

BAIS:6050 Data Management and Visual Analytics



Week 5 – SQL I

- Post-Mortem: Homework 2
- Lecture Recap
- Demonstration
- Exercises



Need Tech Support?
Call 319-335-3925

2

BAIS:6050 Data Management and Visual Analytics



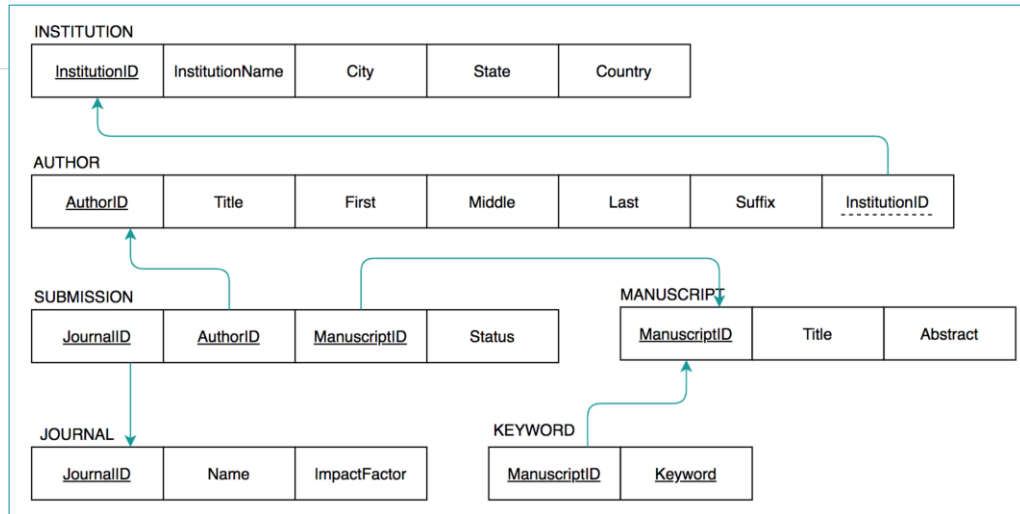
Post-Mortem: Homework 2

- Average = 88%

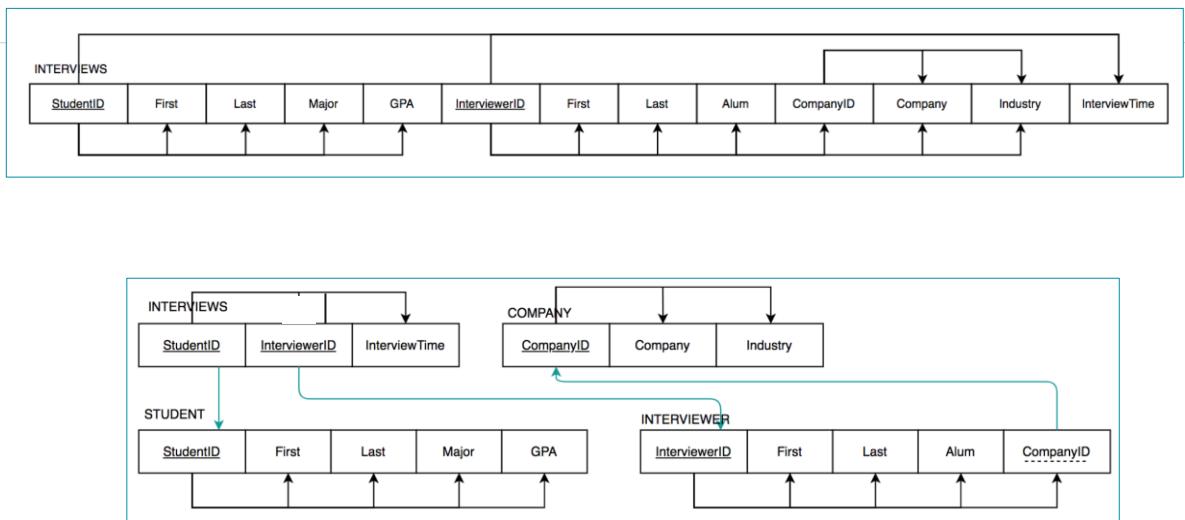


Need Tech Support?
Call 319-335-3925

3



4



5

BAIS:6050 Data Management and Visual Analytics



Lecture Recap

- Lecture 5.1: Data Manipulation Language
 - INSERT, UPDATE, DELETE
- Lectures 5.2-5.5: Single-Table Queries
 - SELECT: Which fields do I want to display in my results?
 - FROM: Which table(s) does the data come from?
 - WHERE: Which rows do I want to include in my results
 - GROUP BY: Do I need to aggregate and/or calculate summary statistics for specific groups?
 - HAVING: Which groups do I want to include in my results
 - ORDER BY: How do I want to sort my results?
 - FETCH: Do I want only the top- or bottom-ranked results? (*Oracle SQL*)



Need Tech Support?
Call 319-335-3925

6

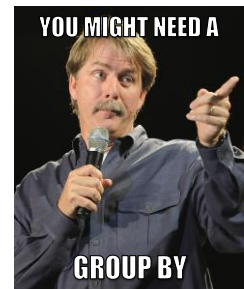
BAIS:6050 Data Management and Visual Analytics



Rules for Queries

- If you call an aggregate function in your query, **and also** select regular (ungrouped) fields, generally need GROUP BY
- It is recommended to rename new fields that will appear in the query results
- Be careful about sorting with text values
 - SELECT * from student order by firstname
- Usually we do NOT **hard-code** PK values to filter data (unless given in question description)
 - Example: Return Britta Perry's student record


```
SELECT * FROM STUDENT
WHERE FirstName = 'Britta'
AND LastName = 'Perry';
```



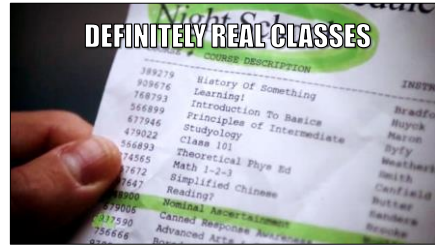
7

BAIS:6050 Data Management and Visual Analytics



Query demo

- Write a simple query to return the full name for all faculty members that have a mentor.
- Write a simple query that returns the course ID and average student grade percentage for all courses that have at least 4 students enrolled. Format the average as a percentage (e.g., 88%).



Need Tech Support?
Call 319-335-3925

8

BAIS:6050 Data Management and Visual Analytics



Solution (Q1)

The full name for all faculty members that have a mentor.

```
SELECT FirstName || ' ' || LastName AS FullName
FROM FACULTY
WHERE Mentor IS NOT NULL;
```

Concatenate First and Last
(with space between)

Rename new field

Filter data (Return any faculty
that HAS a value for Mentor)

FULLNAME
Charlotte Baker
Ian Duncan
Sean Garrity
Marion Holly
Roger DeSalvo



9

BAIS:6050 Data Management and Visual Analytics



Solution (Q2)

The course ID and average student grade percentage for all courses that have at least 4 students enrolled. Format the average as a percentage

```
SELECT CourseID, ROUND(AVG(Percentage) * 100, 0) || '%'
AS AverageGrade
FROM GRADE
GROUP BY CourseID
HAVING COUNT(StudentID) >= 4;
```

Rescale percentage

Round percentage

Append %

Filter data (Return courses that have at least 4 total students)

COURSEID	AVERAGEGRADE
BUS:110	91%
ANT:120	84%
SPA:101	83%
LEI:440	88%



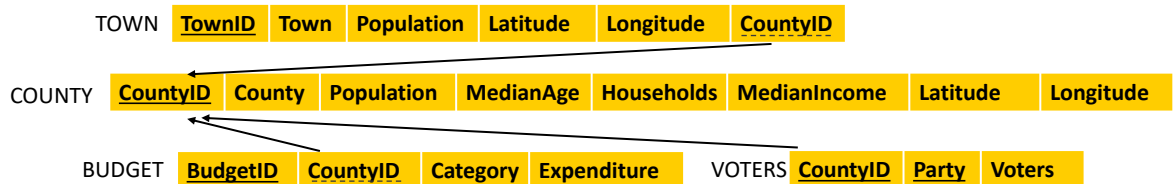
10

BAIS:6050 Data Management and Visual Analytics



Iowa Database

- Download "IOWA.sql" from ICON
- Upload to SQL Scripts and Run
- Iowa Database:
 - COUNTY: 99 rows
 - BUDGET: 891 rows
 - VOTERS: 495 rows
 - TOWN: 41 rows

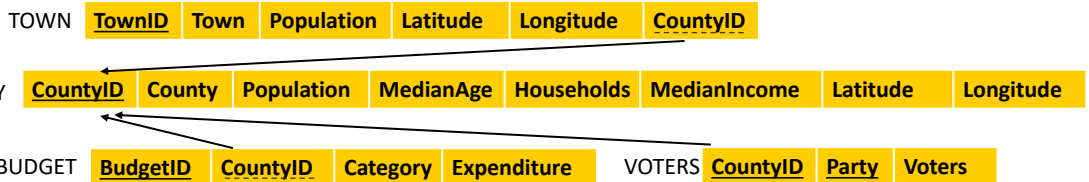


11

BAIS:6050 Data Management and Visual Analytics

**Exercise 1**

- Write a simple query to return the min, max, and median budget amount spent on Public Safety & Legal Services over all counties.
- Which counties have the most towns? For each county that has more than one town (i.e., appears more than once in the TOWN table), return the county ID, total number of towns, and total people living in towns.



12

BAIS:6050 Data Management and Visual Analytics

**Solution (Q1)**

Return the min, max, and median budget amount spent on Public Safety & Legal Services over all counties.

```
SELECT MIN(Expenditure) AS MinSpending,
       MAX(Expenditure) AS MaxSpending,
       MEDIAN(Expenditure) AS MedianSpending
FROM BUDGET
WHERE Category = 'Public Safety & Legal Services';
```

MINSPENDING	MAXSPENDING	MEDIANSPENDING
821290	74288161	2988035



- Filter data by budget category (must match spelling/capitalization exactly)
- Use aggregate functions, but no GROUP BY (no other fields in SELECT)

13

BAIS:6050 Data Management and Visual Analytics

**Solution (Q2)**

For each county that has more than one town (i.e., appears more than once in the TOWN table), return the county ID, total number of towns, and total people living in towns.

```
SELECT CountyID, COUNT(TownID) AS Towns,
SUM(Population) AS People
FROM TOWN
GROUP BY CountyID
HAVING COUNT(TownID)>1;
```

COUNTYID	TOWNS	PEOPLE
19153	8	459312
19111	2	20863
19163	2	138133
19113	2	171628
19103	3	115492
19013	2	109157
19181	2	26886



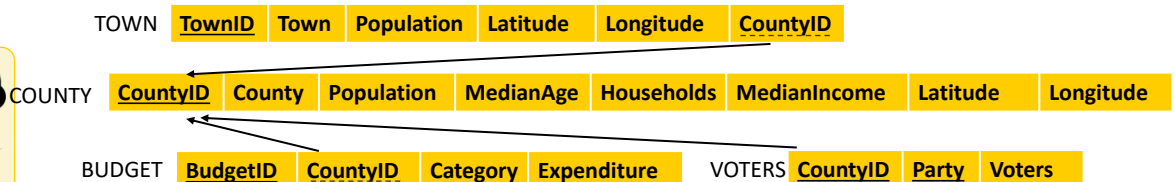
- Group data by CountyID
- Use HAVING to filter data based on aggregate function (COUNT)

14

BAIS:6050 Data Management and Visual Analytics

**Exercise 2**

- Write a simple query to return the number of Iowa towns with more than 100,000 residents
- Write a simple query to return all information about counties whose median age is between 18 to 35 years old (inclusive).



15

BAIS:6050 Data Management and Visual Analytics

**Solution (Q1)**

Return the number of Iowa towns with more than 100,000 residents

```
SELECT COUNT(TownID) AS TotalBigCities
FROM TOWN
WHERE Population > 100000;
```

TOTALBIGCITIES
3



- Filter data by town population (no quotes for numeric values)
- Use aggregate functions, but no GROUP BY (no other fields in SELECT)

16

BAIS:6050 Data Management and Visual Analytics

**Solution (Q2)**

Return all information about counties whose median age is between 18 to 35 years old (inclusive).

```
SELECT *
FROM COUNTY
WHERE MedianAge BETWEEN 18 AND 35;
```

COUNTYID	COUNTY	POPULATION	MEDIANAGE	HOUSEHOLDS	MEDIANINCOME	LATITUDE	LONGITUDE
19013	Black Hawk	132960	34.9	52811	50916	42.4700957	-92.3088197
19103	Johnson	144425	29.9	57423	59965	41.6715511	-91.5880849
19167	Sioux	34692	33.3	12113	66022	43.0826174	-96.1778827
19169	Story	95888	25.9	37106	52671	42.0362415	-93.4650448



- Use * wildcard to select all fields from a table
- BETWEEN operator filters data based on range of numbers
- Could also use MedianAge >=18 AND MedianAge <=35

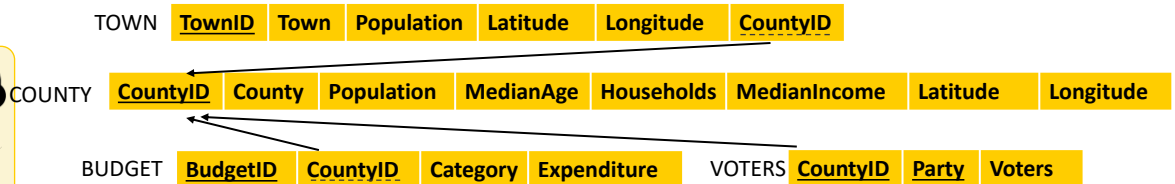
17

BAIS:6050 Data Management and Visual Analytics



Exercise 3

- Which counties have multi-word names? Write a simple query to return the ID and name for any counties whose name has multiple words.
- What are the least-common political affiliations in Iowa? Return the party name and total voters for any parties that have fewer than 100,000 registered voters across the state.



18

BAIS:6050 Data Management and Visual Analytics



Solution (Q1)

Write a simple query to return the ID and name for any counties whose name has multiple words.

```
SELECT CountyID, County
FROM COUNTY
WHERE County LIKE '% %';
```

- Use LIKE and wildcard operators for partial text matching
- Pattern will match any county name with a space anywhere



COUNTYID	COUNTY
19013	Black Hawk
19021	Buena Vista
19033	Cerro Gordo
19057	Des Moines
19147	Palo Alto
19177	Van Buren

19

BAIS:6050 Data Management and Visual Analytics

**Solution (Q2)**

Return the party name and total voters for any parties that have fewer than 100,000 registered voters across the state.

```
SELECT Party, SUM(Voters) AS TotalVoters
FROM VOTERS
GROUP BY Party
HAVING SUM(Voters) < 100000;
```

PARTY	TOTALVOTERS
Libertarian	10300
Other	3220



- Group data by party name
- Use HAVING to filter data based on aggregate function (SUM)

20

BAIS:6050 Data Management and Visual Analytics

**Practice question 1**

- Write a simple query that returns the total number of Independent voters in Iowa (over all counties).
- Find the top 3 counties with the largest budgets. For each, return the county ID and total budget amount (formatted as a currency value, e.g., \$1,240,311).



21

BAIS:6050 Data Management and Visual Analytics

**Solution (Q1)**

```
SELECT SUM(Voters) AS TotalIndependent
FROM VOTERS
WHERE Party = 'Independent';
```

TOTALINDEPENDENT
781011



- Filter data by party name (must match spelling/capitalization exactly)
- Use aggregate functions, but no GROUP BY (no other fields in SELECT)

22

BAIS:6050 Data Management and Visual Analytics

**Solution (Q2)**

```
SELECT CountyID, TO_CHAR(SUM(Expenditure),
'$999,999,999') AS TotalBudget
FROM BUDGET
GROUP BY CountyID
ORDER BY SUM(Expenditure) DESC
FETCH FIRST 10 ROWS ONLY;
```

- Group data by CountyID
- Apply currency formatting with TO_CHAR()
- Use ORDER BY and FETCH to find top 3 counties

COUNTYID	TOTALBUDGET
19153	\$272,192,978
19113	\$107,617,949
19103	\$102,468,933



23

BAIS:6050 Data Management and Visual Analytics



Practice question 2

- Find the top 5 counties by median income. For each, return the county ID, name, and median income (formatted as a currency value, e.g., \$68,019).
- Assume that western Iowa includes all counties west of Des Moines (county location is less than -93.75 degrees of longitude). Write a simple query to return the number of counties and total population of western Iowa.



24

BAIS:6050 Data Management and Visual Analytics



Solution (Q1)

```
SELECT CountyID, County, TO_CHAR(MedianIncome, '$999,999')
AS MedIncome
FROM COUNTY
ORDER BY MedianIncome DESC
FETCH FIRST 5 ROWS ONLY;
```

- Apply currency formatting with TO_CHAR()
- Use ORDER BY and FETCH to find top 5 counties

COUNTYID	COUNTY	MEDINCOME
19049	Dallas	\$82,719
19181	Warren	\$71,514
19129	Mills	\$67,949
19167	Sioux	\$66,022
19017	Bremer	\$65,440



25

BAIS:6050 Data Management and Visual Analytics

**Solution (Q2)**

```
SELECT COUNT(CountyID) AS TotalCounties, SUM(Population)
AS TotalPopulation
FROM COUNTY
WHERE Longitude < -93.75;
```

TOTALCOUNTIES	TOTALPOPULATION
43	800430



- Filter data by longitude (no quotes for numeric values)
- Use aggregate functions, but no GROUP BY (no other fields in SELECT)