run correctly on the Postgres version of the database, and be submitted electronically to CATE as single batch file adb_2016_cw2.sql by the coursework deadline. For full marks, the queries must also run (unaltered) on the SQLServer version of the database. A template version of the file is available on CATE for download. The queries in the file must be given in the order of the questions below, and be separated by semi-colons.

To test your answer against the Postgres version of the database, you should run the command:

```
psql -h db.doc.ic.ac.uk -d uscensus1990 -U lab -W -f adb_2016_cw2.sql
```

Note that 60% of the marks will be awarded for correctness, and 40% of the marks for style, including efficiency, how concise the queries are, appropriate use of indentation, use of Capital letters for keywords, and expressing join conditions by use of JOIN statements in the FROM clause as opposed to using equals in the WHERE clause.

## Questions

1. List as the scheme (state_name,name) the name of the state and the name of all place entries that have a name that ends in 'City', but which do not have the type column set of 'city'. The result must be ordered by state_name,name.

2. Say that a big city is defined as place of type city with a population of at least 100,000. Write an SQL Query that returns the scheme (state_name,no_big_city,big_city_population) ordered by state_name, listing those states which have either (a) at least five big cities or (b) at least one million poeple living in big cities. The column state_name, is the name of the state, no_big_city is the number of big cities in the state, and big_city_population is the number of people living in big cities in the state.

3. Write an SQL query that returns the scheme (type,place,mcd,county) ordered by type where type is the value of the type column appearing in the place, mcd or county tables. The value of place should be the number times the value of type appears in place. The value of mcd should be the number times the value of type appears in mcd. The value of county should be the number times the value of type appears in county.

4. Write an SQL Query that returns the scheme (name,population,pc_population,land_area,pc_land_area), ordered by name, where name is the name of a state. The population is the sum of the all mcd population figures in the state, and pc_population is the percentage of the whole USA population that this figure represents. Similary, land_area is the sum of the all mcd land area figures in the state, and pc_land_area is the percentage of the whole USA land area that this figure represents. Every state must be listed; and the whole USA population and land area figures must be calculated from the mcd table. All percentage values must be rounded to two decimal places.

5. Write a query returning the scheme (state_name,county_name,population), that lists in order of state name, the five most populous county names in each state in decending order of population, together with the population of those counties.

6. Write a query returning the scheme (zip_code,zip_name,name,distance), that lists in order of zip_code and place name the zip code that is closest to place name. The query should be

restricted to just places and zip codes in the state with state code 6. If should be assumed that only zip codes 5 miles or less from entries in place match the place. Your query should assume that the earth is a perfect sphere of radius 3959 miles, and round the distance figure returned to two decimal places.