DBM1: Databases
The 2020 Paralympics

João Marques

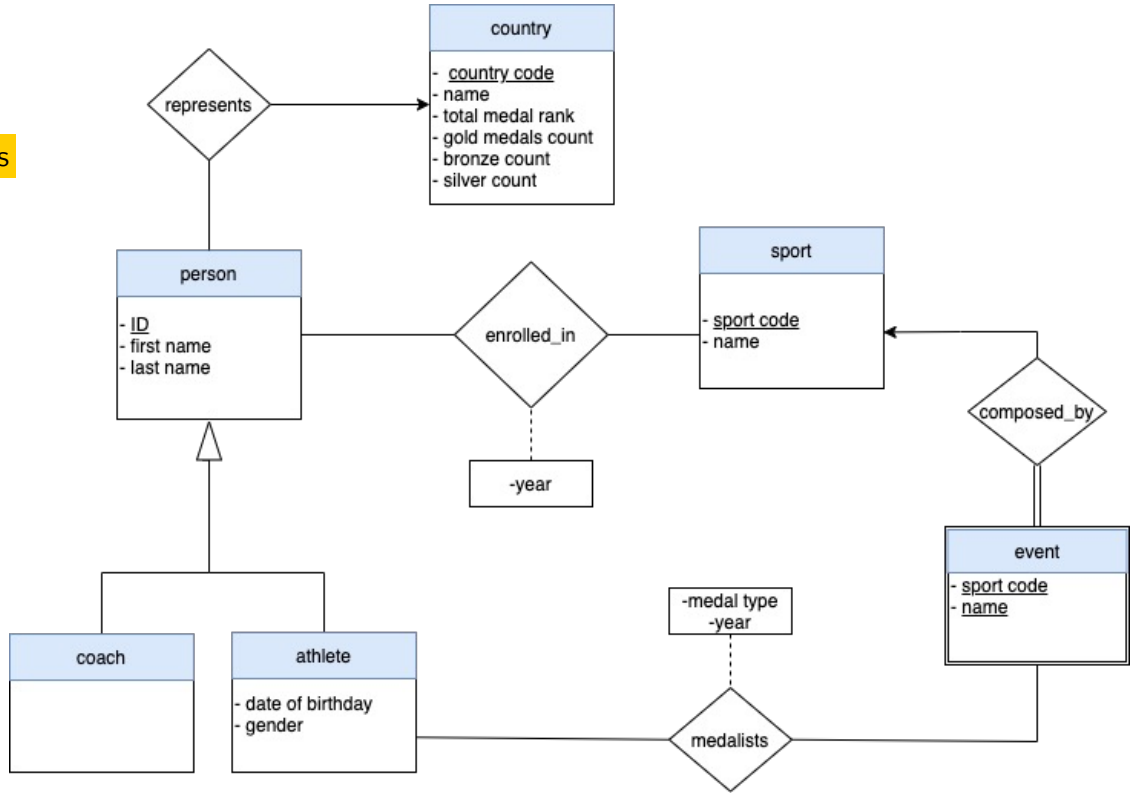
Katharina Alefs

The 2020 Paralympics Dataset

- 4426 athletes, 212 teams from 162 countries are are competing for 2141 medals
- Preprocessing steps: split names, replace gender column with *char*
- Generated mock data for other editions using Mockaroo

ER Diagram

Entities and design choices



Natural Language Queries

- Number of athletes that where enrolled per discipline?
- Which first name was the 2nd most popular among the athletes?
- How many gold, silver and bronze medals did each country win? (medal tally)
- For each year (1980 - 1990), how many athletes were born then and what are their names (sorted alphabetically)?
- The birth year of athletes that won exactly 1 gold and 1 silver and 1 bronze medal?
- From all winning teams across all events, which team has the most members and what are their names?
- Is there a female athlete that won a medal in two different disciplines in two different editions?
- Is there any athlete that skipped one edition but then performed better then his/her former participation?

Number of athletes that where enrolled per discipline?

```
SELECT first_name,
COUNT(first_name) AS counter
FROM person NATURAL JOIN athlete
GROUP BY(first_name)
ORDER BY counter DESC
OFFSET 1 LIMIT 1;
```

The birth year of athletes that won exactly 1 gold, 1 silver and 1 bronze medal?

```
SELECT EXTRACT (YEAR FROM date_of_birthday)
FROM athlete NATURAL JOIN person
WHERE id IN
    (SELECT id
    FROM medalists
    WHERE medal_type = 'Gold'
    GROUP BY id
    HAVING COUNT(id) = 1)
AND id IN (SELECT id
    FROM medalists
    WHERE medal_type = 'Silver'
    GROUP BY id
    HAVING COUNT(id) = 1)
AND id IN (SELECT id
    FROM medalists
    WHERE medal_type = 'Bronze'
    GROUP BY id
    HAVING COUNT(id) = 1);
```

A female athlete that won a medal in two different disciplines in different editions?

```
SELECT DISTINCT(winner.id), winner.first_name,
winner.last_name
FROM (person
NATURAL JOIN athlete
NATURAL JOIN medalists)
AS winner
WHERE winner.gender = 'F' AND winner.id IN(
    SELECT id
    FROM medalists AS m
    WHERE m.id = winner.id
    AND m.year != winner.year
    AND m.sport_code != winner.sport_code
    )
ORDER BY winner.id;
```

Relational Algebra

- First name of all athletes that won a medal in 2020, but did not win a medal in 2024

```
π first_name (person ⋈ athlete ⋈ (σ year ='2020' (medalists)))
- π first_name (person ⋈ athlete ⋈ (σ year ='2024' (medalists)))
```

- All the people (coaches and athletes) who represent Portugal in sport Judo

```
π id (σ represents = 'Portugal' (person) ⋈ σ sports_code = 'JUD' (enrolled))
```

Complications

"Which first name was the 2nd most popular among the athletes?"

```sql
SELECT first_name,
COUNT(first_name) AS counter
FROM person NATURAL JOIN athlete
GROUP BY(first_name)
ORDER BY counter DESC
OFFSET 1 LIMIT 1;
```

```sql
SELECT first_name, COUNT(first_name)
FROM person NATURAL JOIN athlete
WHERE first_name != '' AND first_name NOT IN
    (SELECT first_name
        FROM person NATURAL JOIN athlete
        WHERE first_name IS NOT NULL
        GROUP BY(first_name)
        HAVING COUNT(first_name) >=ALL(
            SELECT COUNT(first_name)
            FROM person NATURAL JOIN athlete
            WHERE first_name IS NOT NULL
            GROUP BY(first_name))
        )
GROUP BY(first_name)
HAVING COUNT(first_name) >= ALL(SELECT COUNT(first_name)
FROM person NATURAL JOIN athlete
WHERE first_name IS NOT NULL AND first_name NOT IN
    (SELECT first_name
        FROM person NATURAL JOIN athlete
        WHERE first_name IS NOT NULL
        GROUP BY(first_name)
        HAVING COUNT(first_name) >=ALL(
            SELECT COUNT(first_name)
            FROM person NATURAL JOIN athlete
            WHERE first_name IS NOT NULL
            GROUP BY(first_name))
        )
GROUP BY(first_name)
ORDER BY first_name DESC
);
```

Thanks!

Any questions ?

Quick Results