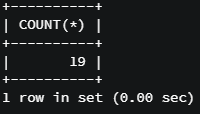
Section slides: <http://webdev.slides.com/coltsteele/mysql-99-101#/80>

* This section will focus on aggregate functions, such as counting, averaging, summing, etc., all of which involves grouping our data in some way
* So rather than working with individual rows, we can group items and perform more interesting calculations on them
  + For example, we can find the average number of sales per year, where we group items by year
  + Or we can group books by author and determine the average number of pages in books written by each author

# The COUNT Function

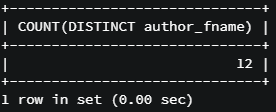
* This function counts whatever you want it to count
* <https://www.w3schools.com/Sql/func_mysql_count.asp>
* We can count the number of books in our database as follows:
  + Within the syntax, the asterisk \* simply means you want to count the number of items, whether it be the number of rows in the table, or the number of “super rows” that are formed after performing GROUP BY (more on that later)





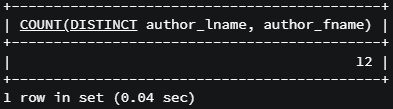
* We can also use it to ask questions such as, how many *unique* author\_fnames are in our database?
  + To solve this, we can combine COUNT and DISTINCT



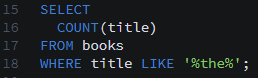


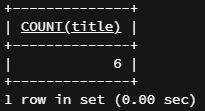
* An edge case shows itself when we try to find the number of *unique* authors in our books table, where we select author\_lname, author\_fname using DISTINCT.
  + This shows that we have 12 unique selections of author\_lname and author\_fname





* What if we want to find how many titles contain the word “the”? To accomplish this, we can use a combination of COUNT with WHERE and LIKE





* Code summary for the COUNT function

SELECT COUNT(\*) FROM books;

SELECT COUNT(author\_fname) FROM books;

SELECT COUNT(DISTINCT author\_fname) FROM books;

SELECT COUNT(DISTINCT author\_lname) FROM books;

SELECT COUNT(DISTINCT author\_lname, author\_fname) FROM books;

SELECT title FROM books WHERE title LIKE '%the%';

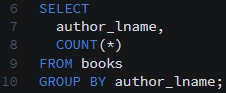
SELECT COUNT(\*) FROM books WHERE title LIKE '%the%';

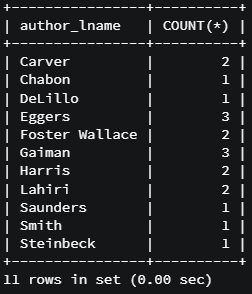
# The GROUP BY Function

* GROUP BY is an aggregate function that works a bit differently than the others that we’ve seen.
  + It does not provide an immediate result. You must use it alongside an aggregate function
  + GROUP BY summarizes or aggregates identical data into single rows
  + It allows us to group things by particular properties and values, and we can use these groupings to do even more analyses
* Check out the syntax below, where we grouping by author\_lname

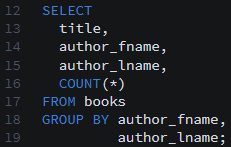


* + If you try this, you will just see a list of unique author\_lnames, alongside just ONE of their book titles
    - In newer versions of MySQL, you’ll get an error since GROUP BY is not really intended to be used this way
    - This is because only\_full\_group\_by is enabled
  + But in the background, the method is actually grouping all of the titles that share the same author into a “super row” underneath one author name. They are grouped together, we just can’t see them
* Suppose we want to count how many books are attributed to each unique author\_lname. We can do the following:
  + This means “give me two columns, the first having author\_lname, and the second having the number of rows under each author\_lname



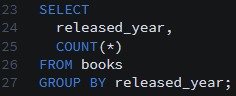


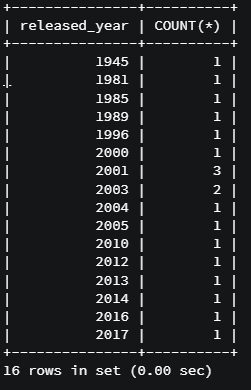
* + We see an issue here, which is that the query is telling us that the author whose last name is Harris has two books. In reality, there are two authors with the last name Harris (Dan and Freida), and they each have one book. We can solve this using the following selection



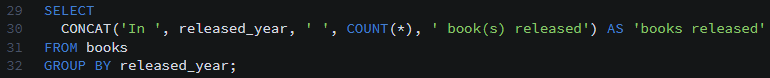


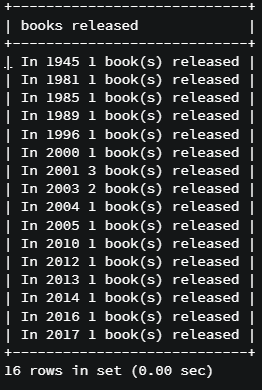
* We can also group by release year and figure out how many books were released each year! Here’s how:





* + We see that 2003 had the most released books out of all years in our table
  + Let’s make it more fancy by throwing in some CONCAT





* Code summary, GROUP BY

SELECT title, author\_lname FROM books;

SELECT title, author\_lname FROM books

GROUP BY author\_lname;

SELECT author\_lname, COUNT(\*)

FROM books GROUP BY author\_lname;

SELECT title, author\_fname, author\_lname FROM books;

SELECT title, author\_fname, author\_lname FROM books GROUP BY author\_lname;

SELECT author\_fname, author\_lname, COUNT(\*) FROM books GROUP BY author\_lname;

SELECT author\_fname, author\_lname, COUNT(\*) FROM books GROUP BY author\_lname, author\_fname;

SELECT released\_year FROM books;

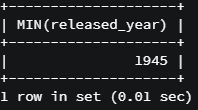
SELECT released\_year, COUNT(\*) FROM books GROUP BY released\_year;

SELECT CONCAT('In ', released\_year, ' ', COUNT(\*), ' book(s) released') AS year FROM books GROUP BY released\_year;

# The Basics of MIN and MAX

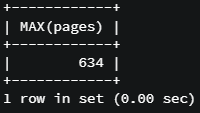
* As you might imagine, the MIN and MAX functions help you identify the minimum or maximum values in a table.
  + Can be used on their own, or can be combined with GROUP BY
* On their own, we can do things like finding the minimum released year of book release





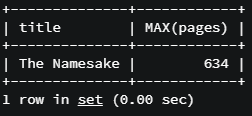
* + Thus, 1945 was the earliest year that a book in our database was released
* Using MAX, we can find the book with the highest number of pages in our table





* Suppose we wanted the title of book with the largest number of pages. To do this, we could try the following:



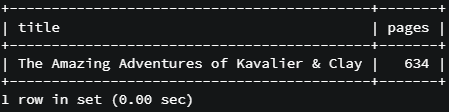


* + This is not entirely correct. While the longest book does have 634 pages, that book is NOT The Namesake. Rather, it is The Amazing Adventures of Kavalier & Clay



* + There are multiple ways around this. One way is to disregard MAX altogether and use ORDER BY and LIMIT





* + We’ll see other ways in the coming lectures
* Code summary: MIN and MAX basics

SELECT MIN(released\_year)

FROM books;

SELECT MIN(released\_year) FROM books;

SELECT MIN(pages) FROM books;

SELECT MAX(pages)

FROM books;

SELECT MAX(released\_year)

FROM books;

SELECT MAX(pages), title

FROM books;

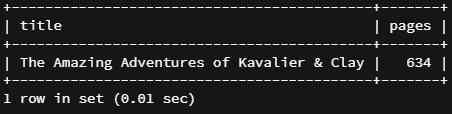
# SUBQUERIES – A Problem with Min and Max

* In the previous lecture we saw that the follow code does not return what we want. It returns the *first* title in the table, and then returns the max number of pages. The two are independent from each other



* We also saw a workaround above, in which we sorted the books by pages and selected (via LIMIT) out the book with the largest number of pages
* But what if we want a solution that utilizes the MAX function? We can do that as well, as the following **subquery** shows
  + A subquery is simply one query nested within another. We’ll see more of this later
  + The “lowest” level query will execute first, followed by the next level up.
  + In this case, the query will first find the largest number of pages, then enter that value into “WHERE pages =”. Then the higher level query will run, selecting the specific title and pages associated with that WHERE condition (the largest number of pages)





* One issue with subqueries is that it can be slow, since the system is essentially running two queries. For that reason, the ORDER BY and ASC LIMIT approach is actually faster and gives the same result
* Code summary

SELECT \* FROM books

WHERE pages = (SELECT Min(pages)

FROM books);

SELECT title, pages FROM books

WHERE pages = (SELECT Max(pages)

FROM books);

SELECT title, pages FROM books

WHERE pages = (SELECT Min(pages)

FROM books);

SELECT \* FROM books

ORDER BY pages ASC LIMIT 1;

SELECT title, pages FROM books

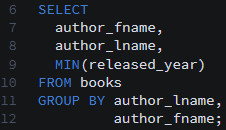
ORDER BY pages ASC LIMIT 1;

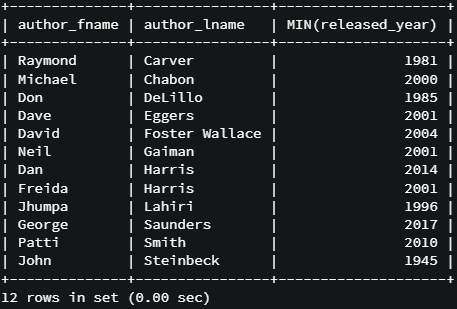
SELECT \* FROM books

ORDER BY pages DESC LIMIT 1;

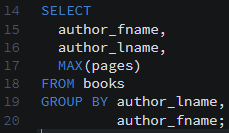
# Using MIN and MIX with GROUP BY

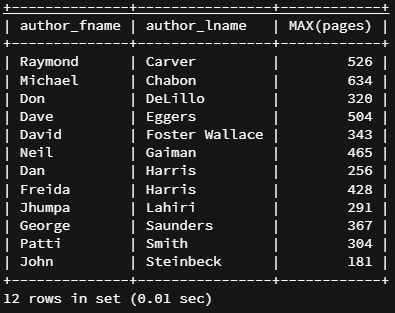
* Using MIN and MIX with GROUP BY is similar to using COUNT
* We’ll illustrate with an example: suppose we want to find the year that each author published his or her first book. How would we do this?
  + What we want to do is find the minimum release year for each author
  + So we’ll need to group the books by author, then find the minimal release year within each of those author “groups”. We’ll want to consider both the author first and last names since there are two authors with the last name Harris



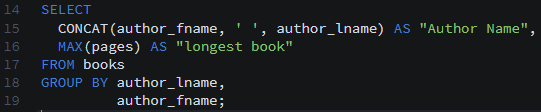


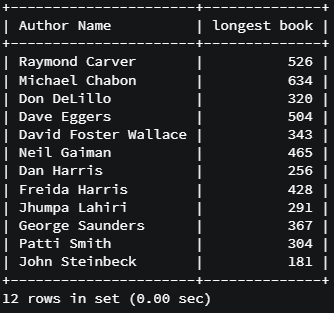
* In another example, we’ll find the *longest* page count for each author. This is very similar to the previous exercise. The result will be a table describing the longest books for each of the authors in our *books* table





* + Let’s make it look a little nicer by using aliases and concatenation





* We’ve been grouping mainly be authors, but we could group by literally any property in the table
* Code summary

SELECT author\_fname,

author\_lname,

Min(released\_year)

FROM books

GROUP BY author\_lname,

author\_fname;

SELECT

author\_fname,

author\_lname,

Max(pages)

FROM books

GROUP BY author\_lname,

author\_fname;

SELECT

CONCAT(author\_fname, ' ', author\_lname) AS author,

MAX(pages) AS 'longest book'

FROM books

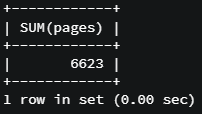
GROUP BY author\_lname,

author\_fname;

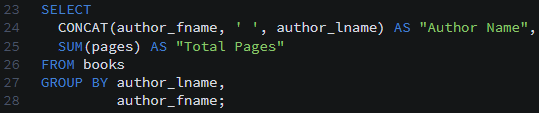
# The SUM Function

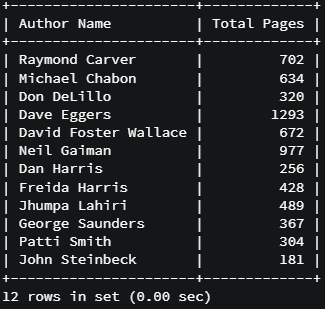
* The SUM function adds things together
* Let’s start by summing all pages in the entire *books* table. This is a very simple command (and not a very useful example)





* But the power of SUM comes with GROUP BY, where we want to know the sum of a quantity within a particular group
* In this example, we’ll sum all the pages that each author has written





* Code summary

SELECT SUM(pages)

FROM books;

SELECT SUM(released\_year) FROM books;

SELECT author\_fname,

author\_lname,

Sum(pages)

FROM books

GROUP BY

author\_lname,

author\_fname;

SELECT author\_fname,

author\_lname,

Sum(released\_year)

FROM books

GROUP BY

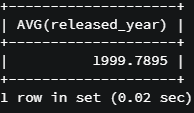
author\_lname,

author\_fname;

# The AVG Function

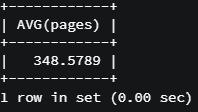
* The AVG function calculates the average value of your selected data
* Let’s calculate the average released year across all books in our table



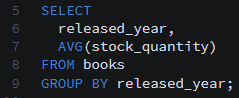


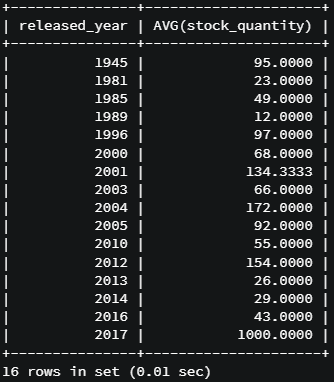
* + Later on we’ll talk about numbers and decimals. But for now when you do average it returns four decimal points due to the datatype that is used here
* Let’s calculate the average pages from all books



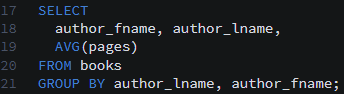


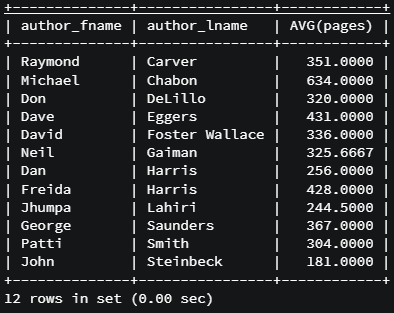
* It is most commonly used in conjunction with GROUP BY
* In this example, we calculate the average stock quantity for books released in the same year
  + Approach: group by year, then calculate average stock quantity





* Let’s print out average pages from each author





* Code summary

SELECT AVG(released\_year)

FROM books;

SELECT AVG(pages)

FROM books;

SELECT AVG(stock\_quantity)

FROM books

GROUP BY released\_year;

SELECT released\_year, AVG(stock\_quantity)

FROM books

GROUP BY released\_year;

SELECT author\_fname, author\_lname, AVG(pages) FROM books

GROUP BY author\_lname, author\_fname;

# Aggregate Functions Challenges

Code summary

SELECT COUNT(\*) FROM books;

SELECT COUNT(\*) FROM books GROUP BY released\_year;

SELECT released\_year, COUNT(\*) FROM books GROUP BY released\_year;

SELECT Sum(stock\_quantity) FROM BOOKS;

SELECT AVG(released\_year) FROM books GROUP BY author\_lname, author\_fname;

SELECT author\_fname, author\_lname, AVG(released\_year) FROM books GROUP BY author\_lname, author\_fname;

SELECT CONCAT(author\_fname, ' ', author\_lname) FROM books

WHERE pages = (SELECT Max(pages) FROM books);

SELECT CONCAT(author\_fname, ' ', author\_lname) FROM books

ORDER BY pages DESC LIMIT 1;

SELECT pages, CONCAT(author\_fname, ' ', author\_lname) FROM books

ORDER BY pages DESC;

SELECT released\_year AS year,

COUNT(\*) AS '# of books',

AVG(pages) AS 'avg pages'

FROM books

GROUP BY released\_year;