Database Query Visualization

An Empirical Study

Zuzel Vera Pacheco – M.Sc. program

Supervisor:
Prof. Gregory V. Wilson

Software Engineering Group
University of Toronto

Scientific Writing Course, 2010

Natural Language

Get the number of wins and the names of all the teams who scored a goal at the 2010 World Cup. Arrange the results in descending order based on the number of wins.

Structured Query Language (SQL)

```
SELECT team, wins FROM world_cups
    WHERE year = 2010 AND goals > 0 ORDER BY wins DESC;
```

Query By Example (QBE)

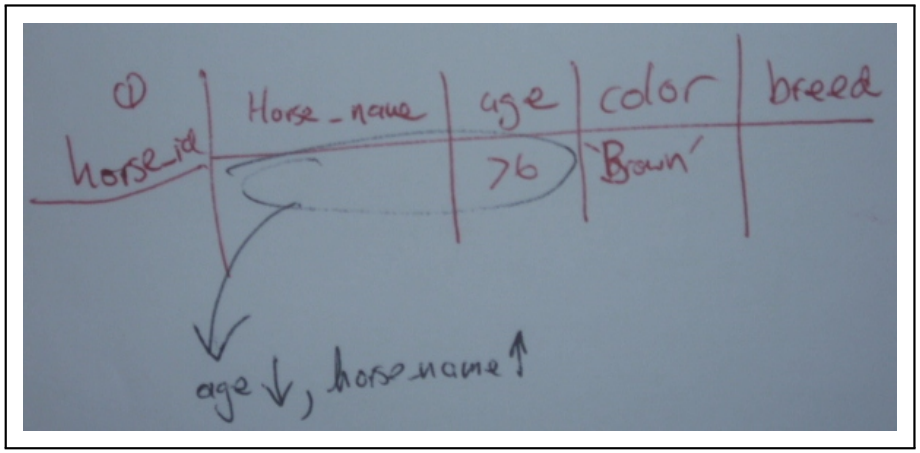
year	team	minutes_played	goals	wins
2010	P.		> 0	P.DO

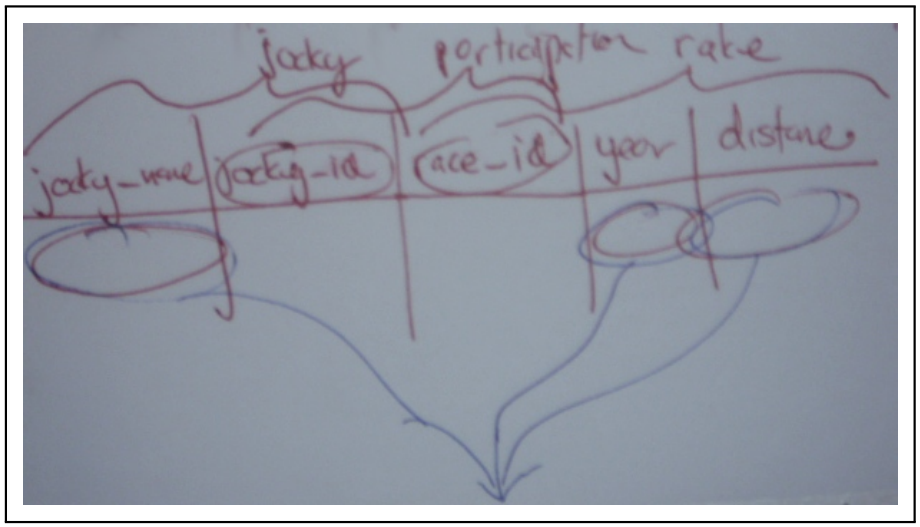
What is the best type of representation for a particular task?

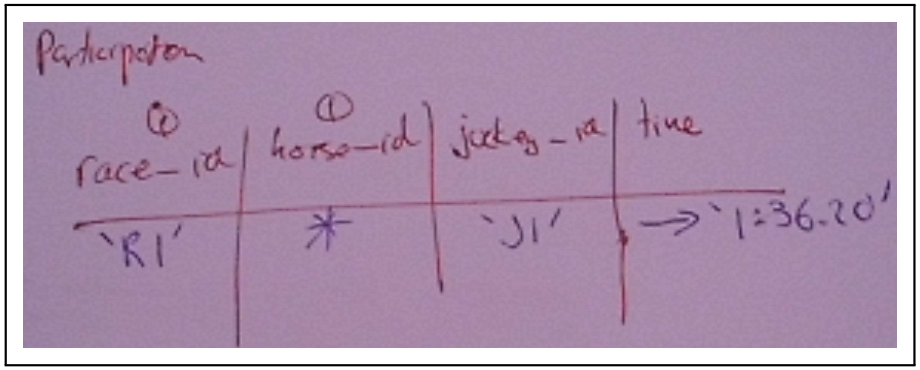
Design of a Diagrammatic Notation

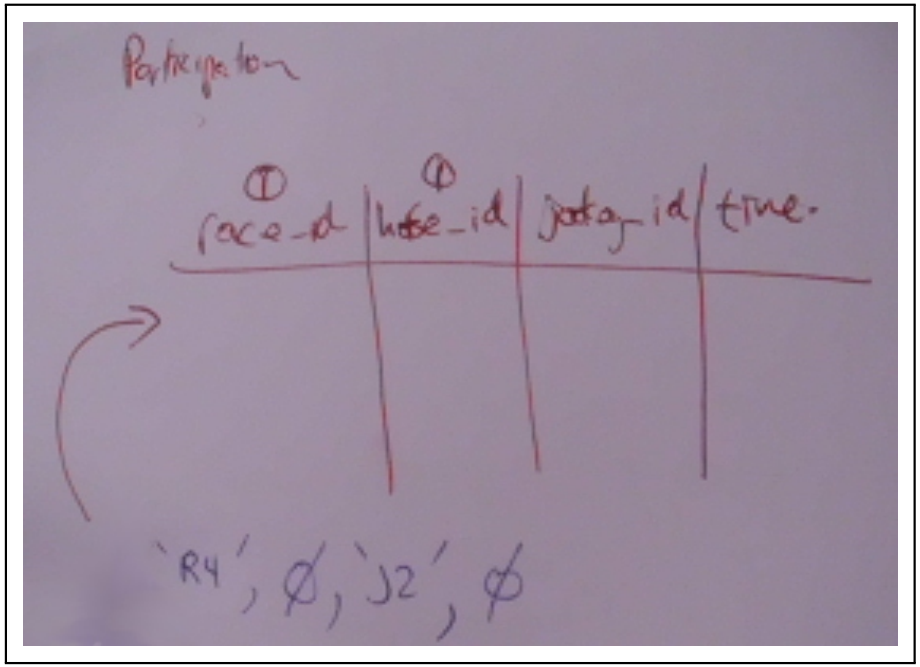
- Programmers use of ad hoc notations instead of standard notations (Cherubini et al., 2007)
- This project: designed and executed an empirical study to determine the common features of ad hoc notations

```
1   SELECT HORSE_NAME, AGE FROM HORSE WHERE COLOR = 'Brown' AND AGE > 6
2     ORDER BY AGE DESC, HORSE_NAME ASC;
3
4
5   DELETE FROM PARTICIPATION WHERE JOCKEY_ID = 'J3' OR RACE_ID = 'R4';
6
7
8   INSERT INTO PARTICIPATION (JOCKEY_ID, RACE_ID) VALUES ('J2', 'R4');
9
10
11  UPDATE PARTICIPATION
12    SET TIME = '1:36.20' WHERE JOCKEY_ID = 'J1' AND RACE_ID = 'R1';
13
14
15  SELECT JOCKEY_NAME, YEAR, DISTANCE FROM (JOCKEY
16      INNER JOIN PARTICIPATION ON PARTICIPATION.JOCKEY_ID = JOCKEY.JOCKEY_ID
17    ) INNER JOIN RACE ON RACE.RACE_ID = PARTICIPATION.RACE_ID;
```









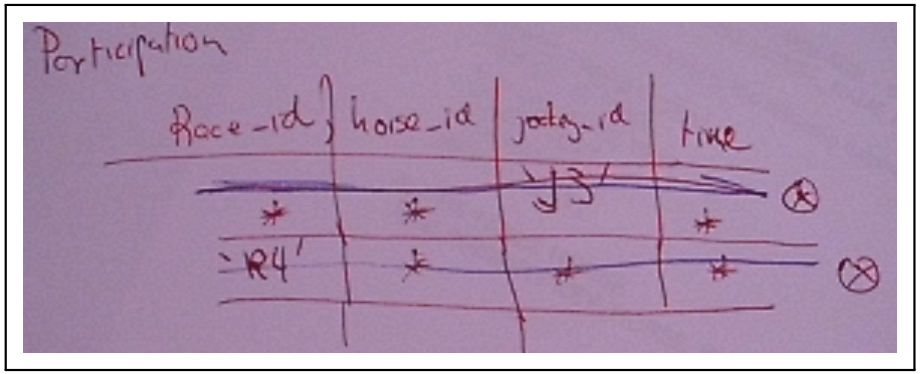
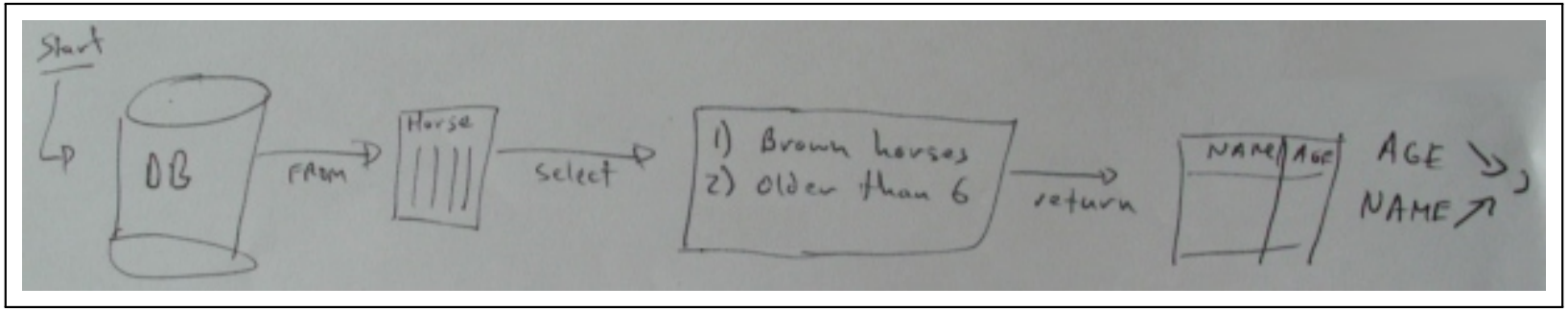
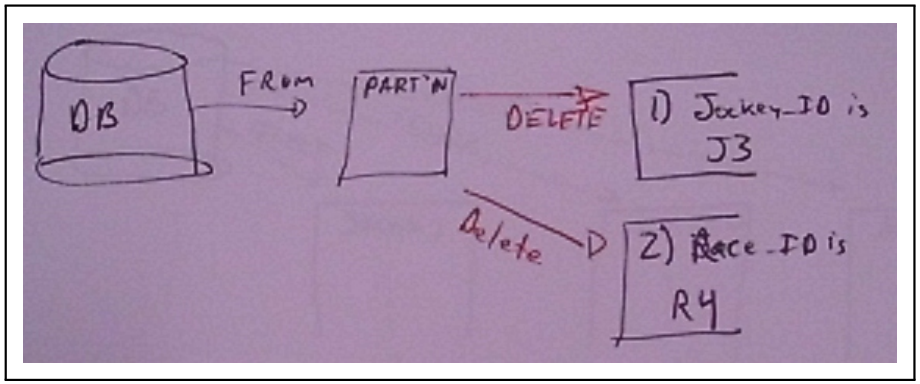
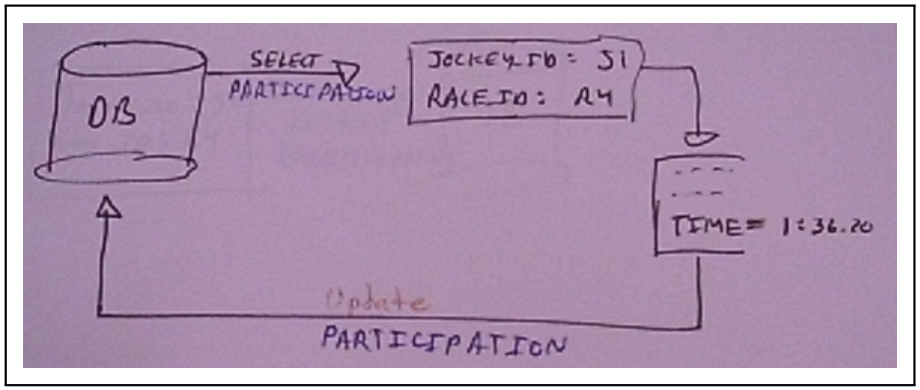
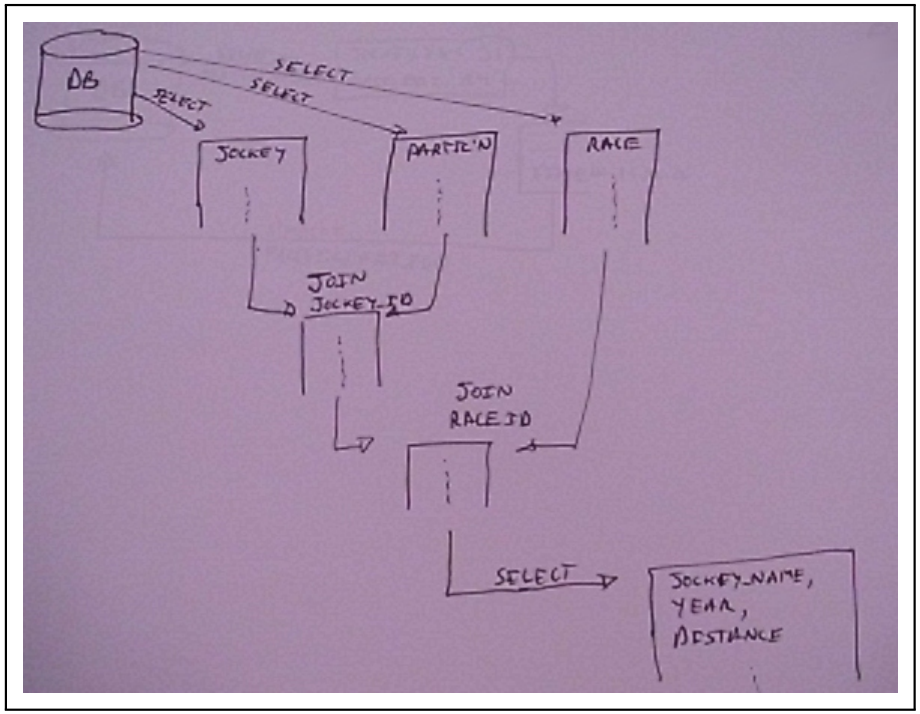


Table-based Notations

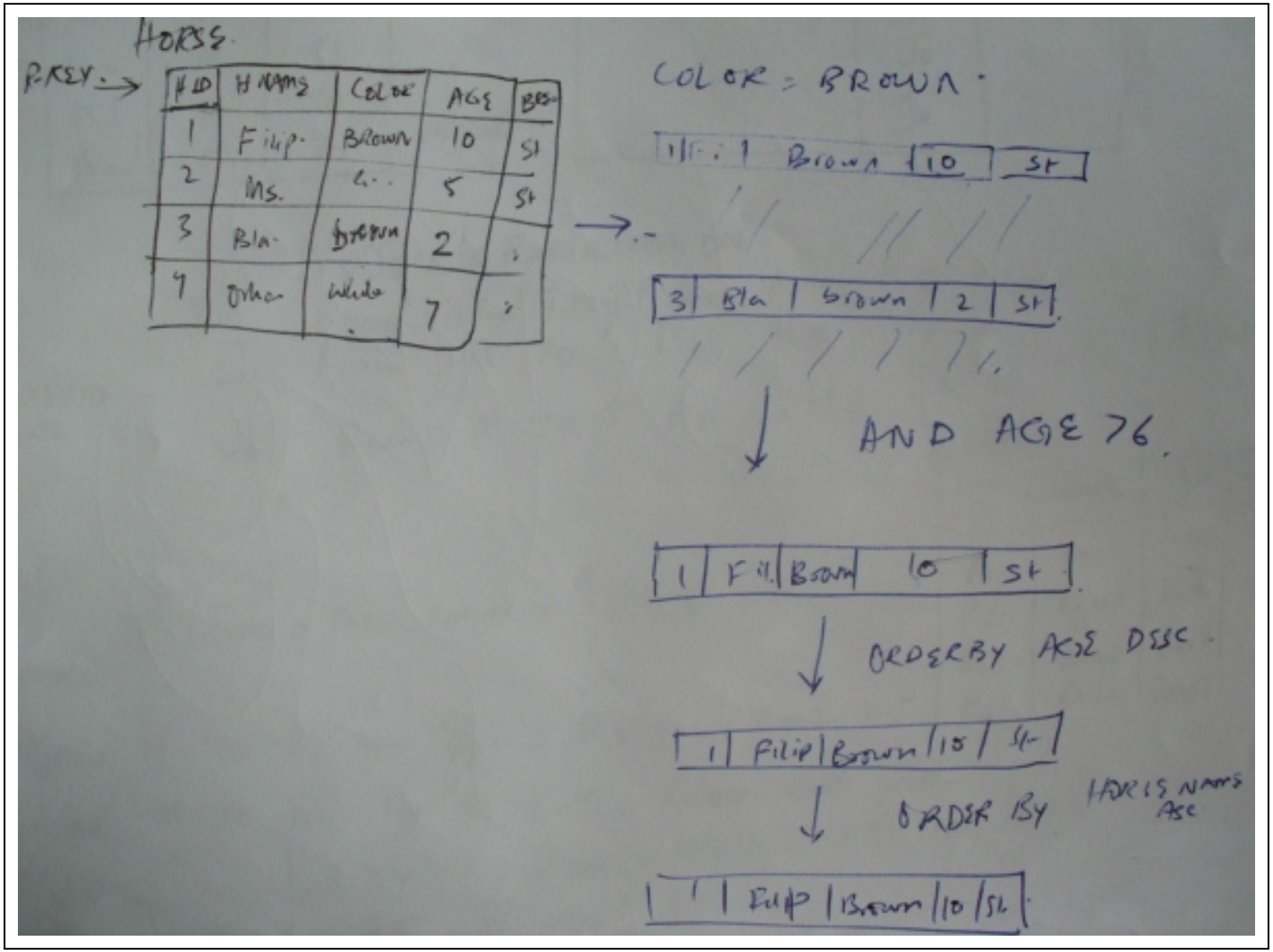




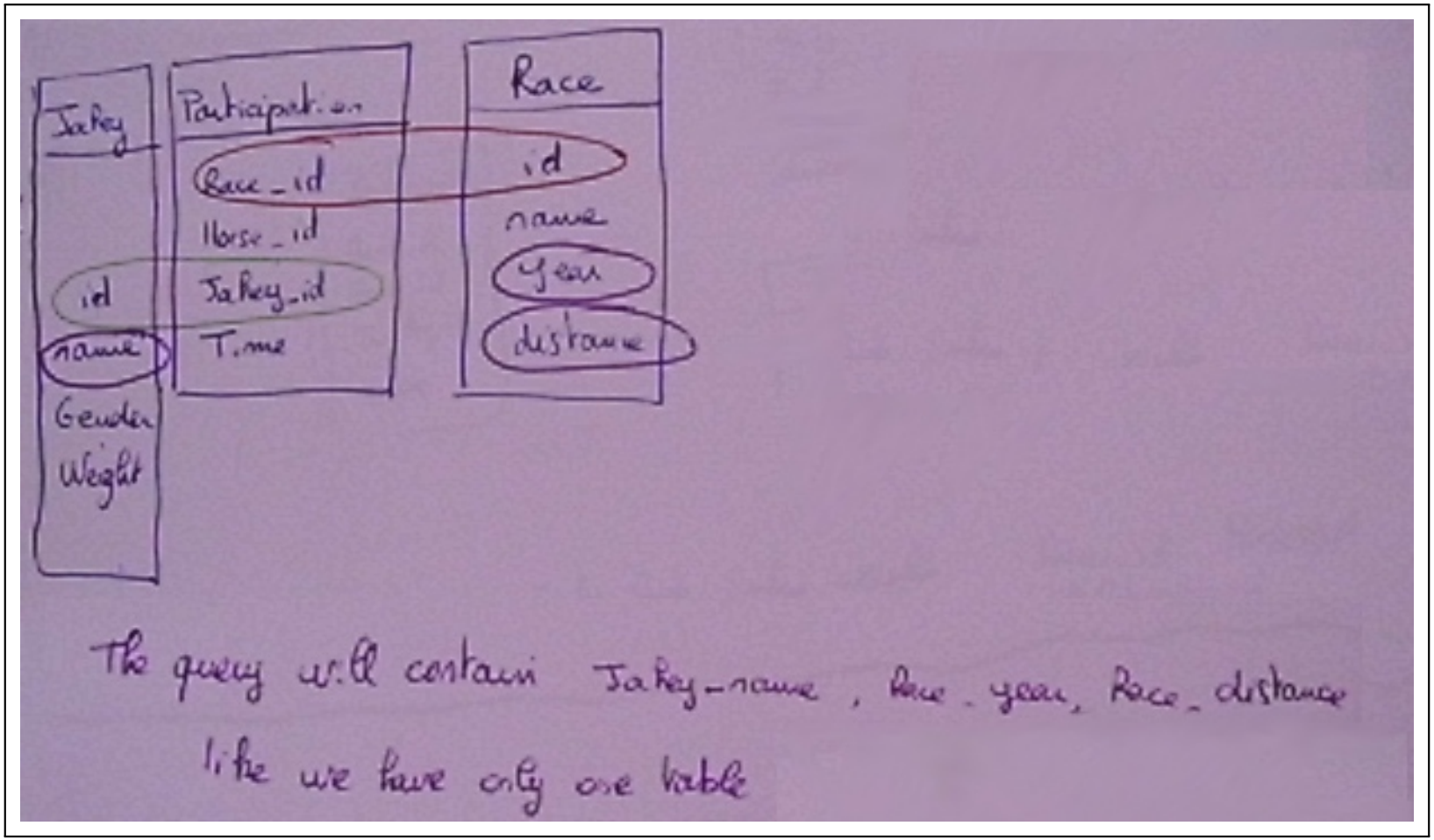




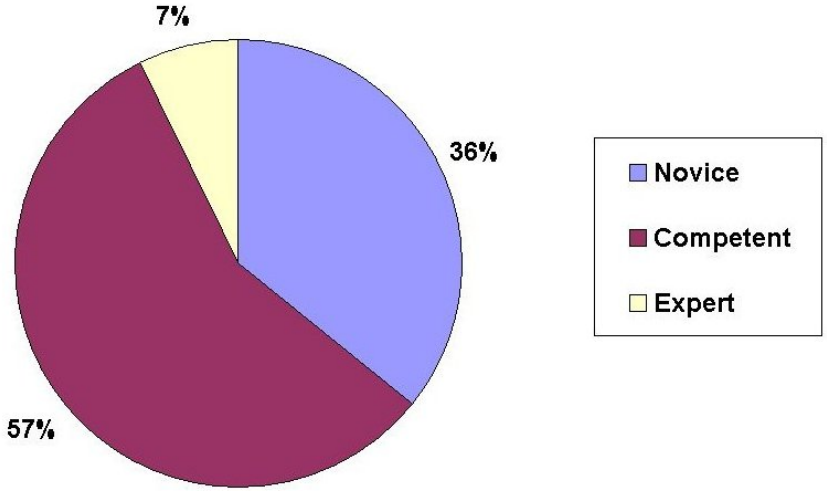
Order-based Notations



Artificial Data Examples



Natural Language Descriptions



Notational Characteristics (between subjects)	Participants		Percentages	
	Yes	No	Yes	No
Table-based notation	19	11	**63%**	37%
Order-oriented notation	10	20	33%	**67%**
Natural language descriptions	5	25	17%	**83%**
Artificial data examples	8	22	27%	**73%**

Preliminary Results

Work in Progress

- Code the data and analyze the order in which the diagrams were drawn
- Differentiate common vs. casual characteristics of the notations

Future Work

- Propose alternative notations for the representation of database queries
- Design studies comparing the proposed notations and SQL
- Develop and evaluate a query visualization tool

Summary

- We want to know what notation should be used to represent database queries.
- After conducting a study involving 30 participants, we are working on the analysis of the data
- Future work includes the development and evaluation of a query visualization tool.

Thanks!

Contact Information

- zuzelvp@cs.toronto.edu
- http://zuzelvp47uoft.wordpress.com
- Bahen Centre, room 3228