



**Karolinska  
Institutet**

# Introduction to PROC SQL in SAS

Ninoa Malki

# Background

SQL - Structured Query Language.

Developed by IBM in the 1970s for querying, updating and modifying relational databases.

SQL is currently used in many relational database software.

# Terminology

## SAS Data Step VS. Proc SQL

SAS Data step	Proc SQL
Dataset	Table
Variable	Column
Observation	Row
Merge	Join

# Introduction PROC SQL

Proc SQL is the SAS implementation of SQL.

Proc SQL is powerful SAS procedure that combines the functionality of the SAS data step with the SQL language.

# Common SQL Statements

1. **SELECT**

2. **INSERT**

3. **UPDATE**

# Syntax

PROC SQL	DATA STEP
<pre>PROC SQL;     Create Table ... AS     Select ....     From ...     Where ... QUIT;</pre>	<pre>DATA ... ;     SET ... ;     WHERE .... ; RUN;</pre>

# SELECT Syntax

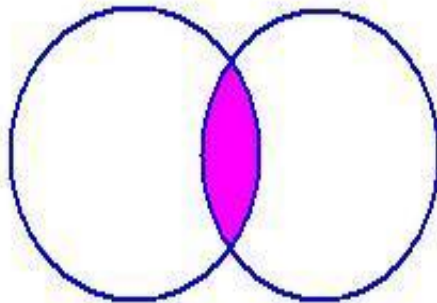
SELECT	Specifies the columns to be selected
FROM	Specifies the tables to be queried
< WHERE	Subset the data based on a condition
< GROUP BY	Classifies the data into groups
< HAVING	Subset groups of data based on a group condition
< ORDER BY	Sort the rows by the values of specific columns

# Join / Merge

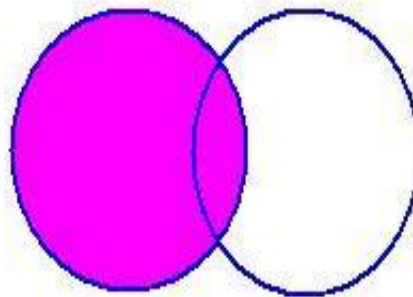
- In SQL the method to connect tables is called Join.
- A join can connect two or more tables together so the data is returned, deleted or updated as one dataset.
- Cartesian product – join many to many observations.
- No need of sorted data.
- No need of similar variable names.



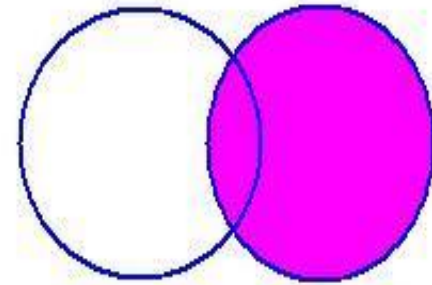
## JOINS AND SET OPERATIONS IN RELATIONAL DATABASES



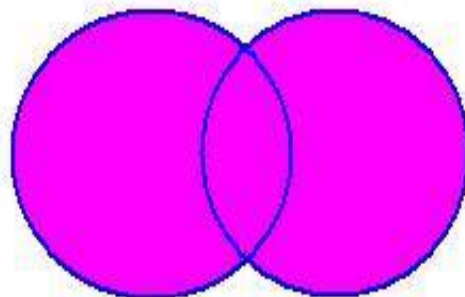
Inner join (result similar  
to Intersect)



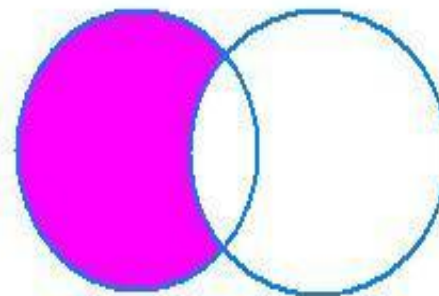
Left outer join



Right outer join



Full outer join



Minus

# Join / Merge

- Inner join / join:

Return rows when there is at least one match in both tables (if a and b).

- Full outer Join:

Returns rows when there is a match in one of the tables (if a or b).

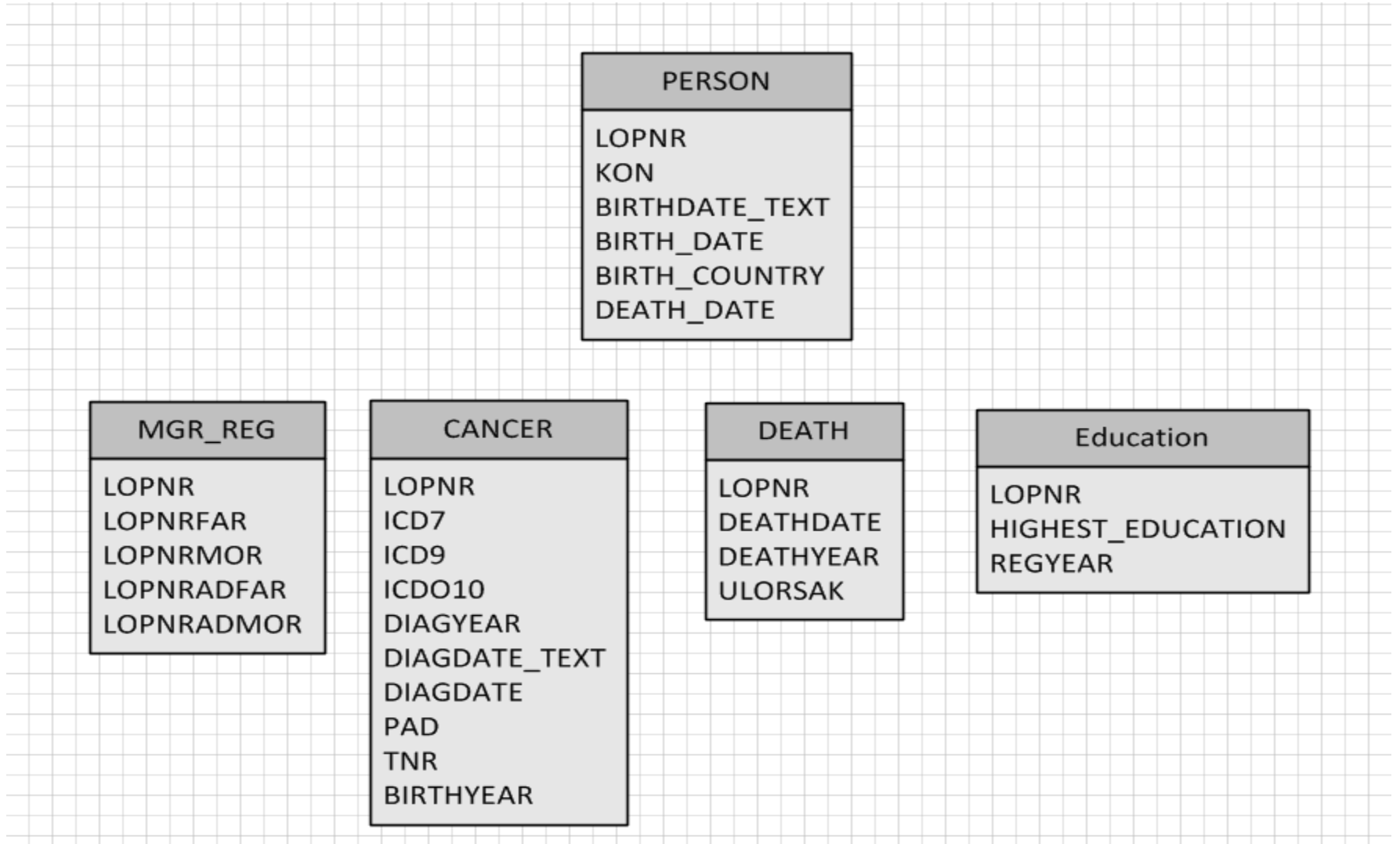
- Left outer Join

return all rows from the left table even if there is no match in the right table (if a).

- Right outer Join

return all rows from the right table even if there is no match in the left table (if b).

# Database



# Advantage / Disadvantage with PROC SQL

## Advantage

- Multiple data steps can be accomplished at same task.
- Less code is required.
- The code is easier to follow.
- Less computer resources.
- Has the ability to read in usorted data.
- M:M Join.

## Disadvantage

- Can't create a table from non relational database data (Excel).
- Can't create multi-tables at same statement