

Statistics Starters

Erin Stellato





<http://www.sqlskills.com/blogs/erin>

erin@sqlskills.com



- **Team of world-renowned SQL Server experts:**

- Erin Stellato (@ErinStellato)
- Glenn Berry (@GlennAlanBerry)
- Jonathan Kehayias (@SQLPoolBoy)
- Kimberly L. Tripp (@KimberlyLTripp)
- Paul S. Randal (@PaulRandal)

- **Instructor-led training: Immersion Events (US & UK)**

- **Online training:  <http://pluralsight.com/>**

- **Consulting: health checks, hardware, performance, upgrades**

- **Remote DBA: system monitoring and troubleshooting**

- **Conferences: SQLIntersection, PASS Summit**

- **Become a SQLskills Insider**

- <http://www.sqlskills.com/Insider>



Abstract

Have you been updating statistics as part of your maintenance plan, even though you're not really sure what they are?

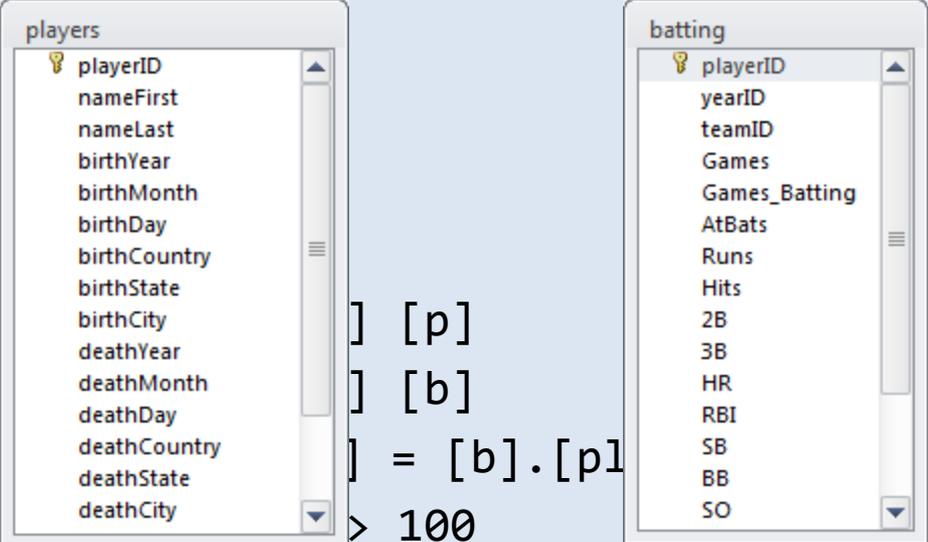
Have you been asking yourself where statistics exist, how to find them, and what people mean when they say "check statistics"?

Do you look at the database options for statistics and wonder if they really matter?

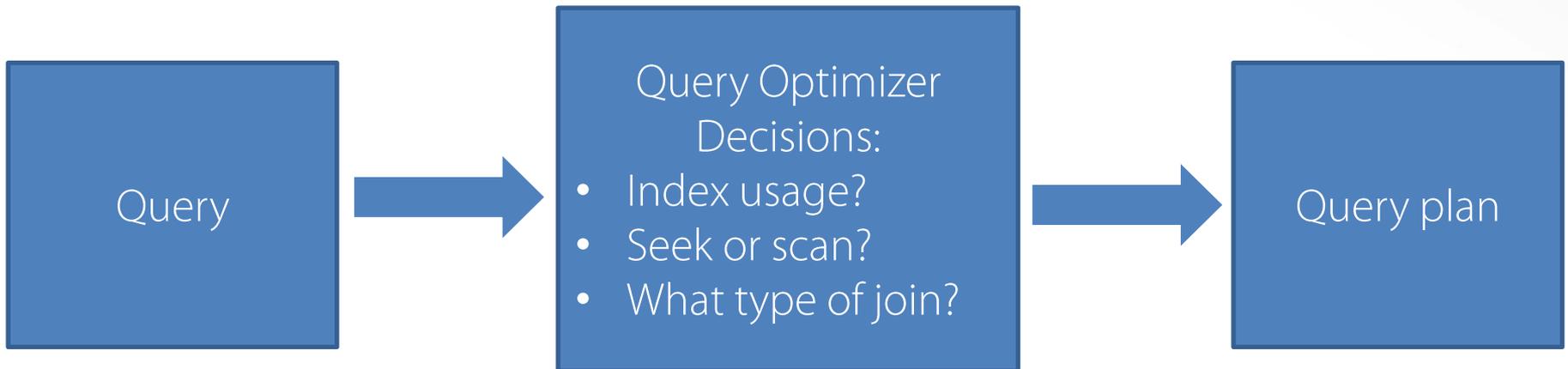
It's time to end the guessing and start understanding the basics about statistics. In this session I'll answer these questions using straightforward demos and real-world examples. And if you haven't been updating statistics as part of your maintenance plan don't worry, I'll get you started.

Why Should We Care About Statistics?

```
SELECT
  [p].[nameFirst],
  [p]
  [b]
  [b]
  [b]
FROM [players] [p]
JOIN [batting] [b]
  ON [p].[playerID] = [b].[playerID]
WHERE ([b].[Games_Batting] > 100
  OR [b].[Hits] > 200)
  AND [p].[nameLast] = 'Williams'
```



Why Should We Care About Statistics?



How Are Statistics Created?

- “...over the supplied column or set of columns of a table or indexed view...”
- **For every index (and indexed view)**
 - Created automatically
- **For individual columns**
 - Created automatically or manually
- **For multiple columns**
 - Created manually
- **For filtered values in a column**
 - Created manually
- **Use the sys.stats catalog view to list all statistics for a table**

(demo)

How Are Statistics Updated?

- **“...Up-to-date statistics allow the optimizer to accurately assess the cost of different query plans, and choose a high-quality plan.”**
- **As part of index rebuilds**
 - Only updates index statistics
 - Equivalent of a full scan (100% sample)
- **Manually (UPDATE STATISTICS or sp_updatestats)**

How Are Statistics Updated?(2)

- **Automatically**
 - A minimum of 500 + 20% of the rows
 - Exception: Trace Flag 2371 (2008R2 SP1+)
 - Not a full scan
 - Can be immediate or asynchronous
 - Index, column or filtered stats

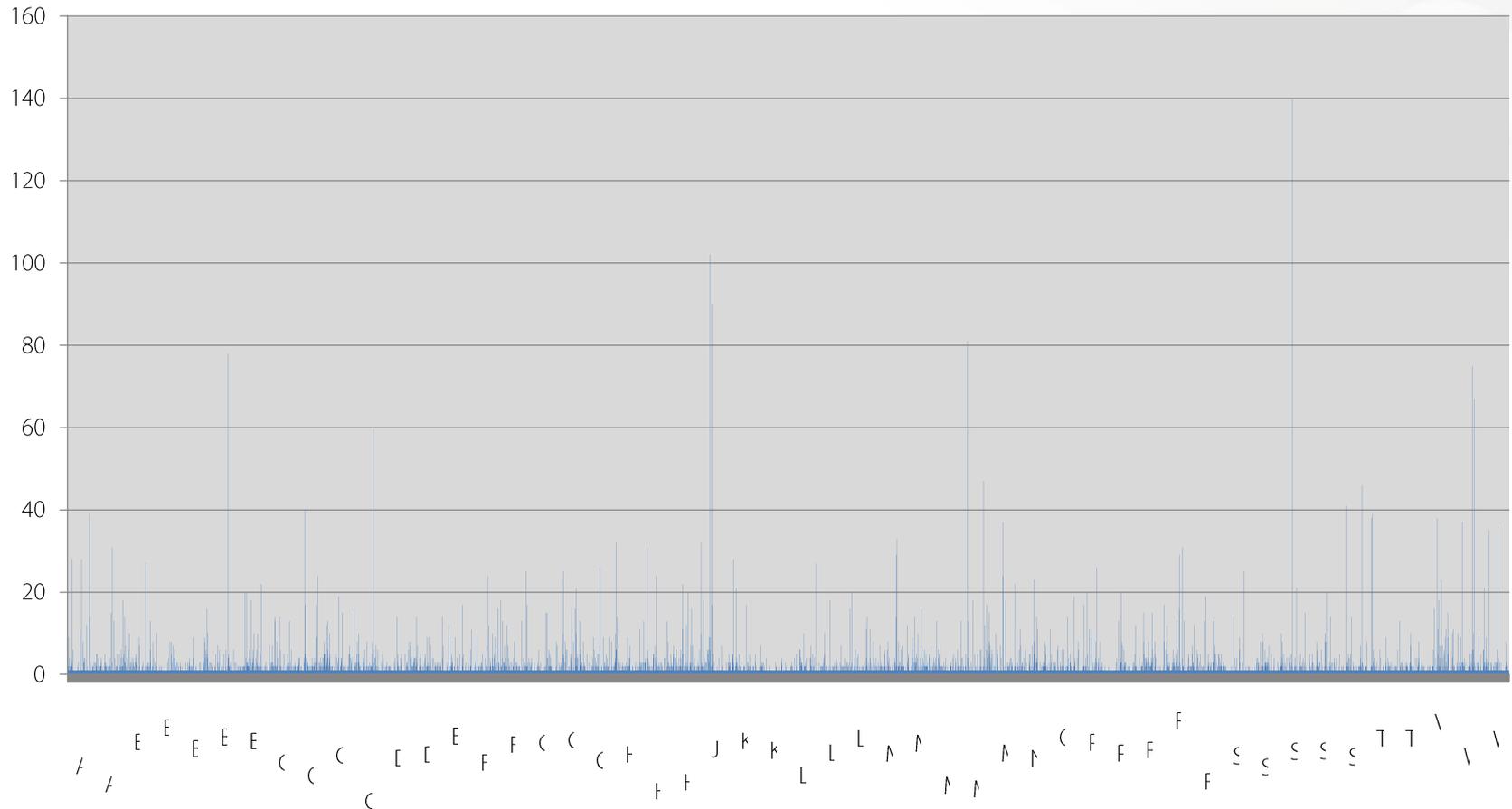
Remember that updating statistics causes queries to recompile*

(demo)

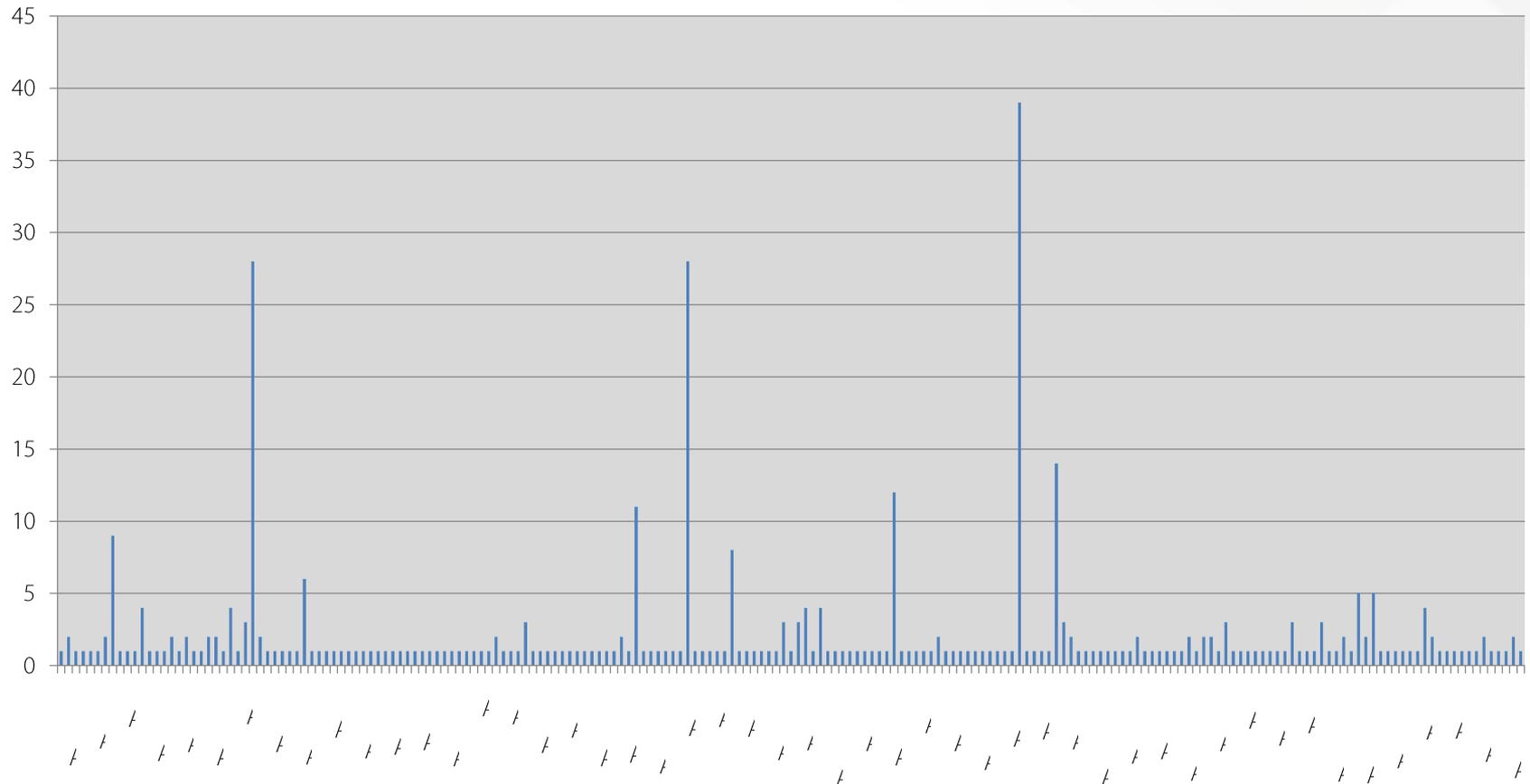
But what **ARE** statistics *really*?

- “...a histogram and associated density groups (collections)...”
- **Density: uniqueness of values with a set of data, calculated as (1/number distinct rows)**
 - $1/5000 = .0002$ density
- **Histogram: represents the distribution of values for a set of data**

Another way to think of a histogram



Another way to think of a histogram(2)



But what **ARE** statistics *really*?

- “...a histogram and associated density groups (collections)...”
- **Density: uniqueness of values with a set of data, calculated as (1/number distinct rows)**
 - $1/5000 = .0002$ density
- **Histogram: represents the distribution of values for a set of data**
- **What if the value doesn't exist in the histogram?**

Statistics and the Optimizer

- **“The query optimizer uses this statistical information to choose the most efficient plan for retrieving or updating data.”**
- **Statistics affect what plan the optimizer chooses**
- **There may be only a few ways to get it right, but thousands of ways to get it wrong**
- **It’s your job to set up the optimizer for success**

How should a DBA manage statistics?

- **Let SQL Server manage statistics**
 - Keep the AUTO UPDATE STATISTICS option enabled
- **Set up a job to manage statistics**
 - Maintenance plans
 - Ola Hallengren's scripts
- **Create a custom job**
 - Need to consider volume, rate of change, skew of data
 - Let SQL Server tell you about your data
 - Profiler, Xevents for auto stats update (volatile)
 - sys.dm_db_index_physical_stats for volume
 - sys.dm_db_index_usage stats for update
 - Remember that updates will occur more often for a new database/new tables
- **There is no perfect option, the goal is to provide SQL Server's optimizer with a better understanding of the data**

Statistics (as defined by BOL)

“...a histogram and associated density groups (collections) over the supplied column or set of columns of a table or indexed view ... The query optimizer uses this statistical information to choose the most efficient plan for retrieving or updating data. Up-to-date statistics allow the optimizer to accurately assess the cost of different query plans, and choose a high-quality plan.”

Thank you!

