UCLA

CS143
Relational Algebra

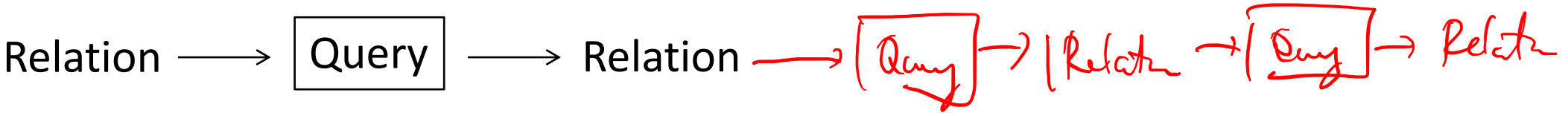
Professor Junghoo “John” Cho

Database Query Language

- What is a “query”?
  - OED: a question, especially one addressed to an official or organization
  - Database jargon for a question
  - Question to get an answer from a database
    - “Who takes a CS class, but no Physics class?”
- Some queries are easy to pose, and some are not

Relational Query Languages

- Formal: relational algebra, relational calculus, datalog
- Practical: SQL, Quel, QBE
- Both input to and output from a query are relations: makes “piping” possible



- Set semantics: no duplicate tuples. Duplicates are automatically eliminated
  - Multiset semantics for SQL for performance reasons. More on this later

Example Database: School Information

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

Q1: All students

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

Student

Q2: Students with age < 18

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$\sigma_{age<18}(Student)$

Q3: Students with GPA > 3.7 and age < 18

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$$\sigma_{GPA>3.7}(\sigma_{age<18}(Student)) = \sigma_{GPA>3.7 \wedge age<18}(Student)$$

Select Operator $\sigma_C(R)$

- Filters out rows in a relation
- *C*: filtering condition as a boolean expression
- *R* can be either a relation or a result from another operator

Q4: sid and GPA of all students

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

project

$\Pi_{sid, GPA}(Student)$

Q5: All departments offering a class

Student(<u>sid</u>, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(<u>dept</u>, <u>cnum</u>, <u>sec</u>, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$\Pi_{dept}(Class)$

$\Pi_{dept,cnum,sec}(Enroll)$

Project Operator $\pi_A(R)$

- Filters out columns in a relation
- *A*: the set of attributes to *keep*

Q6: sid and GPA of students with age < 18

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$$\Pi_{sid,GPA}(\sigma_{age<18}(Student))$$

Questions

- Is it every useful to compose two projection operators next to each other?

$\Pi_{name}(\Pi_{sid, GPA, name}(Student))$ $\Pi_{name}(Student)$

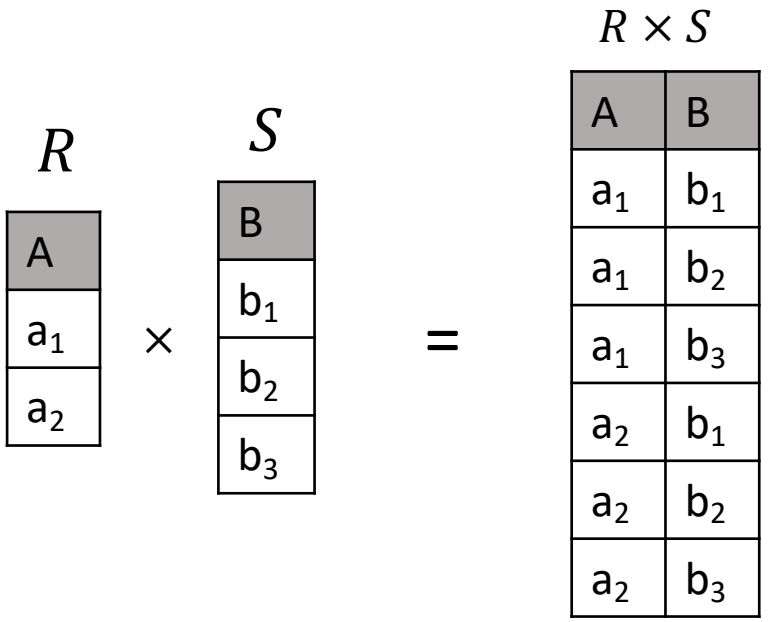
- Is it ever useful to compose to selection operators next to each other?

$\sigma_{GPA>3}(\sigma_{age<18}(Student))$

$= \sigma_{GPA>3 \wedge age<18}(Student)$

Cross Product (Cartesian Product) Operator

- Example: $R \times S$



- Concatenate tuples from two relations horizontally
- Create one output per every pair of input tuples
- If column names conflict, prefix with the table name

Cross Product $R \times S$

- Definition: $R \times S = \{ t \mid t = (r, s)$ for $r \in R$ and $s \in S\}$
- Q: Concatenating two unrelated tables looks odd. Why use it?

Q7: Names of students who take CS classes

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$= \pi_{name}(\sigma_{Student.sid = Enroll.sid \wedge dept='CS'}(Student \times Enroll))$

$\pi_{name}(\sigma_{Student.sid = Enroll.sid}(Student \times \pi_{sid}(\sigma_{dept='CS'}(Enroll))))$

{301, 303, 401}

Question

- If $|R| = r$ and $|S|$ =s, what is $|R \times S|$?

$r \times s$

$(R \times S) \times T$

$= R \times (S \times T)$

$R \times S \times T$

Natural Join Operator ⋈

- Join two tables “naturally”

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

⋈

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

=

sid	name	addr	age	GPA	dept	cnum	sec
301	John	183 Westwood	19	2.1	CS	112	01
301	John	183 Westwood	19	2.1	CS	143	01
303	Elaine	301 Wilshire	17	3.9	EE	143	01
303	Elaine	301 Wilshire	17	3.9	CS	112	01
401	James	183 Westwood	17	3.5	CS	112	01

Natural Join Operator ⋈

- Notation: $R \bowtie S$
  - Concatenate tuples horizontally
  - Enforce equality condition on *all common attributes*
  - Only one copy of the common attributes are kept in the result
  - Most “natural” way to join two tables

Q8: Names of students who take CS classes

Student(<u>sid</u>, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(<u>dept</u>, <u>cnum</u>, <u>sec</u>, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$\Pi_{name}(Student \bowtie \sigma_{dept='CS'}(Enroll))$

Q9: Names of students who take classes offered by “Dick Muntz”

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$$\Pi_{name}(Student \bowtie (Enroll \bowtie \sigma_{instructor='DM'}(Class)))$$

Q10: Names of student pairs who live at the same address

Student(sid, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(dept, cnum, sec, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Rename Operator $\rho_S(R)$

- $\rho_S(R)$: Rename *R* to *S*
- $\rho_{S(A1,A2)}(R)$: Rename *R* to *S(A1, A2)* including attribute names

$\rho_{Student(sid, sname, addr, age, GPA)}(Student)$

Q11: All students and instructors' names

Student(<u>sid</u>, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(<u>dept</u>, <u>cnum</u>, <u>sec</u>, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

$\rho_{Person(name)}(\Pi_{name}(Student)) \cup \rho_{Person(name)}(\Pi_{instructor}(Class))$

Union Operator ∪

- $R \cup S$: Union of tuples from *R* and *S*
- The schemas of *R* an *S* should be the *same*
- No duplicate tuples in the result

Q12: Courses (dept, cnum, sec) that no one takes

Student(<u>sid</u>, name, addr, age, GPA)

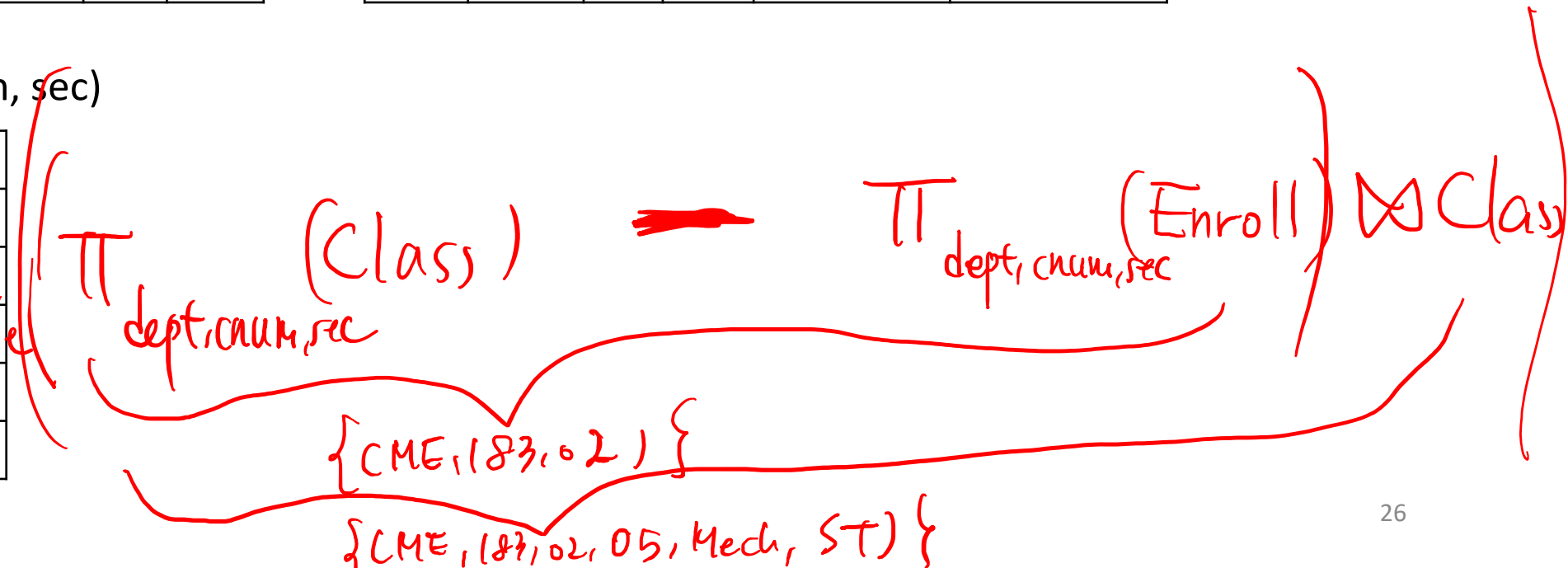
sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(<u>dept</u>, <u>cnum</u>, <u>sec</u>, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
~~CS~~	~~112~~	01	03	Modeling	Dick Muntz
~~CS~~	~~143~~	01	04	DB Systems	John Cho
~~EE~~	~~143~~	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01



Set Difference Operator –

- $R - S$: Tuples in *R* that do not exist in *S*
- The schemas of *R* an *S* should be the *same*
- Q13: What if we want to get the titles of previous courses?

$$\Pi_{title}\left(Class \bowtie \left(\Pi_{dept,cnum,sec}(Class) - \Pi_{dept,cnum,sec}(Enroll)\right)\right)$$

Q14: Instructor names who teach both CS and EE courses

Student(<u>sid</u>, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(<u>dept</u>, <u>cnum</u>, <u>sec</u>, unit, title, instructor)

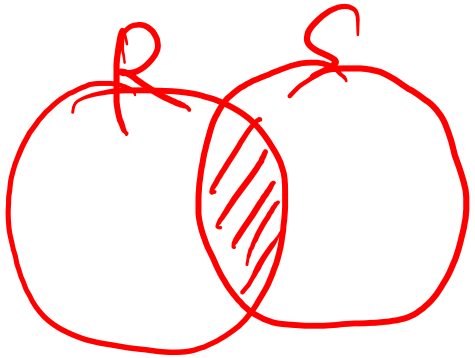
dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

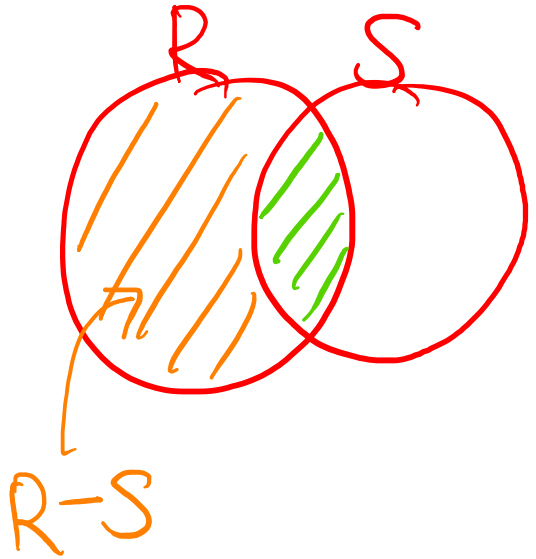
$$\pi_{inst}(\sigma_{dept='CS'}(Class)) \bowtie \pi_{inst}(\sigma_{dept='EE'}(Class))$$

- Q: Can we answer this query without using intersection?

Intersect Operator ∩

- $R \cap S$: Tuples that exist in both *R* and *S*
- The schemas of *R* and *S* should be the *same*
- $R \cap S = R - (R - S)$





Q15: Sids of students who did not take any CS class

Student(<u>sid</u>, name, addr, age, GPA)

sid	name	addr	age	GPA
301	John	183 Westwood	19	2.1
303	Elaine	301 Wilshire	17	3.9
401	James	183 Westwood	17	3.5
208	Esther	421 Wilshire	20	3.1

Class(<u>dept</u>, <u>cnum</u>, <u>sec</u>, unit, title, instructor)

dept	cnum	sec	unit	title	instructor
CS	112	01	03	Modeling	Dick Muntz
CS	143	01	04	DB Systems	John Cho
EE	143	01	03	Signal	Dick Muntz
ME	183	02	05	Mechanics	Susan Tracey

Enroll(sid, dept, cnum, sec)

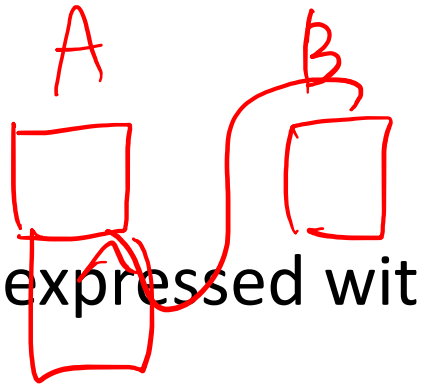
sid	dept	cnum	sec
301	CS	112	01
301	CS	143	01
303	EE	143	01
303	CS	112	01
401	CS	112	01

$$\Pi_{sid}(Student) - \Pi_{sid}(\sigma_{dept='CS'}(Enroll))$$

Advice: When a query is difficult to write, think of its complement!

Core Relational Operators

- σ, π, $\times$, $\bowtie$, ρ, $\cup$, $\cap$, $-$
- Q: which ones are "core" and which ones can be expressed with others?



R S

{a, b, c} {a}

d b

R - S = {b, c}

{b, c, d}

{c}

Summary

- Relational algebra: Formal query language for relational model
  - Theoretical foundation behind SQL
- Both inputs and outputs are relations: “piping” is possible

- Set semantics: duplicates are automatically eliminated
- Operators learned: σ, π, $\times$, $\bowtie$, ρ, $\cup$, $\cap$, $-$
- Suggestion: If a query is difficult to write, think of its complement!!