https://www.youtube.com/watch?v=uAXtO5dMqEI
Do not try and bend the spoon. That's impossible. Instead... only try to realize the truth.
What truth?
There is no spoon.
There is no spoon?
Then you'll see, that it is not the spoon that bends, it is only yourself.

Today

Data Modeling in RDBMS

Real World Entity

Conceptual Level | Entity-Relationship Model (E/R) Level

Conceptual Level | Logical Level | <u>R</u>elational Model

Conceptual Level | Logical Level | Physical Level | SQL

Conceptual Level | Logical Level | Computable Entity

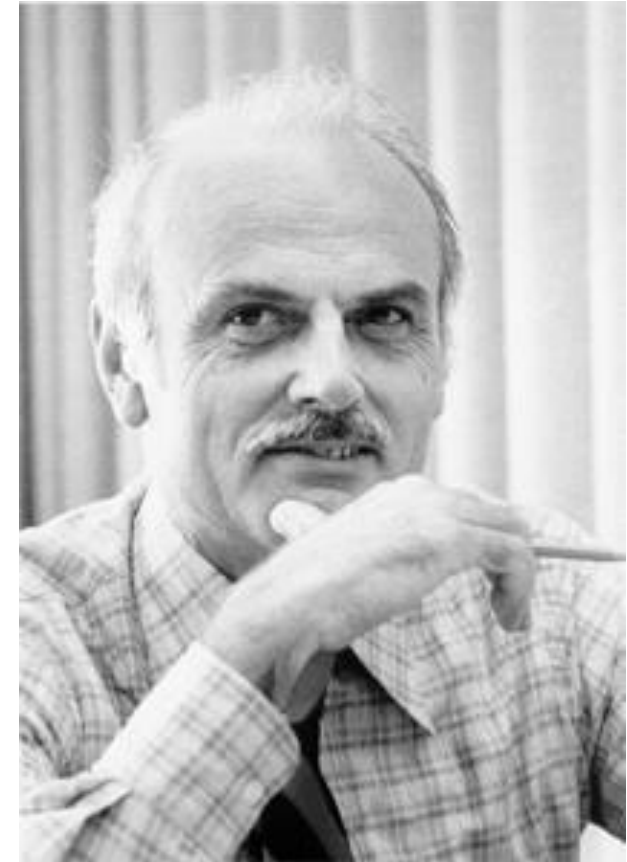Welcome | Relational Algebra | Unary Op | Binary Op

Relational   3

Edgar Frank "Ted" Codd, IBM, 1969, 1970

Information Retrieval

A Relational Model of Data for Large Shared Data Banks

E. F. CODD
IBM Research Laboratory, San Jose, California

Relational Model × Algebra

Given a set, defining operations on elements of the set!

Given Z={integers}={..., -2, -1, 0, 1, 2, ...}
Operators & Operands:
Unary: -(2), 2!
Binary: 2+3, 2*3, 2^3
Closure:
Result is also an element of the set

Discrete Mathematics

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations

Query in Natural Language → Query in Math Formula

Relational Model × Algebra　　6

Who made **'Pulp Fiction'**?

π (σ (Movie × MovieDirector × Director))

Director.FirstName　　Movie.Title='PulpFiction'　　　AND
Director.LastName　　Movie.Id=MovieDirctor.MovieId　　AND
　　　　　　　　　　Director.Id=MovieDirctor.DirectorId

Relational Model × Algebra

Who acted in 'Pulp Fiction'?

π (σ (Movie × StarIn × Actor))

Actor.FirstName   Movie.Title='PulpFiction' AND

Actor.LastName   Movie.Id=StarIn.MovieId AND

         Actor.Id=StarIn.ActorId

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations

Unary Operation	Binary Operation (Set Theory)
π(R) : Project	R1 ∪ R2: Union
σ(R) : Select	R1 ∩ R2: Intersection
ρ(R) : Rename	R1 \ R2 : Set Difference
	R1 × R2: Cartesian Product

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

π(R) : Project

σ(R) : Select

ρ(R) : Rename

Binary Operation (Set Theory)

R1 ∪ R2: Union

R1 ∩ R2: Intersection

R1 \ R2 : Set Difference

R1 × R2: Cartesian Product

Algebra × Project (π)

π, pi, is used to select a subset of <u>attributes</u> (columns) from a relation

$$A = \pi_{<\text{attribute list}>}(R)$$

Vertical Filtering

R is a relation

<attribute list> subset of attributes of R

A is a <u>relation</u> including all tuples in R with only attributes in list

Algebra × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

What are directors' name?

$$A = \pi_{FirstName, LastName}(Director)$$

Algebra × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

What are directors' name?

$$A = \pi_{\text{FirstName, LastName}}(\text{Director})$$

Algebra × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

How many movies each director made?

$$A = \pi_{\text{FirstName, LastName, MovieCount}}(\text{Director})$$

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

How many movies each director made?

$$A = \pi_{\text{FirstName, LastName, MovieCount}}(\text{Director})$$

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation	Binary Operation (Set Theory)
π(R) : Project	R1 $\cup$ R2: Union
σ(R) : Select	R1 $\cap$ R2: Intersection
ρ(R) : Rename	R1 \ R2 : Set Difference
	R1 $\times$ R2: Cartesian Product

Algebra × Selection (σ)

σ, sigma, is used to select a subset of <u>tuples</u> from a relation based on a <u>condition</u> (θ) over relation's attributes.

$$A = \sigma_{\theta} (R)$$

Horizontal Filtering

R is a relation
θ is a Boolean expression on the attributes of R
A is a <u>relation</u> including tuples that make θ true

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which director was born in US?

$$A = \sigma_{\text{PlaceOfBirth='USA'}} (\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which director was born in US?

$$A = \sigma_{PlaceOfBirth='USA'} (Director)$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which director make more than 20 movies?

$$A = \sigma_{MovieCount > 20} (Director)$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which director make more than 20 movies?

$$A = \sigma_{MovieCount > 20} (Director)$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which director has same first and last names?

$$A = \sigma_{\text{FirstName = LastName}} (\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which director has same first and last names?

$$A = \sigma_{\text{FirstName = LastName}} \text{(Director)}$$

$$= \emptyset$$

Algebra × Selection (σ)

θ can be made up of number of Boolean clauses

$A = \sigma_{\theta} (R)$

=
<>
>
>=
<
<=

Attribute

Constant
Attribute

AND
OR
NOT

Not Equal

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which <u>American</u> director made more than 20 movies?

$$A = \sigma_{\text{PlaceOfBirth}='USA'\ \text{AND}\ \text{MovieCount} > 20}(\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which American director made more than 20 movies or is not American?

| Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which American director made more than 20 movies or is not American?

$A1 = \sigma_{\text{PlaceOfBirth}='USA' \text{ AND } (\text{MovieCount} > 20 \text{ OR } \text{PlaceOfBirth}<>'USA')} (\text{Director})$

❌

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which American director made more than 20 movies or is not American?

$$A2 = \sigma_{(PlaceOfBirth='USA' \text{ AND } MovieCount > 20) \text{ OR } (PlaceOfBirth<>'USA')}(Director)$$

Algebra × Selection (σ)

Boolean Gates

Digital Systems
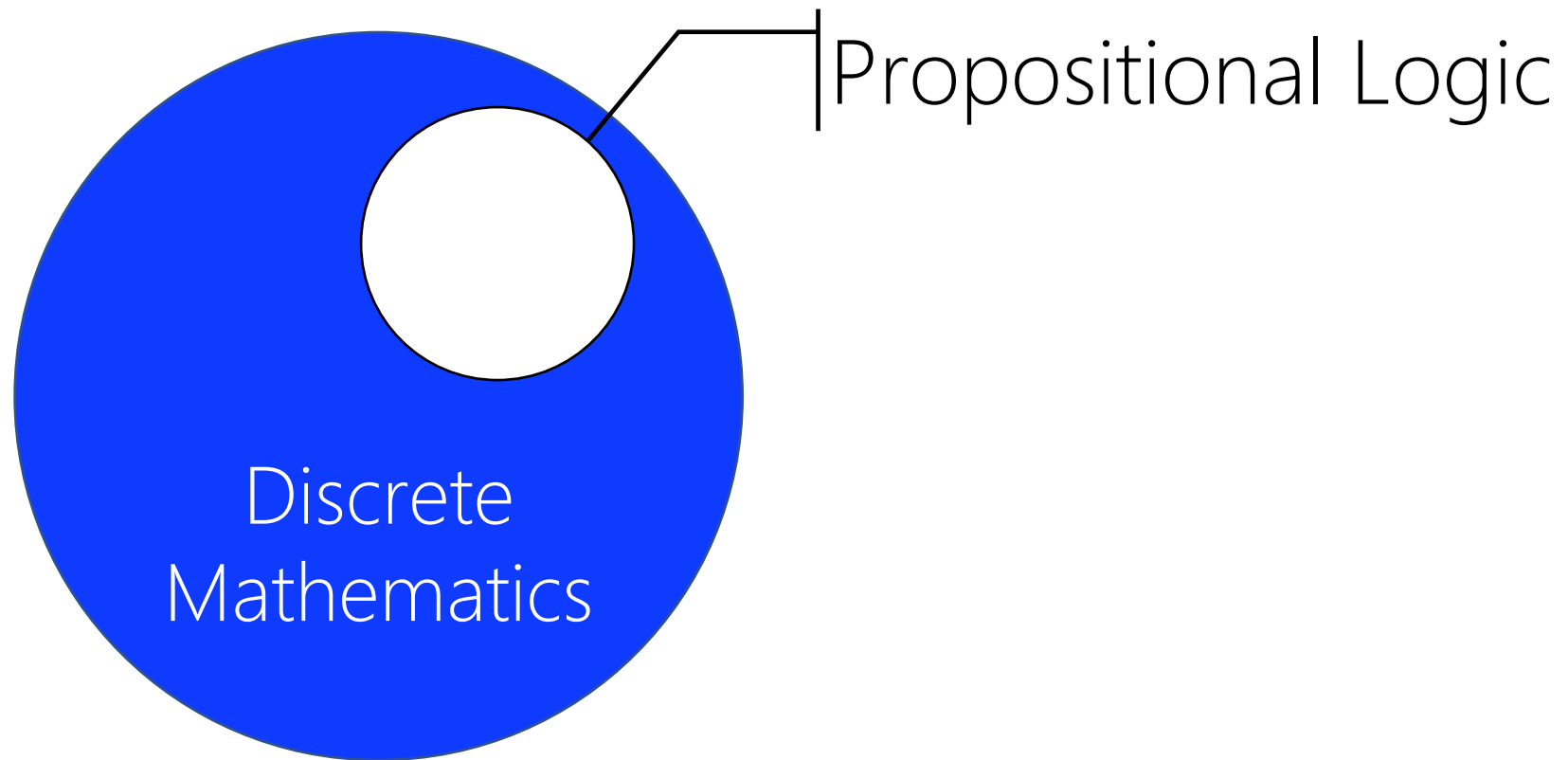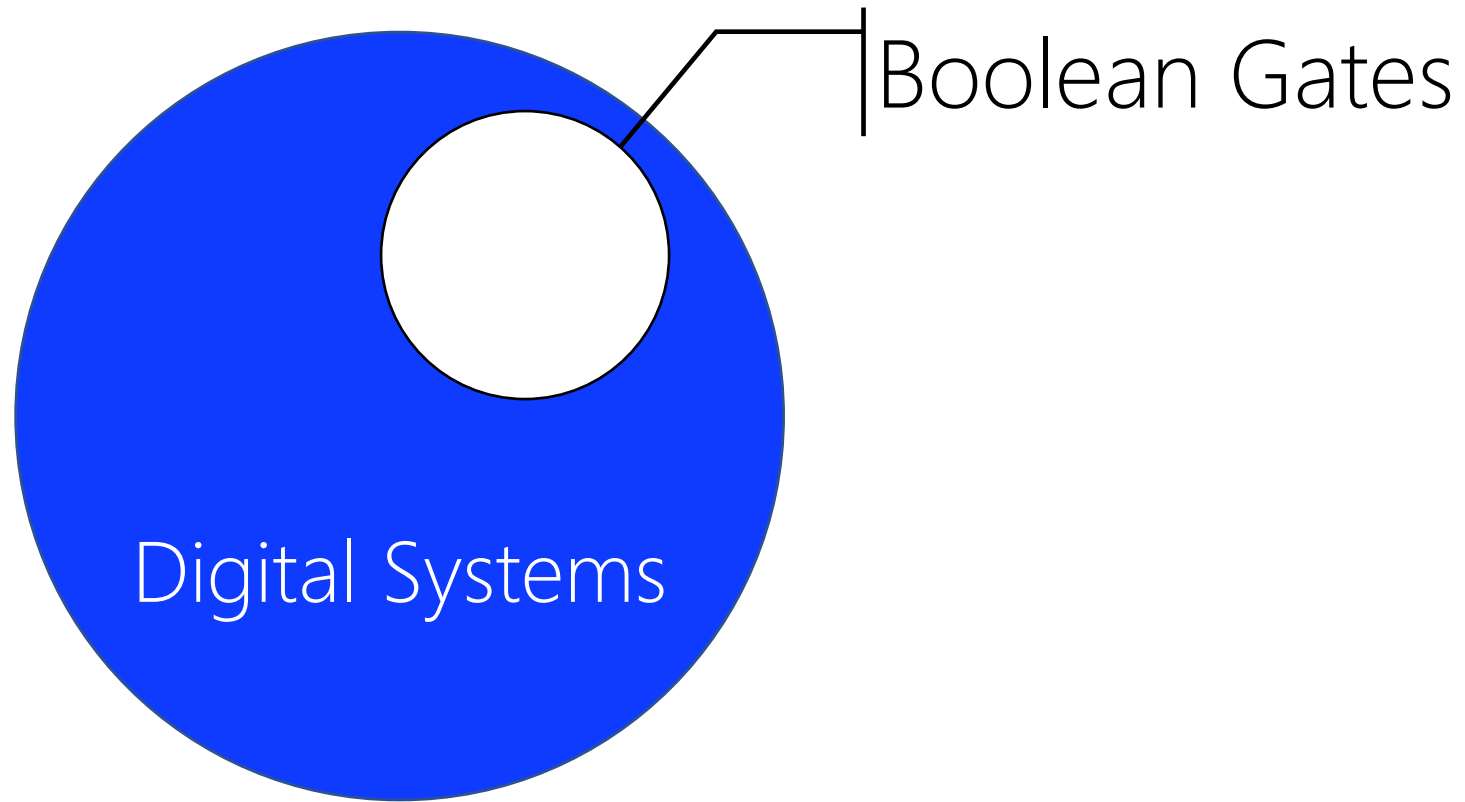
Algebra × Selection (σ)

Operation Precedence for Logical Operations
  () > NOT > AND = OR

Commutative Law
  A AND B = B AND A *(we say A commutes with B under AND)*
  A OR B = B OR A *(we say A commutes with B under OR)*

Associative Law
  A AND (B AND C) = (A AND B) AND C = A AND B AND C
  A OR (B OR C) = (A OR B) OR C = A OR B OR C

Distributive Law
  A AND (B OR C) = (A AND B) OR (A AND C)
  A OR (B AND C) = (A OR B) AND (A OR C)

de Morgan´s Theorem**
  NOT (A AND B) = NOT(A) OR NOT(B)
  NOT (A OR B) = NOT(A) AND NOT(B)

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which <u>American</u> director made more than 20 movies?

$$A = \sigma_{\text{PlaceOfBirth='USA' AND MovieCount > 20}} (\text{Director})$$

$$= \sigma_{\text{MovieCount > 20 AND PlaceOfBirth='USA'}} (\text{Director})$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which American director made more than 20 movies?

$$A1 = \sigma_{PlaceOfBirth='USA'} (Director)$$

$$A = \sigma_{MovieCount > 20} (A1)$$

Algebra × Selection (σ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which American director made more than 20 movies?

$$A = \sigma_{\text{MovieCount} > 20} \ (\sigma_{\text{PlaceOfBirth}='USA'} (\text{Director}))$$

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Which American director made more than 20 movies?

$$A = \sigma_{\text{PlaceOfBirth}='\text{USA}'}(\sigma_{\text{MovieCount} > 20}(\text{Director}))$$

$$= \sigma_{\text{PlaceOfBirth}='\text{USA}' \text{ AND } \text{MovieCount} > 20}(\text{Director})$$

$$= \sigma_{\text{MovieCount} > 20 \text{ AND } \text{PlaceOfBirth}='\text{USA}'}(\text{Director})$$

$$= \sigma_{\text{MovieCount} > 20}(\sigma_{\text{PlaceOfBirth}='\text{USA}'}(\text{Director}))$$

$$\sigma_\theta \left(\sigma_{\theta'}(R) \right)$$

$$=$$

$$\sigma_{\theta \text{ AND } \theta'}(R)$$

$$=$$

$$\sigma_{\theta' \text{ AND } \theta}(R)$$

$$=$$

$$\sigma_{\theta'} \left(\sigma_\theta (R) \right)$$

Commutative Law

Corollary: $\sigma_\theta \left(\sigma_{\theta'} \left(\sigma_{\theta''} \ldots \left(\sigma_{\theta'''}(R) \right) \right) = \sigma_{\theta \text{ AND } \theta' \text{ AND } \theta'' \ldots \text{ AND } \theta'''}(R)$

Algebra × Selection (σ) × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Find directors' <u>name</u> who make more than 20 movies?

Algebra × Selection (σ) × Project (π)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Find directors' <u>name</u> who make more than 20 movies?

$$A1 = \sigma_{MovieCount > 20} (Director)$$

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Find directors' <u>name</u> who make more than 20 movies?

$$A1 = \sigma_{MovieCount > 20}(Director)$$

$$A = \pi_{FirstName, LastName}(A1)$$

$$= \pi_{FirstName, LastName}(\sigma_{MovieCount > 20}(Director))$$

Algebra × Selection (σ) × Project (π) 40

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Find directors' <u>name</u> who make more than 20 movies?

❌ $A = \sigma_{MovieCount > 20} (\pi_{FirstName, LastName} (Director))$

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

$\pi(R)$: Project

$\sigma(R)$: Select

$\rho(R)$: Rename

Binary Operation (Set Theory)

R1 ∪ R2: Union

R1 ∩ R2: Intersection

R1 \ R2 : Set Difference

R1 × R2: Cartesian Product

Algebra × Rename (ρ)

ρ, rho, is used to rename a relation or its attributes or both

$$A = \rho_{<R'(a'/a,\ b'/b,\ ...)>}(R)$$

R is a relation
R' is the new name for R(a, b, ...)
a' is the new name for attribute a of R
b' is the new name for attribute b of R

...

Algebra × Rename (ρ)

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

$$A = \pi_{\text{FirstName, Name}} \left(\rho_{\text{ActiveDirector(Name/LastName)}} \left(\sigma_{\text{MovieCount > 20}}(\text{Director}) \right) \right)$$

We'll c its real use ☺

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

$\pi(R)$: Project

$\sigma(R)$: Select

$\rho(R)$: Rename

Binary Operation (Set Theory)

$R1 \cup R2$: Union

$R1 \cap R2$: Intersection

$R1 \setminus R2$: Set Difference

$R1 \times R2$: Cartesian Product

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

π(R) : Project

σ(R) : Select

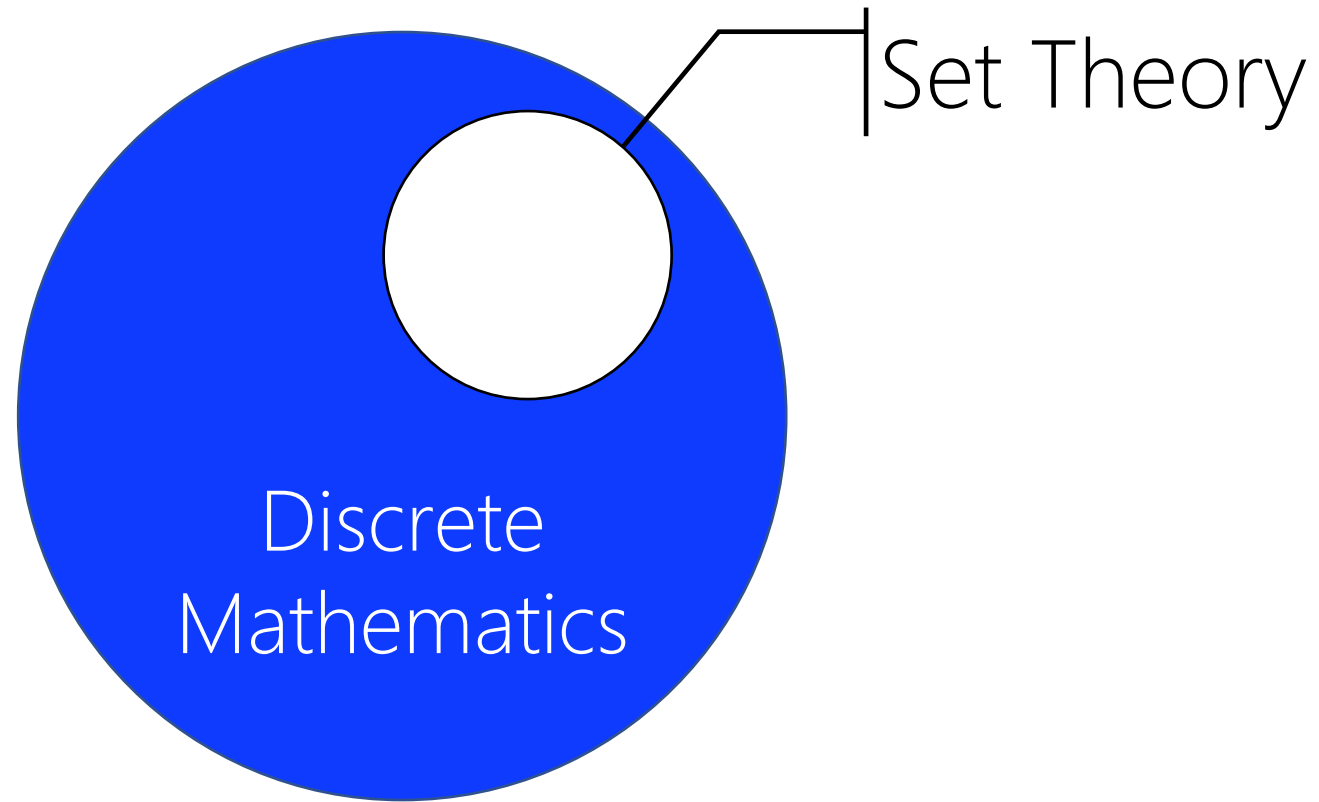ρ(R) : Rename

Binary Operation (Set Theory)

R1 ∪ R2: Union

R1 ∩ R2: Intersection

R1 \ R2 : Set Difference

R1 × R2: Cartesian Product

Algebra × Union (∪)

∪, union, is used to include tuples exist in <u>either</u> relations

$$A = R1 \cup R2$$

R1 and R2 are relations
A has all tuples from R1 and R2

<u>No duplicate</u>!
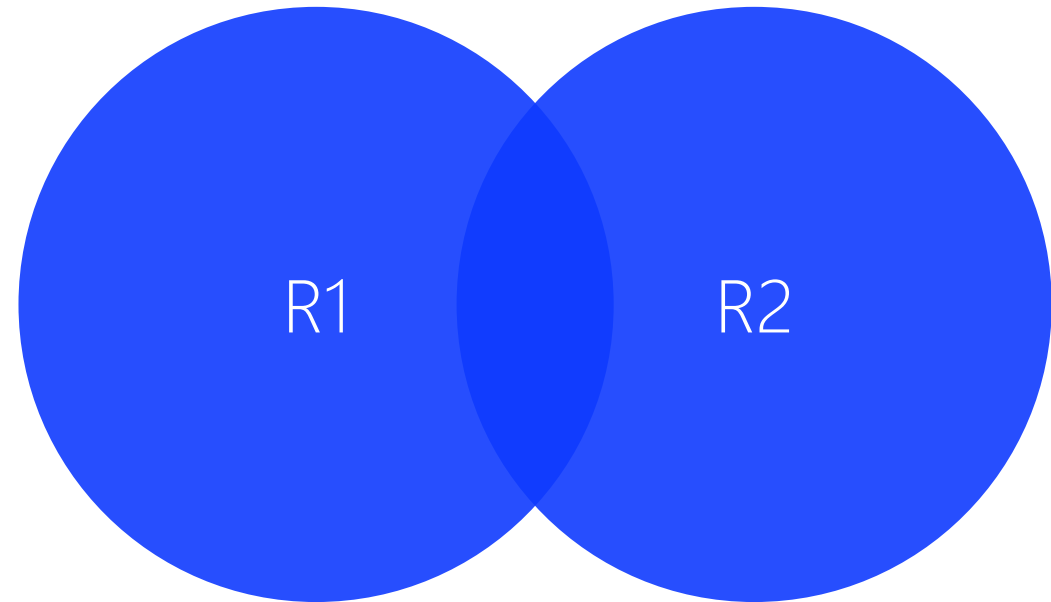
R1   R2

Venn Diagram

Algebra × Union (∪)

∪, union, is used to include tuples exist in <u>either</u> relations

$$A = R1 \cup R2$$

Commutative Law
   $R1 \cup R2 = R2 \cup R1$

Associative Law
   $R1 \cup (R2 \cup R3) = (R1 \cup R2) \cup R3 = R1 \cup R2 \cup R3$

Algebra × Union (∪)   49

Director					
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor					
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

Show all American cast and crews' name?

Algebra × Union (∪)   50

Director					
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor					
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

Actor ∪ Director

Algebra × Union (∪)   50

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

Actor ∪ Director

❌ Attribute Incompatibility (Datatype)

Algebra × Union (∪)   50

Director

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor

Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

Actor ∪ Director

❌ Attribute Incompatibility (Semantic)

Algebra × Union (∪)   52

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

$$A = \pi_{FirstName, LastName} (\sigma_{PlaceOfBirth='USA'} (Actor)) \cup (\pi_{FirstName, LastName} (\sigma_{PlaceOfBirth='USA'} (Director))$$

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

π(R) : Project

σ(R) : Select

ρ(R) : Rename

Binary Operation (Set Theory)

R1 $\cup$ R2: Union

R1 $\cap$ R2: Intersection

R1 \ R2 : Set Difference

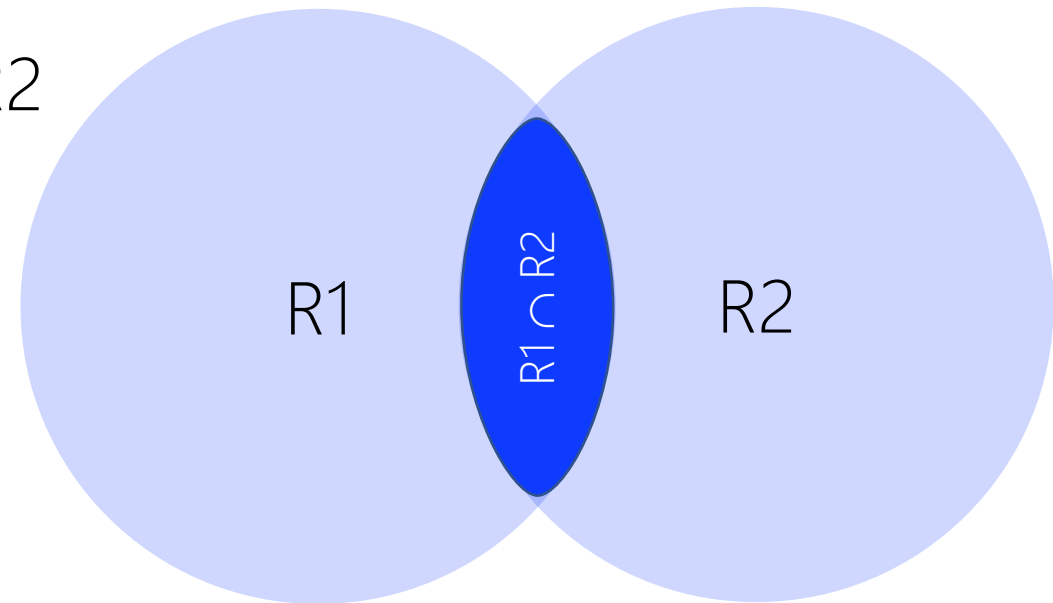R1 $\times$ R2: Cartesian Product

Algebra × Intersection (∩)

∩, intersection, is used to include tuples exist in <u>both</u> relations

$$A = R1 \cap R2$$

R1 and R2 are relations
A has tuples exist both in R1 and R2

Algebra × Intersection (∩)

∩, intersection, is used to include tuples exist in <u>both</u> relations

$$A = R1 \cap R2$$

Commutative Law
$$R1 \cap R2 = R2 \cap R1$$

Associative Law
$$R1 \cap (R2 \cap R3) = (R1 \cap R2) \cap R3 = R1 \cap R2 \cap R3$$

Distributive Law
$$R1 \cap (R2 \cup R3) = (R1 \cap R2) \cup (R1 \cap R3)$$
$$R1 \cup (R2 \cap R3) = (R1 \cup R2) \cap (R1 \cup R3)$$

Algebra × Intersection (∩)

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You …	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w…	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

Which actor has directed a movie?

Algebra × Intersection (∩)

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You …	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w…	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

$$A = \pi_{\text{FirstName, LastName}} (\text{Actor} \cap \text{Director})$$

Algebra ⨯ Intersection (∩)

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You …	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w…	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

$$A = \pi_{\text{FirstName, LastName}} (\textcolor{red}{Actor \cap Director})$$

FirstName	LastName
Stanley	Kubrick
Alfred	Hitchcock
Clint	Eastwood

FirstName	LastName
John	Travolta
Samuel	Jackson
Uma	Thurman
Clint	Eastwood

$$\pi_{\text{FirstName, LastName}}(\text{Director})$$

FirstName	LastName
Stanley	*Kubrick*
Alfred	*Hitchcock*
Clint	*Eastwood*

FirstName	LastName
John	*Travolta*
Samuel	*Jackson*
Uma	*Thurman*
Clint	*Eastwood*

$\pi_{\text{FirstName, LastName}}$ (Actor)

FirstName	LastName
Stanley	Kubrick
Alfred	Hitchcock
Clint	Eastwood

FirstName	LastName
John	Travolta
Samuel	Jackson
Uma	Thurman
Clint	Eastwood

$$A = (\pi_{\text{FirstName, LastName}}(\text{Actor})) \cap (\pi_{\text{FirstName, LastName}}(\text{Director}))$$

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

π(R) : Project

σ(R) : Select

ρ(R) : Rename

Binary Operation (Set Theory)

R1 $\cup$ R2: Union

R1 $\cap$ R2: Intersection
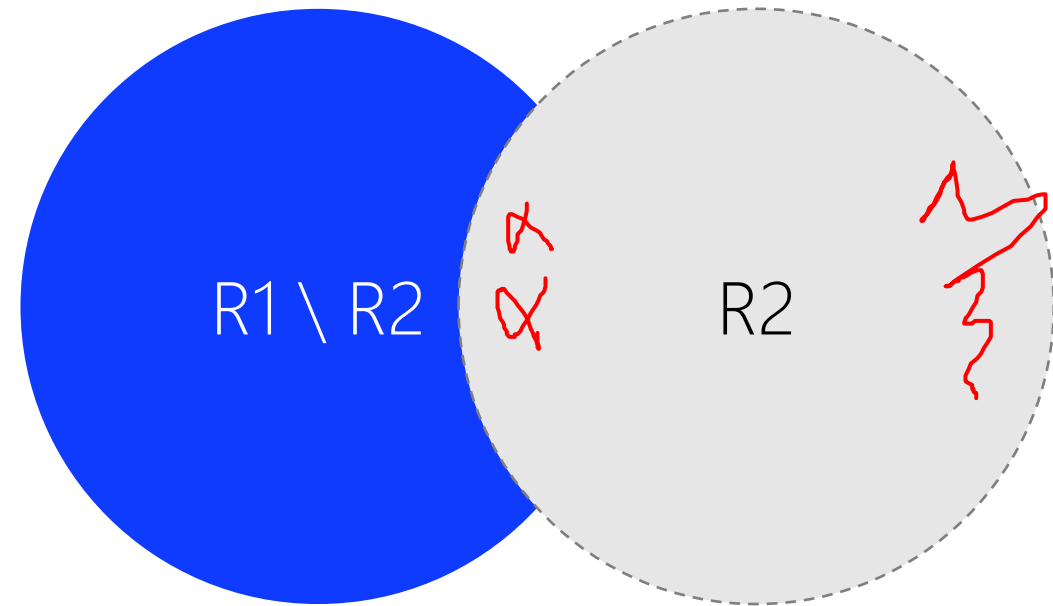
R1 \ R2 : Set Difference

R1 × R2: Cartesian Product

\ , minus, is used to exclude tuples from a relation

$$A = R1 \setminus R2$$

R1 and R2 are relations
A has tuples in R1 but not in R2

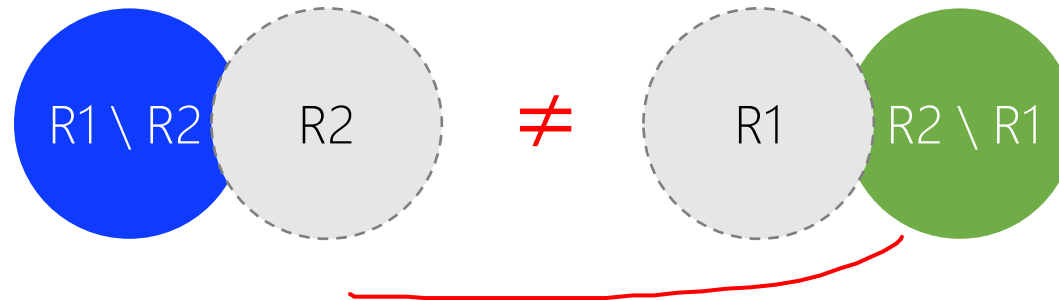# Algebra × Set Difference (\ )

\ , minus, is used to exclude tuples from a relation

$$A = R1 \setminus R2$$

Commutative Law
R1 \ R2 ≠ R2 \ R1

Algebra × Set Difference (\ )

\ , minus, is used to exclude tuples from a relation

$$A = R1 \setminus R2$$

Associative Law

R1 \ (R2 \ R3) ≠ (R1 \ R2) \ R3 (Why?)

Algebra × Set Difference (\ )

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

Which director never appeared in a movie?

Algebra × Set Difference (\ )

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say 'w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ..	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

$$A = \pi_{\text{FirstName, LastName}} (\text{Director} \setminus \text{Actor})$$

Director						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Actor						
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestLine	MovieCount
1	John	Travolta	Feb. 18, 1954	USA	You ...	61
2	Samuel	Jackson	Dec. 21, 1948	USA	Say w...	125
3	Uma	Thurman	Apr. 29, 1970	USA	I believe ...	51
4	Clint	Eastwood	May 31, 1930	USA	A good ..	69

$$A = \pi_{\text{FirstName, LastName}} (\text{Director} \setminus \text{Actor})$$ ❌

Algebra × Set Difference (\ )

FirstName	LastName
Stanley	*Kubrick*
Alfred	*Hitchcock*
Clint	*Eastwood*

FirstName	LastName
John	*Travolta*
Samuel	*Jackson*
Uma	*Thurman*
Clint	*Eastwood*

$\pi_{\text{FirstName, LastName}}$ (Director)

Algebra × Set Difference (\ )

FirstName	LastName
Stanley	Kubrick
Alfred	Hitchcock
Clint	Eastwood

FirstName	LastName
John	Travolta
Samuel	Jackson
Uma	Thurman
Clint	Eastwood

$$\pi_{\text{FirstName, LastName}} (\text{Actor})$$

Algebra × Set Difference (\ )

FirstName	LastName
Stanley	*Kubrick*
Alfred	*Hitchcock*
Clint	*Eastwood*

FirstName	LastName
John	*Travolta*
Samuel	*Jackson*
Uma	*Thurman*
Clint	*Eastwood*

$$A = (\pi_{\text{FirstName, LastName}}(\text{Director})) \setminus (\pi_{\text{FirstName, LastName}}(\text{Actor}))$$

Director					
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Movie			
Id	Title	Language	RunningTime
1	2001: A Space Odyssey	English	142
2	Rosemary's Baby	English	136

Director and Movie do **not** share same

i) Number of attributes

ii) Datatype in attributes

iii) Semantics in attributes

A = Director ∪ Movie

A = Director ∩ Movie

A = Director \ Movie

Algebra × Set Compatibility

Director					
Id	FirstName	LastName	DateOfBirth	PlaceOfBirth	BestMovieId	MovieCount
1	Stanley	Kubrick	Jul. 26, 1928	USA	1	13
2	Alfred	Hitchcock	Aug. 13, 1899	England	203	47
3	Clint	Eastwood	May 31, 1930	USA	803	35

Movie			
Id	Title	Language	RunningTime
1	2001: A Space Odyssey	English	142
2	Rosemary's Baby	English	136

Find movies which are NOT the best movies of any directors?

A = Movie \ Director

BestMovieId
1
203
803

Id
1
2

Find movies which are NOT the best movies of any directors?

$$A = (\pi_{Id}(Movie)) \setminus (\pi_{BestMovieId}(Director))$$

Relational Model × Algebra

Given relational (table) schema filled with actual data instances (rows):
Operations to SELECT Information FROM Relations
Operations to write query

Unary Operation

$\pi(R)$: Project

$\sigma(R)$: Select

$\rho(R)$: Rename

Binary Operation (Set Theory)

R1 ∪ R2: Union

R1 ∩ R2: Intersection

R1 \ R2 : Set Difference

R1 × R2: Cartesian Product