Relational division – how derive

- Given relations R and Q with attribute sets C(R) and C(Q),
  - Attribute set gives names and domains
- Such that C(Q) proper subset of C(R)
- R/Q is a relation with attribute set C(R/C) = C(R) - C(Q)
- A tuple is in R/Q exactly when combining (concatenating) it with *every* tuple in Q yields a tuple in R
  - R/Q is a subset of $\pi_{C(R) - C(Q)}(R)$
- R/Q is expressed with basic relational operations as

$\pi_{C(R) - C(Q)}(R) - \pi_{C(R) - C(Q)}((\pi_{C(R) - C(Q)}(R) \times Q) - R)$

Huh?

Subtract $\pi_{C(R) - C(Q)}((\pi_{C(R) - C(Q)}(R) \times Q) - R)$ from $\pi_{C(R) - C(Q)}(R)$?

- Let $(d_1, \ldots, d_m) \in \pi_{C(R) - C(Q)}(R)$
- Let $(q_1, \ldots, q_n) \in Q$
- $(d_1, \ldots, d_m, q_1, \ldots, q_n) \in (\pi_{C(R) - C(Q)}(R) \times Q)$ may or may not be in R
- If $(d_1, \ldots, d_m, q_1, \ldots, q_n)$ not in R, then $(d_1, \ldots, d_m)$ not in R/Q

  => If $(d_1, \ldots, d_m, q_1, \ldots, q_n)$ in $((\pi_{C(R) - C(Q)}(R) \times Q) - R)$ then $(d_1, \ldots, d_m)$ not in R/Q

  => Correct to subtract

  $(d_1, \ldots, d_m) \in \pi_{C(R) - C(Q)}((\pi_{C(R) - C(Q)}(R) \times Q) - R)$ from $\pi_{C(R) - C(Q)}(R)$

Subtract $\pi_{C(R) - C(Q)}((\pi_{C(R) - C(Q)}(R) \times Q) - R)$ from $\pi_{C(R) - C(Q)}(R)$?

- Let $(d_1, \ldots, d_m) \in \pi_{C(R) - C(Q)}(R)$
- Let $(q_1, \ldots, q_n) \in Q$

But have we subtracted enough from $\pi_{C(R) - C(Q)}(R)$?

- If $(d_1, \ldots, d_m)$ not in R/Q, then there is some $(q_1, \ldots, q_n) \in Q$ such that $(d_1, \ldots, d_m, q_1, \ldots, q_n)$ not in R

  $\Rightarrow (d_1, \ldots, d_m, q_1, \ldots, q_n)$ in $((\pi_{C(R) - C(Q)}(R) \times Q) - R)$

  $\Rightarrow (d_1, \ldots, d_m)$ in $\pi_{C(R) - C(Q)}((\pi_{C(R) - C(Q)}(R) \times Q) - R)$

  Yes, we have subtracted all that is needed

Note that $\pi_{C(Q)}(R)$ may contain elements not in Q
Not affect result.

Board examples

Formal definition

- A relational expression is
  - A relation R in the database
  - A constant relation
  - For any relational expressions E_1 and E_2
    - $E_1 \cup E_2$
    - $E_1 - E_2$
    - $E_1 \times E_2$
    - $\sigma_P(E_1)$ for predicate P on attributes of E_1
    - $\pi_S(E_1)$ where S is a subset of attributes of E_1
    - $\rho(Q(L), E_1)$ where Q is a new relation name and L is a list of (old name → new name) mappings of attributes of E_1
- A query in the relational algebra is a relational expression

Summary

- Relational algebra operations provide foundation of query languages for database systems
- Derived operations, especially joins, simplify expressing queries
- Formal algebraic definition allow for provably correct simplifications, optimizations for query evaluation