

Summary Statistics: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2020

Syntax

- Returning a count of rows in a table:

```
SELECT COUNT(Major)
FROM recent_grads
```

- Returning the minimum value of a table:

```
SELECT MIN(ShareWomen)
FROM recent_grads;
```

- Computing the sum of a column as an integer:

```
SELECT SUM(Total)
FROM recent_grads
```

- Computing the sum of a column as a float value:

```
SELECT TOTAL(Total)
FROM recent_grads
```

- Specifying a name for a column in the results:

```
SELECT (*) AS "Total Majors" -- also works without AS
FROM recent_grads
```

- Returning the unique values of a column:

```
SELECT DISTINCT Major_category
FROM recent_grads
```

- Performing an arithmetic operation on a table:

```
SELECT P75th - P25th quartile_spread  
FROM recent_grads
```

Concepts

- Summary statistics are used to summarize a set of observations.
- Everything is considered a table in SQL. One advantage of this simplification is that it's a common and visual representation that makes SQL approachable for a much wider audience.
- Datasets and calculations that aren't well suited for a table representation must be converted to be used in a SQL database environment.
- Aggregate functions are applied over columns of values and return a single value.
- The `COUNT` clause can be used on any column while aggregate functions can only be used on numeric columns.
- SQL supports that standard arithmetic operators (+, -, *, /).
- Arithmetic between two floats returns a float.
- Arithmetic between a float and an integer returns a float.
- Arithmetic between two integers returns an integer.

Resources

- [Aggregate Functions](#)
- [Summary Statistics](#)

