Name: ____

COS 480 students must answer all questions that are not marked with a ⋆. The points for each question are indicated in parentheses next to the question number. Questions marked with a ⋆ may also be answered, for extra credit.

COS 580 students must answer all questions, including those marked with a ⋆. Each question is worth 2/3 times the points indicated in parentheses.

Some questions use the database instance that was described in the first homework (summarized below). Recall that your answers to questions that ask for queries should work for all instances of databases conforming to the given schema, not only the one depicted below.

Trees

name varchar(50)	botname varchar(50)	ttype varchar(25)	dia float	height float	minz int	maxz int
White Pine	Pinus strobus	coniferous	30.0	90.0	3	8
Pitch Pine	Pinus rigida	coniferous	18.0	35.0	5	7
Bigtooth Aspen	Populus grandidentata	deciduous	15.0	70.0	3	5
Quaking Aspen	Populus tremuloides	deciduous	13.0	67.5	1	8

Places

city varchar(20)	state varchar(20)	pop integer	zone integer	subzone char(1)	mintemp float
Orono	Maine	9112	5	b	-15.0
Bangor	Maine	31473	5	a	-15.1
Bar Harbor	Maine	4820	5	b	-14.0
Caribou	Maine	8312	4	a	-25.4
Van Buren	Maine	2631	3	a	-35.6
Tucson	Arizona	486699	8	a	39.0

For notational convenience in relational algebra, we shall abbreviate as follows:

`Trees(name, botname, ttype, dia, height, minz, maxz)` $T(N, B, T, D, H, M, X)$

`Places(city, state, pop, zone, subzone, mintemp)` $P(C, S, P, Z, Y, L)$

1. (1 pt.) Write your name in the space provided above.

2. (24 pts.) Indicate the result of evaluating each of the following SQL queries.

   (a)
   ```
   select count(*)
   from Places
   where pop > 5000;
   ```

(b)

```
select count(*) as N
from Places
where pop > 5000
group by state
having min(zone) >= 3
order by N;
```

(c)

```
select count(*) as N
from Places, Trees
where pop > 5000
group by state
having min(zone) >= 3
order by N;
```

(d)

```
select count(*) as N
from Places, Trees
where pop > 5000 and minz <= zone and zone <= maxz
group by state
having min(zone) >= 3
order by N;
```

(e)

```
select state, zone, count(*) as N
from Places, Trees
where pop > 5000 and minz <= zone and zone <= maxz
group by state, zone
having min(zone) >= 3
order by state, zone desc, N;
```

3. (25 pts.) Indicate the result of evaluating each of the following expressions, which use the *extended bag algebra*.

   (a) $\tau_H \pi_{NH} \sigma_{M<4 \,\wedge\, H>50} T$

   (b) $\tau_H \pi_{NH} \sigma_{M<4 \,\wedge\, H>50} (T \bowtie \sigma_{Z \neq 5} P)$

(c) $\tau_H \gamma_{N,\text{sum}(H) \to S} \sigma_{M<4 \wedge H>50} (T \bowtie \sigma_{Z \neq 5} P)$

(d) $\tau_H \gamma_{N,sum(H) \to S} \sigma_{M<4 \wedge H>50} (T \underset{Z < M \vee Z > X}{\bowtie} \sigma_{Z \neq 5} P)$

(e) $\tau_H \gamma_{N,sum(H) \to S} \pi_{NH} (\pi_{NHMX} \sigma_{M<4 \wedge H>50} T \underset{Z < M \vee Z > X}{\bowtie} \pi_Z \sigma_{Z \neq 5} P)$

4. (25 pts.) Write SQL queries as directed below.

   (a) Write a SQL query to find the maximum height of trees for each minimum-zone value. The output should consist of tuples of the form (m, h) where h is the maximum height of trees that have a minimum zone of m. The result should be sorted in ascending order of m.

   (b) Write a SQL query to find the average ratio of height to diameter for trees of each type. The output should consist of tuples of the form (t, r) where r is the average value of the height-to-diameter ratio for trees of type t. The result should be sorted in ascending order of t. Use the same units for the numerator and denominator of the ratio.

   (c) Write a SQL query to find the common names of the tallest trees that grow well in each place. The output should consist of tuples of the form (c, s, n) where n is the common name of a tallest tree that grows well in the city c in state s (that is, the zone of city c in state s is in the range of zones for n). Note that there may be multiple tallest trees per place.

(d) Write a SQL query to find the average height of trees that grow well in the zones of each state. The output should consist of tuples of the form (s, h), where s is the name of a state and h is the average height of trees that grow well in one or more places in s. Note that the average is taken over all trees that grow well at some place in a state, irrespective of the number of such places in the state. [Hint: Watch out for trees that grow in several places in a state; each such tree should contribute only once to the average for that state.]

(e) Let p_1 and p_2 be places in zone z with minimum temperatures t_1 and t_2, respectively, such that no place in z has a minimum temperature lower than t_1 or higher than t_2. We say $t_2 - t_1$ is the *minimum-temperature range* of z. Write a SQL query to find the zones that have the largest minimum-temperature range. The output should consist of tuples of the form (z) where z is a zone satisfying the above condition. (Note that z may not be unique.)

5. (25 pts.) For each part below, use the extended bag algebra to write a query that is equivalent to the SQL query in the corresponding part of Question 4.

   (a)

   (b)

   (c)

   (d)

   (e)

6. (50 pts.) $\star$

(a) Write a SQL query that is equivalent to the bag algebra expression in Question 3d.

(b) Write a SQL query that is equivalent to the bag algebra expression in Question 3e.

(c) Prove or disprove the equivalence of the bag algebra expressions of Questions 3d and 3e.

(d) Prove or disprove the equivalence of your answers to Questions 6a and 6b by proving each query equivalent to the corresponding algebra expression.

(e) Prove or disprove the equivalence of your answers to Questions 6a and 6b directly, *without mapping to algebra*; rather, use the direct interpretation of SQL queries.