*-- warmup query: what proceedings in 2016?*

**SELECT** proc\_id, proc\_title

**FROM** proceedings

**WHERE** *EXTRACT*(**YEAR FROM** proc\_start\_date) = 2016;

**SELECT** proc\_id, proc\_title

**FROM** proceedings

**WHERE** *EXTRACT*(**YEAR FROM** proc\_start\_date) >= 2016;

**SeLEct** proc\_id, proc\_title

**FROM** proceedings

**WHERE** proc\_start\_date >= **'2016-01-01'**

**AND** proc\_start\_date < **'2017-01-01'**;

*-- string queries*

**SELECT** proc\_id, proc\_title, proc\_start\_date

**FROM** proceedings

**WHERE** proc\_title **LIKE '%Extended Abstracts%'**;

*-- let's aggregate*

**SELECT** *COUNT*(*\**)

**FROM** proceedings;

*-- what are the first and last proceedings?*

**SELECT** *MIN*(proc\_start\_date) **AS first**, *MAX*(proc\_start\_date) **AS last**

**FROM** proceedings;

*-- what does this do? what is it averaging?*

*-- will whoever finds out post to Piazza?*

**SELECT** *AVG*(proc\_start\_date) **AS** average

**FROM** proceedings;

*-- but this is somewhat more sensical*

**SELECT** *AVG*(*EXTRACT*(**YEAR FROM** proc\_start\_date)) **AS** average\_year

**FROM** proceedings;

*-- grouping makes them sing*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(proc\_id) **AS** num\_confs

**FROM** proceedings

**GROUP BY year**;

*-- count(\*) also works here*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(*\**) **AS** num\_confs

**FROM** proceedings

**GROUP BY year**;

*-- difference: count(\*) counts rows, count(col) counts non-null values in col*

*-- missing from GROUP*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

proc\_start\_date,

*COUNT*(proc\_id) **AS** num\_confs

**FROM** proceedings

**GROUP BY year**;

*-- oops - start date isn't grouped by!*

*-- you can group by more than one thing*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*EXTRACT*(**MONTH FROM** proc\_start\_date) **AS month**,

*COUNT*(proc\_id) **AS** num\_confs

**FROM** proceedings

**GROUP BY year**, **month**;

*-- SQL does not guarantee order. If you want it, ORDER BY*

*-- you can group by more than one thing*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*EXTRACT*(**MONTH FROM** proc\_start\_date) **AS month**,

*COUNT*(proc\_id) **AS** num\_confs

**FROM** proceedings

**GROUP BY year**, **month**

**ORDER BY year**, **month**;

*-- let's sort years by conference count! Most conferences first (DESC)*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(*\**) **AS** num\_confs

**FROM** proceedings

**GROUP BY year**

**ORDER BY** num\_confs **DESC**;

*-- two tables. what articles in 2016?*

**SELECT** proc\_id, proc\_title, article\_id, title

**FROM** article

**JOIN** proceedings **USING** (proc\_id)

**WHERE** *EXTRACT*(**YEAR FROM** proc\_start\_date) = 2016;

*-- this is identical:*

**SELECT** proceedings.proc\_id, proc\_title, article\_id, title

**FROM** article

**JOIN** proceedings **ON** article.proc\_id = proceedings.proc\_id

**WHERE** *EXTRACT*(**YEAR FROM** proc\_start\_date) = 2016;

*-- so is this:*

**SELECT** proceedings.proc\_id, proc\_title, article\_id, title

**FROM** article, proceedings

**WHERE** article.proc\_id = proceedings.proc\_id

**AND** *EXTRACT*(**YEAR FROM** proc\_start\_date) = 2016;

*-- we can define table aliases*

**SELECT** p.proc\_id, proc\_title, article\_id, title

**FROM** article a, proceedings p

**WHERE** a.proc\_id = p.proc\_id

**AND** *EXTRACT*(**YEAR FROM** proc\_start\_date) = 2016;

*-- AS works too (I like that)*

**SELECT** p.proc\_id, proc\_title, article\_id, title

**FROM** article **AS** a, proceedings **AS** p

**WHERE** a.proc\_id = p.proc\_id

**AND** *EXTRACT*(**YEAR FROM** proc\_start\_date) = 2016;

*-- Let's look at people - person and pet names*

**SELECT pers\_name**, pet\_name

**FROM** people.person, people.pet

**WHERE** owner\_id = **pers\_id**;

*-- Nice syntax:*

**SELECT pers\_name**, pet\_name

**FROM** people.pet

**JOIN** people.person **ON** owner\_id = **pers\_id**;

*-- let's break things*

**SELECT pers\_name**, pet\_name

**FROM** people.person, people.pet;

*-- query planner, what dost thou think?*

**EXPLAIN SELECT** proc\_id, proc\_title, article\_id, title

**FROM** article

**JOIN** proceedings **USING** (proc\_id)

**WHERE** *EXTRACT*(**YEAR FROM** proc\_start\_date) = 2016;

*-- let's count articles by year*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(article\_id) **AS** article\_count

**FROM** proceedings

**JOIN** article **USING** (proc\_id)

**GROUP BY year**;

*-- let's count conf proceedings and articles by year*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(proc\_id) **AS** conf\_count,

*COUNT*(article\_id) **AS** article\_count

**FROM** proceedings

**JOIN** article **USING** (proc\_id)

**GROUP BY year**;

*-- oops, let's peek inside. this is the table from previous query*

*-- before grouping is applied*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

proc\_id, article\_id

**FROM** proceedings

**JOIN** article **USING** (proc\_id);

*-- many article IDs for the same proc ID*

*-- fix that by counting \*distinct\* proc IDs*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(**DISTINCT** proc\_id) **AS** conf\_count,

*COUNT*(article\_id) **AS** article\_count

**FROM** proceedings

**JOIN** article **USING** (proc\_id)

**GROUP BY year**;

*-- let's find all years with at least 10 conferences*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(proc\_id) **AS** conf\_count

**FROM** proceedings

**WHERE** conf\_count >= 10

**GROUP BY year**;

*-- we haven't grouped!*

*-- HAVING to the rescue*

**SELECT** *EXTRACT*(**YEAR FROM** proc\_start\_date) **AS year**,

*COUNT*(proc\_id) **AS** conf\_count

**FROM** proceedings

**GROUP BY year**

**HAVING** conf\_count >= 10;