Name 1: Date:
Name 2:

In the IMDB database, we can find all the productions directed by Christopher Nolan using this query:

$\Pi_{id, year}[P \bowtie D \bowtie (\Pi_{pid} \sigma_{firstname='Christopher' \text{ and } lastname='Nolan'} Per)]$

id	year
Interstellar (2014)	2014
The Dark Knight Rises (2012)	2012
Inception (2010)	2010
The Dark Knight (2008)	2008
The Prestige (2006)	2006
The Exec (2006) {{SUSPENDED}}	2006
Batman Begins (2005)	2005
Insomnia (2002)	2002
Memento (2000)	2000
Following (1998)	1998
Doodlebug (1997)	1997

However, when we want to know the rank (rating) for each of these productions we get less tuples (2 productions are missing):

$\Pi_{id, year}[R \bowtie P \bowtie D \bowtie (\Pi_{pid} \sigma_{firstname='Christopher' \text{ and } lastname='Nolan'} Per)]$

id	year	votes	rank
The Dark Knight Rises (2012)	2012	721037	8.6
Inception (2010)	2010	893674	8.8
The Dark Knight (2008)	2008	1098312	9
The Prestige (2006)	2006	528756	8.5
Batman Begins (2005)	2005	634331	8.3
Insomnia (2002)	2002	164363	7.2
Memento (2000)	2000	582937	8.6
Following (1998)	1998	46433	7.6
Doodlebug (1997)	1997	6393	7.1

a) Why is the second query returning less tuples than the first?

Because some productions are not present in the ratings table

b) how do we "fix" our second query so we get the same number of tuples as in the first query?

We "right" join:

$\Pi_{id, year}[R \bowtie_R P \ldots$

↑ right join. We can also use full join.

same as before