

Aggregation

Aggregation

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What is the average instructor salary?

We have already seen how to express this in SQL

Aggregation

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What is the average instructor salary?

We have already seen how to express this in SQL

```
sqlite> select round(avg(salary)) from instructor;  
74833.0
```

Aggregation

Aggregation functions:

- ❑ avg
- ❑ min
- ❑ max
- ❑ sum
- ❑ count

“Aggregation”
functions because
they aggregate
multiple values into
a single value

```
sqlite> select min(salary) from instructor;  
40000
```

Combining aggregation with other SQL

teaches

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-101	1	Spring	2010
15151	MUS-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010

How many instructors teaching in the Spring 2010 semester?

Exercise: try this

Combining aggregation with other SQL

teaches

ID	course_id	sec_id	semester	year
10101	CS-101		1 Fall	2009
10101	CS-315		1 Spring	2010
10101	CS-347		1 Fall	2009
12121	FIN-101		1 Spring	2010
15151	MUS-199		1 Spring	2010
22222	PHY-101		1 Fall	2009
32343	HIS-351		1 Spring	2010
45565	CS-101		1 Spring	2010
45565	CS-319		1 Spring	2010
76766	BIO-101		1 Summer	2009
76766	BIO-301		1 Summer	2010
83821	CS-190		1 Spring	2009
83821	CS-190		2 Spring	2009
83821	CS-319		2 Spring	2010

How many instructors teaching in the Spring 2010 semester?

Exercise: try this

```
sqlite> select count(distinct(ID))  
        from teaches  
        where semester="Spring" and year="2010";  
6
```

Aggregation with grouping

course

course_id	title	dept_name	credits
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principals	Physics	4

How many courses offered for each dept?

Do we already know how to express this in SQL?

dept_name	#courses

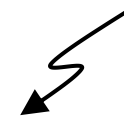
Aggregation with grouping

course

course_id	title	dept_name	credits
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principals	Physics	4

How many courses offered for each dept?

Do we already know how to express this in SQL?



```
sqlite> select dept_name, count(*) from course group by dept_name;
Biology,3
Comp. Sci.,5
Elec. Eng.,1
Finance,1
History,1
Music,1
Physics,1
```


Exercise

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What is the average salary by department?

Exercise

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What is the average salary by department?

```
sqlite> select dept_name, round(avg(salary)) from instructor group by dept_name;
Biology,72000.0
Comp. Sci.,77333.0
Elec. Eng.,80000.0
Finance,85000.0
History,61000.0
Music,40000.0
Physics,91000.0
```

Grouping by multiple columns

teaches

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-101	1	Spring	2010
15151	MUS-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010

How many different instructors teaching in each semester/year?

Grouping by multiple columns

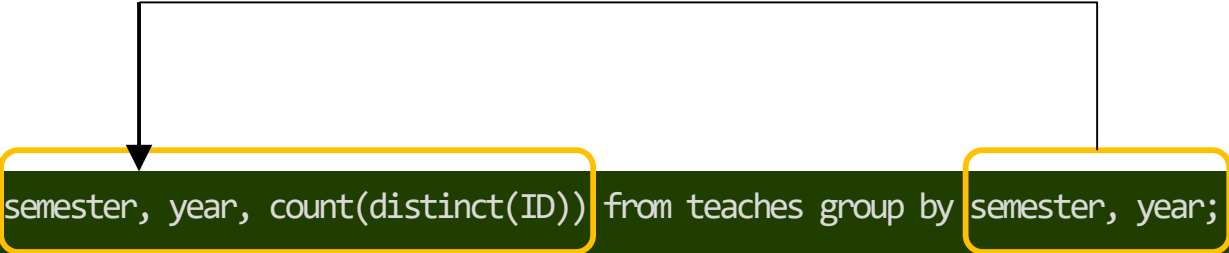
teaches

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-101	1	Spring	2010
15151	MUS-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010

How many different instructors teaching in each semester/year?

```
sqlite> select semester, year, count(distinct(ID)) from teaches group by semester, year;  
Fall,2009,2  
Spring,2009,2  
Spring,2010,6  
Summer,2009,1  
Summer,2010,1
```

The “group by” rule



```
sqlite> select semester, year, count(distinct(ID)) from teaches group by semester, year;  
Fall,2009,2  
Spring,2009,2  
Spring,2010,6  
Summer,2009,1  
Summer,2010,1
```

The “having” clause

instructor

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

What is the average salary by department, for departments where the average salary is more than \$42,000?

‘having’ like
‘where’, but
applies to
groups

```
sqlite> select dept_name, round(avg(salary))
from instructor group by dept_name having avg(salary) > 42000;
Biology,72000.0
Comp. Sci.,77333.0
Elec. Eng.,80000.0
Finance,85000.0
History,61000.0
Physics,91000.0
```