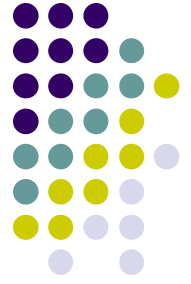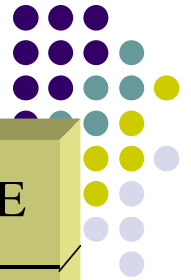# ROLAPs, Rollups and Cubes:
an Introduction to Super Groups

CS240A Notes from

A Complete Guide to DB2 Universal Database, by Don Chamberlin, Morgan Kaufmann Publishers, Inc., 1998

The Census Database

NAME	CITY	COUNTY	STATE	BIRTHDATE	SEX	INCOME
Joe	Miami	Dade	FL	08/20/55	M	32100
Chen	Miami	Dade	FL	06/05/57	M	40200
Bob	Hialeah	Dade	FL	03/21/57	M	33500
Karen	Hialeah	Dade	FL	08/23/55	F	43900
Jim	(null)	Dade	FL	10/24/56	M	29600
Joan	(null)	Dade	FL	11/15/56	F	36300
Dave	Orlando	Orange	FL	09/25/57	M	38000
Linda	Orlando	Orange	FL	05/13/55	F	46700
Jeff	Taft	Orange	FL	02/08/57	M	32600
Pat	Taft	Orange	FL	10/30/57	F	26500
Sam	Baytown	Harris	TX	03/02/55	M	28500
Bill	Baytown	Harris	TX	12/21/56	M	32800
Mary	Houston	Harris	TX	(null)	F	44700
Susan	Houston	Harris	TX	04/30/55	F	(null)
Alex	Houston	Harris	TX	07/11/57	M	30900
John	Austin	Travis	TX	01/06/56	M	38400
Fred	Austin	Travis	TX	10/25/56	M	43500
Anne	(null)	Travis	TX	08/17/55	F	34800

ROLLUP

SELECT state, avg(income) AS avg_income
FROM census
GROUP BY ROLLUP(state)

STATE	AVG__INCOME
FL	35940
TX	36085
(null)	36000
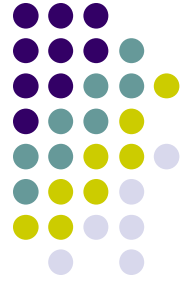
ROLLUP: a more complex example

The total population and the average income in each city, county, and state, and the census as a whole:

```
SELECT      state, county, city
            count(*) AS  population,
            avg(income) AS  avg_income
FROM census
GROUP BY ROLLUP (state, county, city);
```
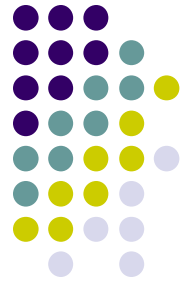
ROLLUP more complex example: Result

STATE	COUNTY	CITY	POPULATION	AVG_INCOME
FL	Dade	Hialeah	2	38700
FL	Dade	Miami	2	36150
FL	Dade	(null)	2	32950
FL	Orange	Orlando	2	42350
FL	Orange	Taft	2	29550
TX	Harris	Baytown	2	30650
TX	Harris	Houston	3	37800
TX	Travis	Austin	2	40450
TX	Travis	(null)	1	34800
FL	Dade	(null)	6	35933
FL	Orange	(null)	4	35950
TX	Harris	(null)	5	34225
TX	Travis	(null)	3	38566
FL	(null)	(null)	10	35940
TX	(null)	(null)	8	36085
(null)	(null)	(null)	18	36000

Getting Rid of nulls
by using the grouping(A) function

```
SELECT
CASE grouping(state)
    WHEN 1 THEN '(-all-)' ELSE state END AS state,
CASE grouping(county)
    WHEN 1 THEN '(-all-)' ELSE county END AS county,
CASE grouping(city)
    WHEN 1 THEN '(-all-)' ELSE city END AS city,
count(*) AS pop,
avg(income) AS avg_income

FROM census

GROUP BY ROLLUP(state, count, city);
```
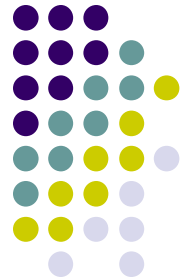
Result of getting rid of nulls

| STATE	COUNTY	CITY	POPULATION	AVG_INCOME
FL	Dade	Hialeah	2	38700
FL	Dade	Miami	2	36150
FL	Dade	(null)	2	32950
FL	Orange	Orlando	2	42350
FL	Orange	Taft	2	29550
TX	Harris	Baytown	2	30650
TX	Harris	Houston	3	37800
TX	Travis	Austin	2	40450
TX	Travis	(null)	1	34800
FL	Dade	(-all-)	6	35933
FL	Orange	(-all-)	4	35950
TX	Harris	(-all-)	5	34225
TX	Travis	(-all-)	3	38566
FL	(-all-)	(-all-)	10	35940
TX	(-all-)	(-all-)	8	36085
(-all-)	(-all-)	(-all-)	18	36000

Using WHERE & HAVING in ROLLUPs

```
SELECT
CASE grouping(state)
      WHEN 1 THEN '(-all-)' ELSE state END AS
state,
CASE grouping(county)
   WHEN 1 THEN '(-all-)' ELSE county END AS county,
CASE grouping(city)
      WHEN 1 THEN '(-all-)' ELSE city END AS city,
count(*) AS pop,   avg(income) AS avg_income

FROM census
WHERE sex= 'F'
GROUP BY ROLLUP(state, count, city)
HAVING count(*) >=2;
```

Result of WHERE and HAVING on ROLLUP
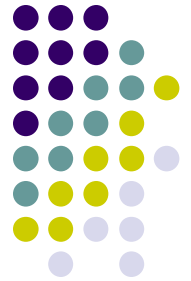
STATE	COUNTY	CITY	F_POP	AVG_F_INCOME
TX	Harris	Houston	2	44700
FL	Dade	(-all-)	2	40100
FL	Orange	(-all-)	2	36600
TX	Harris	(-all-)	2	44700
FL	(-all-)	(-all-)	4	38530
TX	(-all-)	(-all-)	3	39750
(-all-)	(-all-)	(-all-)	7	38816
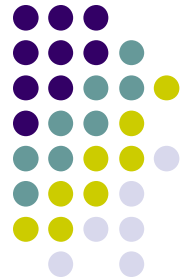
The CUBE

Effect of sex and birth date on income. Four possible ways to group:

1. by sex and birthdate,
2. by sex only
3. By birthdate only
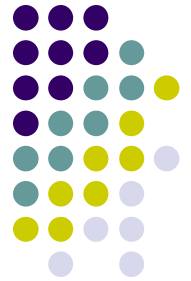4. By nothing—the table becomes one big group

```
SELECT sex,
       year(brirthdate) AS birth_year
       max(income) AS max_income
FROM census

GROUP BY CUBE(sex, year(birthdate));
```

Result of Cubing on Sex and Birth Year

SEX	BIRTH_YEAR	MAX_INCOME
F	1955	46700
F	1956	36300
F	1957	26500
F	(null)	44700
M	1955	32100
M	1956	42500
M	1957	40200
F	(null)	46700
M	(null)	42500
(null)	1955	46700
(null)	1956	42500
(null)	1957	40200
(null)	(null)	44700
(null)	(null)	46700

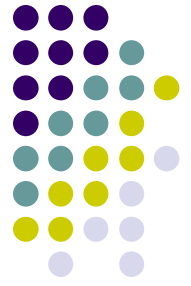# A 3-D Cube: size and avg income for groups of size ≥ 4

```
SELECT CASE grouping(state)
        WHEN 1 THEN '(-all)' ELSE state END AS state,
  CASE grouping(sex)
        WHEN 1 THEN '(-all)' ELSE sex END AS sex,
  CASE grouping(year(brirthdate))
        WHEN 1 THEN '(-all)'
        ELSE char(year(brirthdate)) END AS birth_date
count(*) AS count,  avg(income) AS avg_income

FROM census
GROUP BY CUBE(state, sex, year(birthdate))
HAVING count(*) >= 4;
```

Size and Income for groups of size ≥4:result

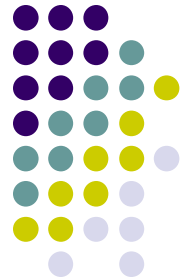| STATE	SEX	BIRTH_YEAR	COUNT	AVG_INCOME
FL	M	1957	4	36075
FL	F	(-all-)	4	38350
FL	M	(-all-)	6	34333
TX	M	(-all-)	5	34620
FL	(-all-)	1957	5	34160
FL	(-all-)	(-all-)	10	35940
TX	(-all-)	(-all-)	8	36085
(-all-)	F	1955	4	41800
(-all-)	M	1956	4	35040
(-all-)	M	1957	5	38816
(-all-)	F	(-all-)	7	34463
(-all-)	M	(-all-)	11	37200
(-all-)	(-all-)	1955	6	35920
(-all-)	(-all-)	1956	5	35920
(-all-)	(-all-)	1957	6	33616
(-all-)	(-all-)	(-all-)	18	36000

Grouping Sets

```sql
SELECT
CASE grouping(state)
    WHEN 1 THEN '(-all-)' ELSE state END AS state,
 CASE grouping(sex)
        WHEN 1 THEN '(-all-)' ELSE sex END AS sex,
 CASE grouping(year(birthdate))
        WHEN 1 THEN '(-all-)'
        ELSE char(year(birthdate)) END AS birth_year,
        count(*) AS count,
        avg(income) AS avg_income
FROM census
GROUP BY
  GROUPING SETS (state, sex), year(birthdate), ();
```

Result of Grouping Sets

STATE	SEX	BIRTH_YEAR	COUNT	AVG_INCOME
FL	F	(-all-)	4	38350
FL	M	(-all-)	6	34333
TX	F	(-all-)	3	39750
TX	M	(-all-)	5	34620
(-all-)	(-all-)	1955	6	37200
(-all-)	(-all-)	1956	6	35920
(-all-)	(-all-)	1957	6	33616
(-all-)	(-all-)	(null)	1	44704
(-all-)	(-all-)	(-all-)	18	36000

Multiple Grouping Specifications

```
SELECT
    CASE grouping(state)
    WHEN 1 THEN '(-all-)' ELSE state END AS state,
    CASE grouping(county)
    WHEN 1 THEN '(-all-)' ELSE county END AS county,
     sex,
     count(*) AS pop,
          avg(income) AS avg_income

FROM census
GROUP BY ROLLUP(state, county), sex;
```

Result of Multiple Grouping

STATE	COUNTY	SEX	POP	AVG_INCOME
FL	Dade	F	2	40100
FL	Dade	M	4	33850
FL	Orange	F	2	36600
FL	Orange	M	2	35300
FL	(-all-)	F	4	38350
FL	(-all-)	M	6	34333
TX	Harris	F	2	44700
TX	Harris	M	3	30733
TX	Travis	F	1	34800
TX	Travis	M	2	40450
TX	(-all-)	F	3	39750
TX	(-all-)	M	5	34620
(-all-)	(-all-)	F	7	38816
(-all-)	(-all-)	M	11	34463