WQ2 due **TOMORROW!!!** (Friday, Oct 13th at 11:59 PM)

HW3 due **Friday, Oct 20th** at 11:00 PM

Set up Azure ASAP

Last day to turn in HW2 (with late days)
Azure Setup

- Set up early to avoid any problems!
  - Instructions in hw3 spec

- You should have received an email with an invitation from Microsoft to create an Azure account/database
  - If you haven’t received an invitation email or are having problems setting up your database, please post on Piazza or contact course staff at cse344-staff@cs.washington.edu
How to Run HW3 Queries

- Azure portal
  - Still in preview mode; limited features
- SQL Server Management Studio (SSMS)
  - Fully-featured
  - Enables viewing of query execution plans
  - Windows only (lab computers)
- DataGrip
  - [https://www.jetbrains.com/student/](https://www.jetbrains.com/student/)
  - Setup instructions posted on Piazza
Azure Set-up: Demo
Group By

- Powerful tool to handle “categories”
  - Treat rows with a same attribute as a category
- Careful when selecting
  - Only select attributes appeared in GROUP BY or aggregates
  - SQLite will guess (arbitrarily pick a value)
  - SQL Server will throw an error
Group By - Examples

Do these queries work?

<table>
<thead>
<tr>
<th>Enrolled(stu_id, course_num)</th>
</tr>
</thead>
<tbody>
<tr>
<td>johndoe</td>
</tr>
<tr>
<td>johndoe</td>
</tr>
<tr>
<td>maryjane</td>
</tr>
<tr>
<td>maryjane</td>
</tr>
<tr>
<td>maryjane</td>
</tr>
</tbody>
</table>

SELECT stu_id, course_num
    FROM Enrolled
    GROUP BY stu_id

SELECT stu_id, count(course_num)
    FROM Enrolled
    GROUP BY stu_id
Group By - Examples

Do these queries work?

Enrolled(stu_id, course_num)

<table>
<thead>
<tr>
<th>johndoe</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>maryjane</td>
<td>?</td>
</tr>
</tbody>
</table>

SELECT stu_id, course_num
FROM Enrolled
GROUP BY stu_id

SELECT stu_id, count(course_num)
FROM Enrolled
GROUP BY stu_id
Group By - Examples

Do these queries work?

Enrolled(stu_id, course_num)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>johndoe</td>
<td>2</td>
</tr>
<tr>
<td>maryjane</td>
<td>3</td>
</tr>
</tbody>
</table>

SELECT stu_id, course_num
FROM Enrolled
GROUP BY stu_id

SELECT stu_id, count(course_num)
FROM Enrolled
GROUP BY stu_id
Group By - Examples

What happens when we try to do:

```
SELECT attr_1, attr_2, ..., attr_n
FROM ...
GROUP BY attr_1, attr_2, ..., attr_n;
```
Group By - Examples

What happens when we try to do:

```
SELECT attr_1, attr_2, ..., attr_n
FROM ...
GROUP BY attr_1, attr_2, ..., attr_n;
```

This is like `SELECT DISTINCT ...`
Witnessing (i.e. argmax)

Find the student who is taking the most classes.

Student(stu_id, id_num)
Enrolled(id_num, class)

<table>
<thead>
<tr>
<th>Student</th>
<th>973</th>
<th>CSE 311</th>
</tr>
</thead>
<tbody>
<tr>
<td>johndoe</td>
<td>973</td>
<td>CSE 344</td>
</tr>
<tr>
<td>maryjane</td>
<td>712</td>
<td>CSE 311</td>
</tr>
<tr>
<td>alsmith</td>
<td>899</td>
<td>CSE 351</td>
</tr>
</tbody>
</table>

SELECT S.stu_id
FROM Student S, Enrolled E
WHERE S.id_num = E.id_num
GROUP BY S.stu_id
HAVING COUNT(E.class) >= ALL(
    SELECT COUNT(E1.class)
    FROM Enrolled E1
    GROUP BY E1.id_num);
Nested Queries

- Avoid when possible
- Danger of making simple queries slow and complicated
- Just because you can do it, doesn’t mean you should
Subquery in SELECT

SELECT DISTINCT C.cname, (SELECT count(*)
    FROM Product P
    WHERE P.cid=C.cid)

FROM Company C
Unnest using JOIN and GROUP BY

SELECT C.cname, COUNT(P.cid)  
FROM Company C  
LEFT OUTER JOIN Product ON C.cid = P.cid  
GROUP BY C.cname;
Subquery in FROM

```
SELECT  X.pname
FROM    (SELECT  *
         FROM    Product
         WHERE   price > 20) AS  X
WHERE   X.price < 500
```

More readable: WITH <name> AS <subquery>
Subquery in FROM

Unnest using `WHERE`

```sql
SELECT X.pname
FROM Product AS X
WHERE X.price < 500 AND X.price > 20;
```
SELECT DISTINCT C.cname
FROM Company C
WHERE EXISTS (SELECT *
               FROM Product P
               WHERE C.cid = P.cid AND P.price < 200)
SELECT DISTINCT C.cname
    FROM Company C, Product P
WHERE C.cid = P.cid AND P.price < 200
Subquery in WHERE Syntax

- `SELECT ........ WHERE EXISTS (sub);`
- `SELECT ........ WHERE NOT EXISTS (sub);`
- `SELECT ........ WHERE attribute IN (sub);`
- `SELECT ........ WHERE attribute NOT IN (sub);`
- `SELECT ........ WHERE attribute > ANY (sub);`
- `SELECT ........ WHERE attribute > ALL (sub);`
(Non-)monotonic Queries

- “Can we take back outputs by looking at more data?”
- Is this a monotonic query?

```
SELECT count(*)
    FROM T1
GROUP BY T1.attr
```
(Non-)monotonic Queries

- “Can we take back outputs by looking at more data?”
- Is this a monotonic query?

```
SELECT count(*)
FROM T1
GROUP BY T1.attr
```

No! This query does not satisfy set containment.

Ex:
Current output: {6, 23, 10}
After more data: {6, 23, 11}

{6, 23, 10} ⊄ {6, 23, 11}
To Nest or Not to Nest

- Not an exact science
- Figuring out what is actually wanted will help you find simpler solutions (best way is to practice)
- Trigger words to use sub-querying
  - Every, All (universal quantifiers)
  - No, None, Never (negation)
  - Only